Housing Subsidies for Refugees: Experimental Evidence on Life Outcomes and Social Integration in Jordan*

ⓡAbdulrazzak Tamim[2] Emma Smith[3] I. Bailey Palmer[4]
Edward Miguel[5] Samuel Leone [6] Sandra V. Rozo[7] Sarah Stillman[8]

January 17, 2025

Abstract

Refugees require assistance for basic needs like housing but local host communities may feel excluded from that assistance, potentially affecting community relations. This study experimentally evaluates the effect of a housing assistance program for Syrian refugees in Jordan on both the recipients and their neighbors. The program offered full rental subsidies and landlord incentives for housing improvements, but saw only moderate uptake, in part due to landlord reluctance. The program improved short-run housing quality and lowered housing expenditures, but did not yield sustained economic benefits, partly due to redistribution of aid. The program unexpectedly led to a deterioration in child socio-emotional well-being, and also strained relations between Jordanian neighbors and refugees. In all, housing subsidies had limited measurable benefits for refugee well-being while worsening social cohesion, highlighting the possible need for alternative forms of aid.

JEL Classification: D22, J61, O17
Keywords: Refugees, Housing, Forced Migration, Social Integration

[1]Author order is Certified Random ⓡ (AEA Confirmation Code gEv7QgDTbi7f). We are grateful to Hasan Ebussuutoglu, Joaquin Fuenzalida, Jonathan Gorham, Peter Grinde-Hollevik, Mansi Kalra, Meghana Kumar, Charlotte McClelland, Ei Thandar Myint, Jonathan Old, Gufran Pathan, Benjamin Shenouda, and Andy Theocharous for their outstanding support as research assistants. We are also grateful for comments received from seminar participants at University of California Davis, ESOC, World Bank, George Mason University, Georgetown University, American University, and the BITSS Forecasting in the Social Sciences Workshop. University of California Berkeley IRB approval #2019-01-11713. This study is registered in the AEA RCT Registry (AEARCTR-0006141). We gratefully acknowledge financial support from Innovation for Poverty Action's Peace and Recovery Fund and the World Bank Development Research Group. We thank our Jordan-based implementing partner, and survey firm, Mindset Social & Marketing Research, and in particular, Mohammad Qardan, Rana Samara, and Mohammad Nababteh.

[2]University of California Berkeley, atamim@berkeley.edu
[3]Georgetown University, emma.smith@georgetown.edu
[4]University of California Berkeley, bailey.palmer@berkeley.edu
[5]University of California Berkeley and NBER, emiguel@berkeley.edu
[6]University of California Berkeley.
[7]World Bank, Development Research Group, sandrarozo@worldbank.org
[8]London School of Economics, s.v.stillman@lse.ac.uk

1 Introduction

Targeted assistance programs typically direct resources to specific vulnerable groups but may create tensions from non-recipients. This trade-off is of particular concern for refugee assistance: most refugees rely on humanitarian or government aid to meet many of their basic needs, but are also at risk of experiencing social exclusion or even xenophobia from hosting country citizens. This tension could reduce public support for assistance programs, and might set recipients back by eroding important social relationships with neighbors, landlords, and employers. Supporting refugees' needs thus requires understanding both the effectiveness of assistance programs at directly raising refugees' socioeconomic conditions as well as the social costs they may incur. We study this important and under-studied trade-off though a field experiment of a refugee-targeted housing subsidy program.

Housing is a particularly important issue for refugees and other displaced people who often face precarious shelter situations, yet there remains limited evidence on how best to improve housing stability outside of camp settings (Kumar 2021; Agness 2023). This is an issue that is, unfortunately, likely to increase in importance: the total number of forcibly displaced people has risen from 38 million in 2000 to approximately 118 million by 2024 (UNHCR 2023c), and is poised to increase further in the coming decades due to the adverse global impacts of conflict and climate change, especially in low- and middle-income countries (LMICs, Burke et al. 2015). Beyond the humanitarian assistance sector, there remains an active debate about how best to provide secure housing for low-income populations, with growing evidence on a variety of policies including housing vouchers, public housing construction, and transfer of home titles (Chetty et al. 2016; Kling et al. 2007; Kumar 2021; Ludwig et al. 2013).

This study examines the impacts of a randomized housing assistance program in Jordan on both the economic outcomes of Syrian refugee recipients and their social cohesion with Jordanian neighbors. Over 80% of Syrian refugees in Jordan live outside camps, making the issue of housing

security particularly salient.[1] The intervention, carried out in partnership with one of the largest humanitarian aid organizations in Jordan, randomized rental subsidies across 2,870 refugee households residing in Irbid and Mafraq governorates in Jordan. The randomization was implemented over geographic unit clusters and designed with sufficient statistical power to detect moderate impacts on important life outcomes, including living standards measures, labor market performance, and subjective well-being.

The housing subsidy covered, on average, a year of rent, combined with funding for landlords to upgrade the refugees' housing quality. Subsidies were offered for recipients' current housing arrangement, and thus their physical location (and by extension, their neighborhood) was typically held constant. To measure effects on refugee recipients as well as Jordanian neighbors, the study collected both in-person and phone surveys over a total of three and a half years since the start of implementation, with refugee households surveyed three times and neighbors once. The data collection succeeded in attaining relatively high sample tracking rates, overcoming some challenges associated with the COVID-19 pandemic. Although 86% of the refugee sample was surveyed at least once, there was some differential attrition between the treatment and control refugee samples, and therefore we estimate bounded effects following Lee (2009) (and the text highlights the results for which both estimated bounds take the same sign unless otherwise noted). These data are analyzed largely following the econometric models and primary outcomes specified in an AEA pre-registration (AEARCT#0006141) and associated pre-analysis plan, while making note of additional and exploratory results.

Another core contribution of this study is to examine the host community reaction to refugee assistance, utilizing a detailed survey of the attitudes and experiences collected among a representative sample of the Jordanian neighbors (N=2,146) of both treatment and control households. To our knowledge, this is among the first studies to experimentally examine how humanitarian assistance to refugees affects the views of the local communities who do not directly benefit. Specifically, we examine whether refugee-targeted transfers — in this case via rental payments

[1] Statistic calculated directly from data on the universe of Syrian refugees in Jordan registered with UNHCR.

and housing improvements — affect host communities' policy views, altruism, and interactions with refugees.[2] The study relates to a literature on transfers and potential negative spillover effects on non-recipients (Haushofer and Shapiro 2016, Haushofer and Shapiro 2018). But unlike other settings where transfer recipients are members of the local community, this study measures non-recipients' reactions to refugees — who are outsiders — receiving in-kind transfers. The closest concurrent studies to this one are Baseler et al. (2023) and Beltramo et al. (2024), both of which examine the effects of refugee assistance programs that also directly benefit host communities by, for example, granting them cash transfers. Of course, due to financial constraints, most humanitarian assistance is limited to forcibly displaced populations and not host community members, as in our setting.

Take-up of the housing assistance was moderate at 33%, and multiple factors contributed to this relatively weak first stage relationship. For one, almost a fifth of homes were deemed ineligible for assistance, typically because the implementing partner viewed the housing shelter as too precarious to support (i.e., temporary structures), and occasionally because it was determined that the residence did not need physical upgrades. There is also qualitative evidence that low take-up was at least in part due to some landlords' reluctance to make signed legal commitments to refugees for the lease, and with the implementing partner for the necessary construction. This is a setting in which most rental contracts are informal and landlords have ample discretion over their terms and a largely free hand to evict tenants. Despite the appeal of guaranteed rent for a year plus funding for housing upgrades, some landlords preferred not to "bind" themselves to the program and the particular recipient refugee household currently residing in their property.[3]

One of the study's main empirical findings is that we detect no significant positive impacts of the housing assistance program for refugees along a range of pre-specified primary outcomes, with the exception of housing expenditures where there is the expected (and somewhat mechanical) drop in spending. Beyond housing expenditures, the other primary pre-specified outcomes in-

[2] We do this by surveyed neighbors but excluding landlords, who were often involved in the program.

[3] The data underlying these findings is from our implementing partner's Integrated Assessment Shelter Analysis.

clude an aggregate housing quality index, total household consumption, respondent mental health, and child socio-emotional well-being, measured using the standardized Strengths and Difficulties Questionnaire (SDQ). In fact, the analysis unexpectedly shows that the socio-emotional well-being of children in treated households decreases substantially, by 0.34 standard deviations on average; we return to interpretation of these patterns below. Due to the non-compliance noted above, estimated instrumental variable (IV) treatment on the treated (TOT) effects are relatively imprecise. But the primary analysis, which pools across multiple rounds of follow-up survey data, does allow us to reject the existence of some moderate positive treatment effects.

A second main finding of the study is that the refugee-targeted housing subsidy program led to a meaningful and statistically significant deterioration in community perceptions towards Syrian refugees among the Jordanian neighbors of treatment households. We estimate a 0.33 standard deviation unit decrease in an index of their social attitudes and perceptions of refugees. The analysis also shows that neighbors become better informed about how much assistance Syrian refugees receive. There is no evidence of effects on neighbors' own economic outcomes. We speculate that the substantial transfers received by treatment refugee households were observed by Jordanian neighbors and led to tensions in some cases, perhaps especially given that these neighbors also live in the same relatively poor neighborhoods as the refugees and are typically not well-off themselves. The easy visibility of assistance in this case – including construction activity in the treated housing units – may have raised the salience of the intervention relative to some other forms of aid (i.e., mobile money transfers or health care subsidies) that are easier to keep private.

The estimated null and even negative results for most pre-specified primary outcomes are in contrast to the predictions of experts. Building on an increasingly common practice in economics and other social sciences (DellaVigna and Pope 2017; DellaVigna et al. 2020), the research team gathered forecasts from both researchers and policy experts regarding the primary impacts of the program. Forecasters generally predicted positive but modest program impacts on refugees, and predicted zero effects on average for Jordanian neighbors. They were thus were somewhat more optimistic about likely program impacts than the estimated effects, though in most cases we are

unable to reject equality between the estimated average impacts and mean forecasts.

Disaggregating the data by survey round assists in understanding the predominantly null impacts on consumption and adult well-being. In the short run (while the program was ongoing), treated households experienced meaningful gains in housing quality and financial stability: they reported increased access to clean water and less reliance on loans. These findings indicate that program investments were able to achieve some of their immediate goals, specifically, in terms of improving the housing quality experienced by recipients, and preventing them from accumulating more debt (much of which is rental related debt in this population). In fact, on average the control group holds 4 to 5 months of rental debt, so potentially reducing this burden is a meaningful benefit for treated households.

Yet both food security and self-reported health worsened: household hunger increased for several measures and more household members exhibited COVID-19 symptoms. We find several pieces of evidence that these negative effects follow from unanticipated redistribution of aid. In particular, we document a meaningful and statistically significant reduction in food aid that treatment households report receiving (from both formal and informal sources), roughly equivalent to 10% of monthly pre-pandemic income. There is also suggestive evidence that treated households received an inflow of additional household members (especially adolescents), further diluting program benefits in per capita terms.

Data from the second round of surveys (collected shortly after assistance had ended) indicates that treated households continued to report reduced housing expenditures as well as increased savings, but at the same time reported a significant decrease in subjective well-being. In the final survey round, one to two years after assistance had ended, the only significant treatment effect that survives a multiple testing adjustment is the substantial decrease in child socio-emotional well-being, noted above.

The effects of the refugee housing assistance program over time suggest a possible link between the negative impacts on child well-being and neighbor attitudes. While all other treatment effects

on refugee households dissipate by the final round of surveying (conducted more than a year after assistance had ended), the negative impacts on child outcomes persist, and the negative impacts on neighbors' perceptions towards refugees are collected at a similar time point (about a year after program implementation had ended). For example, two components of the neighbor social attitudes index that are statistically significant capture whether the children of the Jordanian household have Syrian refugee friends, and whether the adult Jordanian neighbor has close Syrian friends. Taken together, the timing of child effects and the data from neighbors offer suggestive evidence that the deterioration in child well-being may be driven by worse social relations with the treated refugee households' Jordanian neighbors.

In all, the empirical findings indicate that this direct housing assistance program offered limited short-run economic improvements that dissipated following the program, while negative psychological and social cohesion effects proved more lasting. Not only was take-up of the housing assistance quite low (at around one third), but impacts were minimal or even negative for some measures. The negative response from neighbors parallels existing research on targeted transfers, which has documented negative psychological impacts on nearby non-recipients of cash transfers in some cases, although the evidence is notably mixed (Haushofer and Shapiro 2016, Haushofer et al. 2015, Baird et al. 2013, Egger et al. 2022). Unlike the above studies, this paper finds that tensions among non-recipients persist more than a year after the transfer implementation period, and that these negative effects might even undermine the well-being of program recipients and their children via reduced social ties between Jordanians and refugees, highlighting the fragility of social cohesion between refugees and host country citizens (Zhou 2019).

If programs like this one are in fact not improving refugees' lives, an urgent question revolves around the feasibility of other policy approaches. A growing body of literature demonstrates that direct cash transfer programs yield meaningful improvements for households, at least while assistance is ongoing (Hidrobo et al. 2014; Özler et al. 2021; Quattrochi et al. 2022; Moussa et al. 2022; Aygün et al. 2024; Haushofer and Shapiro 2016), and cash is one immediate alternative that would likely be less observable to non-recipients. In line with this, and echoing the largely null

evaluation results, a large majority (70%) of refugee respondents in this study stated in the endline survey that they would have preferred cash assistance of equivalent value to the rental subsidy. At the same time, such cash assistance would not do much to improve the structural impediments to economic and social integration facing refugees in most settings.

2 Background and Context

2.1 Syrian Refugees in Jordan

As of 2023, the Syrian crisis was one of the largest displacement crises in the world, resulting in 6.8 million internally displaced Syrians and another 6.5 million Syrian refugee abroad (UNHCR 2023c). The war began in 2011 with pro-democracy demonstrations against President Bashar al-Assad, whose government responded with lethal force against the protesters, leading to escalating violence that drove the country into civil war. Considering Syria's 2010 population was 21 million, more than half of its population was displaced during the study period (UNHCR 2023b).

Jordan hosts roughly 650,000 registered Syrian refugees, accounting for about 6% of its population of 11.1 million people (UNHCR 2023a). While the Jordanian government is not party to the 1951 Refugee Convention nor its 1967 Protocol, which ensure dignity and rights for refugees in signatory countries, it does uphold the principle of *non-refoulement*, which mandates that refugees cannot be forcibly returned to countries where they face persecution.

In response to the influx, the Jordanian government opened multiple refugee camps in coordination with UNHCR, including *Za'atari*, *Azraq*, and *Mrajeeb Al Fhood*. Despite the availability of refugee camps, individuals were granted free mobility and, in fact, over 80% of Syrian refugees currently live in local communities and not in camps (UNHCR 2023a).

2.2 Humanitarian Shelter Program

In Jordan, there is limited secure and affordable housing available for Syrian refugees. The UNHCR Quarterly Analysis of Refugees in Jordan found that in 2022, 25% of Syrian households perceived a current threat of eviction.[4] Additionally, one in four refugees applying for rental assistance from a major Jordanian humanitarian organization had moved at least three times in the previous year (NRC 2015), demonstrating the widespread precariousness of their housing situation. Refugees often rely on humanitarian assistance to manage rising rental costs, but these funds are insufficient to cover the full expense, and hence 70% of Syrian households living in local communities reported having rental debt (Hagen-Zanker et al. 2018; Lehmann and Masterson 2014). Frequent relocation due to rental affordability issues and a lack of stable rental contracts may also make it challenging for refugees to maintain necessary documentation and access public services (NRC 2015).

Given this situation, the implementing partner designed the Shelter Program to secure more stable housing for Syrian refugees living in host communities. The intent was also to use the program to facilitate social cohesion between refugees and hosts by including direct benefits for the landlords renting to refugees. It focuses on Syrian refugees in the country's northern governorates of Irbid and Mafraq, which are adjacent to the Syrian border and contain a large share of the country's displaced persons.

The overall Shelter Program is composed of several different assistance modalities, including projects focused on water and sanitation, eviction protection, and "inclusion kits" for individuals with disabilities. This study evaluates the impacts of a housing shelter modality, which we will refer to in what follows as the Housing Subsidy Program (HSP).[5]

Under HSP, the implementing partner directly compensated Jordanian landlords for rent on behalf of their refugee tenants, with the refugee household receiving rent-free housing for 9 to

[4]Determined using the Vulnerability Assessment Framework, a bi-annual survey of registered refugees in Jordan.

[5]The pre-analysis plan proposed the evaluation of the far less generous energy subsidy program as well. That will be the subject of future research.

18 months, depending on negotiations with the landlord. These negotiations were one source of variation in how the subsidies were spent: some landlords agreed to participate in the rental subsidy program for 12 months while others agreed to a shorter or longer period. The program was provided to refugees in their existing rental arrangements, building on established tenant-landlord relationships.

The program offered several secondary benefits in addition to direct rental payment. For instance, the implementing partner paid for physical housing improvements, such as door and lock installation, roof and wall construction, and mold removal. The program thus benefited not just the Syrian refugee tenants but also their Jordanian landlords, as well as the overall Jordanian housing stock. Together, the rental assistance plus the construction improvements were valued at USD $2,200 per unit on average, a considerable sum in this context, where per capita income was roughly USD $4,100 in 2019.[6] The program also formalized refugees' tenancy agreements: Jordanian landlords and Syrian refugees signed standard tenancy agreements in line with Jordanian law, and the implementing partner and the landlord entered into a contract for the construction work on the property. Finally, the program required landlords to not raise rent for an additional year following the end of assistance, possibly generating further housing savings for treatment households.

3 Experimental Design and Data

This study employs a Randomized Control Trial (RCT) with 2,870 Syrian refugee households, and surveys of 2,146 neighboring host-community households, to evaluate the direct impacts of the housing subsidy program on refugees as well as the spillover effects onto host community neighbors. The study additionally assesses the accuracy of forecasts of program impacts of the program from research experts and policymakers gathered via the Social Science Prediction Platform. The

[6]This is according to the World Bank's World Development Indicators.

research design and primary hypotheses were pre-registered before program roll-out began.[7] Given the ethical complexities involved in conducting research in humanitarian settings, the research team has written an ethics appendix detailing the decisions made to maintain a high standard of ethical conduct in study design, implementation, and data collection (see Ethics Appendix).

Study eligibility. Participants were identified by the implementing organization which disseminated information about the program through their networks, resulting in a pool of refugee applicants that exceeded the sample they had resources to assist. Given this over-subscription, applicants' opportunity to receive HSP assistance was randomized. The implementing partner then conducted a standardized vulnerability assessment. Households found to be in the most vulnerable 10% of the sample were guaranteed access to the program (i.e., were not in the randomization), while the least vulnerable 10% of households were excluded from the program (and the randomization). The remaining 80% of households with more typical levels of vulnerability were then eligible to receive the program if residing in a treatment community, as described below.

Treatment Randomization: The housing subsidy program was randomized geographically at the community level.[8] In the first step, 158 communities in Irbid and Mafraq were randomized into treatment or control for HSP assistance, stratifying on governorate and district population quartile. One third of communities were randomly assigned to treatment, while the remaining two thirds were randomly assigned to control. All eligible applicants living in the treatment communities were assigned to treatment. In all analysis below, error terms are clustered at the community level. There were several reasons for the cluster randomized design, including the ability to streamline implementation (and thus lower program costs), to reduce conflict among refugee households assigned to treatment versus control, and to improve the analysis of treatment spillovers onto host community members by boosting the local saturation of program assistance, thereby making it more salient to neighbors.

Refugee Sample: The survey sample consists of all refugees assigned to treatment and an

[7] See AER RCT Registration Number AEARCTR-0006141.

[8] Communities correspond to the Jordanian government's administrative unit of "localities". Communities in Mafraq have 3,866 people on average, while those in Irbid are somewhat larger at 14,626 individuals on average.

equal number of randomly selected households in the control communities who had applied for the program. Figure 1 illustrates the location of the treated and control communities in Jordan.

Jordanian Host Community Sample:

A key concern with refugee assistance programs is the potential for social backlash from host communities. The HSP was intentionally designed to mitigate such effects among landlords, by disbursing funds directly to landlords for housing improvements, as noted above. The social cohesion component of the design focuses instead on measuring the impacts on attitudes of Jordanian neighbors in the community, who are were not eligible for HSP assistance, rather than on landlords. The Jordanian neighbor sample consists of 1,455 Jordanians living near the treated home location at the time of implementation. The sample was identified in 2022, a year after program aid had ended, through in-person surveying.

Enumerators used a randomized algorithm to select a Jordanian neighbor living close to the original home of the refugee participant (from among the set of nearest neighbors). This selection was based on the refugee participant's home location at the time of study randomization, irrespective of any subsequent moves. The algorithm was successful at identifying and surveying a close neighbor: the median distance between the selected neighbor and the original refugee participant's home is just 63 meters (roughly 200 feet).

3.1 Data on refugee outcomes

The study collected three rounds of surveys tracking participants' outcomes during ("midline"), immediately after ("endline"), and 1.5 years after all assistance was delivered ("follow-up"). The outcomes to be examined were specified in detail in the pre-analysis plan and divided in three main groups including:

1. *Proximate outcomes:* This comprises outcomes that were expected to change directly due to program implementation, and include housing expenditures and a housing quality index. As

described above, HSP households received not only full rental subsidies for 9 to 18 months but also physical upgrades to the home environment. These proximate outcome measures can be seen as capturing successful program implementation.

To measure physical improvements to the shelter, we constructed an index of self-reported floor, roof, and wall quality; electricity and water access; and crowding. Housing expenditures were measured in two ways: First, in the midline phone survey, respondents reported their total monthly housing expenditures, including the payments provided by the implementing organization. Second, in the endline in-person survey, once assistance had ended, participants reported their monthly out-of-pocket housing expenditures (excluding the amount paid by the implementer). The latter is our preferred measure of the direct effect of the program on housing expenditures, because it allows us to test whether treated households experienced rental savings. Because the endline measure was collected after assistance ended, any observed expenditure reductions would most likely be an underestimate of the benefits experienced during the program. Recall that HSP also required the landlord to not raise rent for an additional year after the assistance ended, which could help account for any persistent effects on housing expenditures in the treatment group.

2. *Primary well-being outcomes:* This category includes three main outcomes that seek to capture refugee household well-being, including household consumption, mental health (measured through the Center for Epidemiological Studies Depression Scale, CES-D), and an index summarizing the 25-item child Strengths and Difficulties Questionnaire (SDQ) that measures emotional and conduct problems, inattention, peer relations, and prosocial behaviors. Both the CES-D and SDQ are validated tools that have been utilized in a variety of international contexts (Park and Yu 2021, Woerner et al. 2004). In each of the three survey rounds, the primary respondent completed the CES-D depression screening. The SDQ was completed by an adult respondent regarding a randomly selected child aged 3 to 8 years old in the endline survey, as well as for the same child in the one and a half year follow-up. Total household consumption was measured only at endline (which was collected in person).

3. *Secondary outcomes among refugees:* These include a broader set of variables that were grouped into 14 broad families in the PAP including: (1) dwelling characteristics and household structure, (2) consumption and expenditure, (3) financial participation, (4) earnings, labor, and occupational choice, (5) migration, (6) physical, mental health, and sleep, (7) marriage and fertility, (8) child outcomes, (9) social capital, (10) political attitudes, (11) time use, (12) education and cognition, (13) behavioral games and preferences, and finally, (14) specific COVID-19 related outcomes. The three primary outcomes noted above are a subset of these measures. Certain outcomes were collected in each of the three survey rounds, for instance, the food security measures (the number of meals eaten and the frequency of going to bed hungry). Respondents also reported information on child school attendance, and completion of learning activities when schools were closed due to the COVID-19 pandemic. In the midline survey collected in 2020 (during the pandemic), the physical health outcomes included additional questions focused specifically on COVID-19 symptoms and treatment but these were dropped in later rounds as the pandemic had eased.

3.2 Measures of social cohesion among neighbors

We pre-specified four primary outcomes regarding the attitudes of Jordanian neighbors:

1. *Interpersonal social attitudes and perceptions:* This was assessed using an index derived from questions about social ties between the Jordanian respondents and Syrian refugees, attitudes regarding social proximity, and opinions on refugees' contributions to society.

2. *Economic attitudes and perceptions:* Focused on Jordanians' perceptions of the impact of Syrian refugees on the Jordanian economy, based on survey questions.

3. *Altruism:* Measured through a dictator game, this outcome gauged altruistic behavior towards refugees, with real monetary incentives provided for a random subset of the sample to

promote genuine responses.[9]

4. *Policy preferences:* This index summarized the respondents' stances on various policies related to Syrian refugees, including their living arrangements (e.g., a hypothetical requirement to live in camps) and employment rights.

In addition, the neighbor survey collected comparable measures of economic and psychological well-being to those measured among the refugee sample, including housing expenditures, total consumption, and mental health, to assess program spillovers.

3.3 Data sources

The study employs four different data sources, with the first obtained from partners at the implementing organization, and the other three original survey data sets collected by the research team. They include:

1. *Baseline administrative data:* The implementing organization collected baseline data from program applicants to establish their eligibility for the HSP. This assessment form collected information on household demographics, housing quality, health and disability, employment, and education, which were used to construct a housing vulnerability index employed as a baseline covariate in the empirical analysis.

2. *Surveys collected among refugee households:* The study collected surveys during program implementation (the midline survey), immediately after (endline survey), and 1.5 years after all assistance was delivered (the follow-up survey) (see Figure 2). In the midline survey, 1,619 participants were surveyed by phone in 2020, during the first year of the COVID-19 pandemic. The endline survey in 2021 collected in-person data from 1,534 participants shortly after HSP assistance had ended. Finally, in the 1.5 year follow-up collected in late

[9] A random one third of respondents were offered financial incentives whereas the others were told it was hypothetical. We cannot reject that the same choices were made in both groups on average.

2022 to early 2023, 1,444 participants were reached for a phone survey. In total, the three rounds of data cover a wide range of outcome families including those listed out above and several others; these outcomes are described in further detail in the appendix and associated pre-analysis plan.

3. *Neighbor surveys:* A round of in-person data collection of the Jordanian neighbors of the study participants was gathered in 2022 roughly one year after the program had ended. In total, 1,455 Jordanian neighbors of refugee participants were surveyed.

4. *Forecast short surveys:* Two rounds of brief 10 minute forecasts were collected to gather the predictions of both academic and non-academic experts on the likely impacts of the program on the primary outcomes. Each round of predictions collected information for approximately 60 individuals, with the first round focused on impacts on study participants, and the second on impacts among neighbors.

Figure 2 depicts the data collection timeline and the response rate of each of the surveys. Overall, 86% of the main refugee study sample was surveyed at least once: 80.3% were surveyed at midline, and at least 75% were surveyed in each of the endline and follow-up rounds. Among the treatment group, survey completion was 6 to 8.6 percentage points higher in each survey round, as shown in Table 3, and this differential attrition is statistically significant. As such, beyond the pre-specified TOT results presented, the analysis was expanded to include Lee bound estimates of the treatment effects (following Tauchmann 2014).

3.4 Descriptive statistics

We present characteristics of the sample using baseline vulnerability assessment in Table 1.[10] Several patterns are noteworthy. First, a quarter of respondents likely have a disability, calculated

[10]The table only includes data for 1,619 out of the 2,017 households sampled for the study since the implementing partner only facilitated data on this assessment for the individuals found in the midline sample. The table also presents statistics for gender and age, which are taken from the midline survey.

using the Washington Group Short Set on Disability.[11] Sharing a home is common, likely as an economic coping mechanism, with on average of 1.3 families living in the same house. Housing quality is low, with most houses in some state of disrepair or incomplete construction.

Table 1 also illustrates descriptive statistics for the treatment versus control households to assess balance across groups, where the third column presents a mean difference test. Respondents across the study arms are largely balanced by age (with an average age of 34 years), marital status (84% married), and incidence of a disability (Panel A). There is a slight imbalance in the share of respondents who are female, with the proportion slightly higher in the control group. Moreover, average household size is just above five members, with nearly three children on average, and these are balanced between treatment and control, as is the number of families per residential unit (Panel B). Refugee households in the sample face challenging and precarious housing conditions, but these characteristics are generally balanced across groups (Panel C). For example, only 66% of households have access to piped water, 22% have functional windows, and 44% completed floors. While the majority of refugees plans to stay in the same housing unit (92%), a large share have moved shelter in the last year (with an average of 0.44 and 0.47 moves in the control and treatment groups, respectively).

Across all baseline characteristics presented in the table, the p-value on the hypothesis that all differences are zero is 0.102 (using a Chi-squared test), not quite significant at traditional confidence levels. This finding together with the fact that there are few economically meaningful differences across treatment arms in terms of respondent, household and shelter characteristics – out of the 18 covariates examined only two exhibit a statistically significant difference at 95% confidence – indicates that the study randomization generated largely comparable groups of households.

[11] We use the recommended cut-off (level 3) to classify respondents as being disabled or not based on six domains: seeing, hearing, walking, cognition, self-care, and communication.

3.5 Forecasts of Program Impacts

The forecasters were mainly researchers working on topics in migration and development economics, or those with expertise in humanitarian program implementation, and were asked to predict effects on the program's primary outcomes. For round 1, a total of 61 individuals, comprising 37 researchers and 23 non-researchers, responded to surveys about the program's effects on refugee well-being. For round 2, 63 respondents, including 53 researchers and 10 non-researchers, provided predictions on the effects on neighbors' social cohesion responses. Gathering forecast data allows us to assess whether estimated impacts were in line with the prior beliefs of research and policy experts. Table 2 presents the mean predicted effects, as well as the interval from the 10th to 90th percentiles of predictions. Generally, forecasters anticipated modest improvements in refugee well-being, particularly regarding housing outcomes, while predicting no average impact on the social cohesion measures collected among neighbors.

4 Empirical Strategy

4.1 First stage compliance with program assignment

Table 3, Panel B presents the first stage analysis and the HSP take-up rate of 33%. While statistically significant, the compliance rate is lower than expected, especially given the substantial funding offered by the program. It is likely attributable to several factors, including implementation challenges during the COVID-19 pandemic, uncooperative or reluctant landlords (who may have feared being locked into long-term contracts with refugee tenants, as noted above), and the implementing partner's policies that led some homes to be excluded due to their condition.

4.2 Estimating program impacts on refugee well-being

The primary approach to estimate program treatment effects on the treated (TOT) is instrumental variables (IV) estimation, as represented by the following two equations:

$$T_{ic} = \alpha_0 + \alpha_1 Z_c + X_c' \Lambda_1 + W_{ic}' \Gamma_1 + \mu_{ict} + \eta_{ic} \tag{1}$$

$$y_{ict} = \beta_0 + \beta_1 \widehat{T}_{ic} + X_c' \Lambda_2 + W_{ic}' \Gamma_2 + \mu_{ict} + \epsilon_{ict} \tag{2}$$

Equation (1) estimates the first-stage effect of assignment to treatment on the household's take-up status (with results presented above), and equation (2) estimates the TOT effects. To enhance statistical power, data from the midline, endline, and follow-up surveys are combined in the main analysis when possible, as pre-specified. The data includes multiple rounds t of survey data for each household, and the data are stacked. T_{ic} denotes the treatment take-up for household i in community c. Z_c is an indicator variable signifying whether community c was randomly assigned to HSP treatment, where all eligible individuals were then assigned to the treatment. The outcome variable of interest is y_{ict}, and the predicted treatment take-up, $\widehat{T}_{ic}$, is derived from equation (1).

The vector X_c encompasses community stratification variables used in the randomization, including indicators for whether a community is in Mafraq or Irbid and the community's district's population quartile[12] The vector W_{ic} includes individual baseline covariates, accounting for household-level demographic variables from the integrated assessment; in particular the vulnerability quartile, month of assessment, household size, number of children, and respondent gender and age. μ_{ict} is a vector of enumerator-round fixed effects. η_{ic} and ϵ_{ict} are error terms and are clustered at the community level, which is the level of randomization.

To assess dynamics, the analysis also estimates TOT effects separately with the midline, endline, and follow-up data, while the regression specification remains effectively the same as previously described, with enumerator fixed effects (rather than enumerator-round fixed effects). The

[12] As noted before, communities are defined by the Jordanian administrative unit "locality". "District" is a more aggregated, administrative geographic unit used in Jordan.

round-by-round results discussed below tend to focus on estimates that survive a multiple testing adjustment within each domain of outcomes, namely the false discovery rate (FDR) adjustment (Anderson 2012), with a q-value of 0.10 or less.[13]

4.3 Bounding estimated effects in the presence of non-random attrition

Although an overall high tracking rate was achieved, with 86% of the refugee sample surveyed in at least one survey round, there was differential attrition between the treatment and control refugee samples. As detailed in Table 3, the gaps in attrition are statistically significant, and therefore we estimate bounds following Lee 2009 and report these in the appendix. The text highlights results for which both estimated bounds take the same sign; when the bounds are on opposite sides of zero, this is noted in the text and the corresponding results are discussed as suggestive.

4.4 Estimating impacts of the program on neighbor's attitudes

The estimation of HSP effects on neighbors is similar to the IV approach carried out in the refugee sample, where the treatment status of the refugee household is applied to their surveyed neighbor:[14]

$$T_{ic} = \delta_0 + \delta_1 Z_c + X_c'\Phi_1 + W_{ic}'\Psi_1 + \mu_{ic} + \rho_{ic} \tag{3}$$

$$y_{ic} = \theta_0 + \theta_1 \widehat{T}_{ic} + X_c'\Phi_2 + W_{ic}'\Psi_2 + \mu_{ic} + \epsilon_{ic} \tag{4}$$

T_{ic} equals one if the neighbor's associated refugee household received treatment and zero otherwise, and Z_c equals one if the community was randomized to treatment. The X_c and W_{ic} terms are as above, μ_{ic} are enumerator fixed effects, and standard errors are again clustered at the commu-

[13] In a few instances, a result with a q-value greater than 0.10 is discussed to contextualize a related result and these cases are noted in the text. Throughout the paper, when two similar outcomes have significant q-values, we tend to report only one of them in the main paper to streamline exposition; refer to the online appendix for estimated effects for all pre-specified outcomes.

[14] While it is conceptually possible (as outlined in the pre-analysis plan) that the program might generate economic and labor market spillovers to untreated refugee households and neighbors, the relatively low saturation level of the experiment (at a share of around 0.5% of refugees treated) makes this unlikely.

nity level. While we focus on this pre-specified analysis of neighbor impacts, the PAP contained several additional secondary or more exploratory analyses, including the estimation of ITT effects, the evaluation of the effects of the program using a continuous treatment variable (measured in either months of treatment and the total cash value of the transfer[15]), the assessment of treatment effects on refugees' economic convergence with their neighbors, and estimation of heterogeneous effects (based on various demographics, see the full set of pre-specified results in the Appendix).

4.5 Estimating the accuracy of forecasts

To determine whether mean forecasts differed significantly from the estimated program impact, the analysis takes into account both the estimation error in the estimated program impact ($\widehat{\beta_1}$ and $\widehat{\theta_1}$ in the second-stage estimation equations above) as well as the observed range of forecasts, using 1000 simulated draws of each. For each draw, the difference between the two values is taken, yielding the distribution of the differences. Statistical significance is determined by inspecting whether zero lies within the 95% confidence interval of the distribution of this difference; if zero lies outside the interval, we reject the hypothesis that the mean forecast and estimated treatment effect are equal.

5 Housing Subsidy Program Impacts on Refugee Life Outcomes

We first estimate program impacts on refugees pooling survey rounds, and then explore dynamic effects by rounds. As noted, we generally focus on results that both survive a multiple hypothesis testing adjustment and do not change sign when bounded to account for differential attrition.

[15] Since the implementing organization negotiated the length of the contract and the amount of reduced rent with each landlord, there is some non-random variation in both across treatment households.

5.1 Average program impacts on refugees pooling survey rounds

Table 4 presents the estimated treatment effects on refugees for the pre-specified primary outcome measures, as well as statistically significant effects on other outcomes.

In terms of primary outcomes, the estimated treatment effect on the housing quality index is positive and moderate in size, at 0.31 standard deviation units, though not statistically significant at traditional confidence levels. Regarding rental payments, there is a significant reduction in households' out-of-pocket housing expenditures of 82.05 USD (SE 32.07) and the estimate survives the False Discovery Rate (FDR) multiple testing adjustment (q-value=0.07). This measure was reliably collected at the endline survey when the subsidy itself was ending for some households but the program's contractual rental level controls should still have been in place. Taken together, the two findings indicate that the program at least partly achieved some of its main proximate objectives of improving housing quality and reducing the rental burden for refugee households, although the housing quality estimate is suggestive. As we are unable to directly measure the reduction in out-of-pocket housing expenditures during the midline survey (as noted above), the endline estimate likely understates the true program benefit to treatment households in terms of reduced rental payments.

In a central finding of the study, we next show that the average estimated program effect on consumption across all survey rounds is small and not statistically significant (Table 4). Specifically, across all rounds for which data are available (which varies across the measures, see notes in Table 4), there are no significant impacts on household food consumption nor log total household consumption expenditures. Likewise there is no significant effect on respondent depression as assessed using the widely used CES-D scale. It appears that a substantial year-long housing subsidy plus shelter upgrade was not sufficient to meaningfully improve these key living standards measures, although it is worth noting that, as with the housing quality effects, the relatively weak first stage increases standard errors on these IV estimates. Despite this, these estimated null effects are still fairly precise: the associated standard errors suggest that an effect of 0.16 log points on

consumption expenditures would be statistically significant at 95% confidence, and analogously a 0.25 standard deviation effect on the CES-D depression scale would be statistically significant.

Unexpectedly, the one large and statistically significant result suggests that the program led to worse outcomes: the housing subsidy program leads to a decrease in the socio-emotional well-being of children (measured using the SDQ scale) of 0.34 standard deviation units, with a multiple testing adjusted q-value of 0.07.

Beyond the primary pre-specified outcomes, all other outcomes with statistically significant estimated program effects pooled across rounds that survive a multiple hypothesis testing adjustment are reported in Table 4, Panel B. One striking finding is the large negative impact on treatment households' food security. In particular, adult respondents in the treatment group were 18 percentage points more likely to go to sleep hungry in the past week (46% more likely) than those in the control group (FDR q-value < 0.01), and children and other adults in treatment households were 15 and 13 percentage points more likely to go hungry in the past week (q-values 0.03 and 0.01), respectively. This may be surprising since we did not document a significant reduction in household food expenditures. However, the midline results discussed below highlight important changes in food aid received, household structure, and COVID-19 incidence that appear to have contributed to this finding. The hunger results are not fully robust to using Lee bounds to account for non-random attrition (Panel B, Table A.1), so we interpret these findings as suggestive.

Yet at the same time – and perhaps due to the fact that the program reduced housing expenditures – the program had positive impacts on household savings: pooled over the endline and follow-up survey rounds, treatment households were 8 percentage points more likely to have 30 JDs (roughly 95 USD PPP) in savings, on a base of just 10 percent in the control group (and the q-value of the difference is 0.05, although the finding is less robust to Lee bounds). Though the magnitude of savings is not large, the evidence that households both increased savings and reported more hunger suggests that households may have valued future savings above and beyond current consumption, perhaps as a precaution against anticipated future negative shocks. There is

also an insignificant but suggestive reduction in rental debt in the endline, equivalent to approximately a quarter of average rental debt in the control group, where the average household owes the equivalent of four months of rent. There is also a smaller but marginally significant reduction in rental debt in the follow-up 1.5 years after implementation. Additionally, we find some indication that monthly rent paid to landlords (including from the implementing organization) increases in the treatment group, consistent with more reliable rental payments through the program (see Appendix).

There were no statistically significant impacts on other outcome measures that were collected across multiple rounds, including for household composition, migration, and employment. However, there are some meaningful effects in particular survey rounds and we turn to them next.

5.2 Impacts in the midline round

Recall that the midline phone survey was conducted while the HSP was ongoing and many households were still receiving assistance. The midline survey data indicates that HSP improved several dimensions of housing quality in the short run (Panel A, Table 5): households experienced a 21 percentage point increase in access to clean water (SE 0.06, control mean 17%) and a 0.51 room increase in dwelling size (SE 0.17), likely the result of door installation and other construction. Yet, as in the pooled results, there is not a statistically significant change in the overall housing quality index.

Food security outcomes at midline echo those in the pooled sample and also shed light on the possible mechanisms (Panel B, Table 5). As in the pooled results, respondents were more likely to report that they and others went to sleep hungry in the past 7 days, with a robust 23 percentage point (q-value = 0.01) increase in respondents' hunger, and other adults and children in the household being 17 and 24 percentage points more likely, respectively, to go to sleep hungry (SE 0.06 and SE 0.07). Most importantly for understanding why, treatment households report substantial decreases in food aid received, consuming 53 USD PPP less in food from assistance

(q-value = 0.03), which represents a 24% decrease in food assistance relative to control. This is a nontrivial proportion of households' budgets: it is equivalent to over 10% of households' total monthly labor income pre-pandemic (as control households reported 218 USD PPP in average monthly food assistance and pre-pandemic monthly labor income of 503 USD PPP). It is worth noting that this measure of assistance includes both formal assistance, such as from the UN High Commissioner for Refugees and the World Food Program, as well as informal assistance from local organizations and community members. Regardless of the channel, this substantial decrease in the household's food budget seems likely to have contributed to worse household food security. This is in line with conversations that the research team had with several humanitarian aid organizations during the pandemic, in which some implementers sought to allocate emergency assistance to households that were not already benefiting from large forms of assistance, such as the HSP.

The midline survey data also indicates that there were several other unexpected short-run changes in household outcomes. For instance, it appears that changes in household composition may have further contributed to household hunger (Panel C, Table 5). Treated households at midline report significantly more children than control households (point estimate 0.3 children, q-value=0.03), and this effect is driven by an increase of 0.16 more adolescent boys (q-value=0.33). This movement between households can be understood as a form of redistribution within the Syrian refugee community, by which treated households shared some of the program benefits of the HSP but at the potential cost of food security to the full household. Despite the possible negative effect on household food security, the study cannot rule out that recomposition might have improved the well-being of the new household members who moved in (and the households they came from), if those adolescents came from relatively worse-off households. (Similarly, the targeting of food assistance away from HSP treatment households noted above could have benefited other refugee households that received additional assistance.) Nonetheless, these individuals' joining the treated households appears to have likely come at a cost for other treated household members. The effects of the program on household recomposition, however, is not fully robust when correcting for non-random attrition (Panel C, Table A.2), and as such should be viewed as suggestive.

Treated households also report experiencing increased COVID-19 exposure during the mid-2020 period when the survey was collected, which coincided with a peak period of the pandemic (Panel D, Table 5). Treated participants displayed an average of 0.84 additional COVID-19 symptoms compared to a control group mean of 0.47 (q-value $<$0.01), and treated households had 0.46 more symptomatic adults and 0.17 more symptomatic children (q-value$<$0.01 and q-value=0.01, respectively). And perhaps related to this health shock, there is lower weekly adult income in the treatment group, which decreased by 20 USD PPP (SE 7.19, q-value = 0.02), during Jordan's first COVID-19 lockdown. This reduction in income, however, did not persist after lockdowns were lifted (see Appendix). These households were also 16 percentage points more likely to report non-adherence to social distancing guidelines (q-value = 0.08). Increased household size, as noted above, may have contributed to increased COVID-19 exposure, especially if individuals from other households were likely to be moving in and out of the home. Another more speculative explanation is that the HSP may have allowed household members to travel outside the home more often, for instance, by improving the security of the home through better window, door and lock installations. As noted above, analysis also shows an increase in women's physical mobility (Table B.1). Perhaps surprisingly, treated households were significantly less likely to have visited healthcare institutions at midline, with 0.56 fewer visits on average compared to 1.14 in control (q-values $<$ 0.05). A definitive explanation for the overall pattern of health and COVID-19 results remains elusive.

Despite these negative effects, there were also several indications of improved household outcomes due to the HSP. Consistent with the subsidy program relaxing the household's budget constraint, midline survey data indicate that there were reductions in credit usage among treated participants (Panel E, Table 5): treated participants reported taking fewer loans before and during Jordan's COVID-19 lockdown by 20 and 34 percentage points, respectively. We largely view this outcome in a positive light considering the high debt burden in the sample: control households hold average total rental debt of 806 USD PPP at endline.

Finally, there are mixed midline estimates on child outcomes. Children in treated households

were reported to engage in 0.19 more learning activities (q-value = 0.08) but attend 0.43 fewer days of school on average, at a time when most in-person schooling was suspended due to the COVID-19 pandemic. Children in treated households also experienced a decrease in reported alertness of 0.73 standard deviations (q-value < 0.01). Beyond child outcomes, other significant effects at midline include the probability that the respondent is living with a disability that complicates self-care (0.38 SD higher difficulties with self-care, q-value < 0.08). Yet only the child alertness reductions and days of school attendance are robust to accounting for non-random attrition (Table A.2).

5.3 Impacts in the endline survey round

The endline survey was carried out in-person in 2021 as pandemic conditions eased, shortly after program assistance had ended for most households. As noted above, we find that the program significantly reduced housing expenditures by 82.05 USD PPP at endline relative to the control group (q-value=0.07). Effects on measured housing quality, on the other hand, largely dissipate by this point, with no significant differences between the treatment and control groups in any dimension of the housing quality index. Some of this could have been due to catch-up investments in the control group, or the possibility that housing improvements like window repairs and mold removal depreciate relatively quickly. Likewise the endline survey data reveals no statistically significant effects on any of the other pre-specified primary outcomes: program impacts on total household consumption, respondent depression levels, and child well-being are all negative, in fact, although relatively small in magnitude and none are significant at traditional confidence levels.

By the endline survey, most impacts observed at midline – including on food security, household composition, child learning outcomes, and COVID-19 – were no longer statistically significant. Four statistically significant effects survive multiple hypothesis testing adjustments (Table 6). First, there was a notable reduction in the amount of time respondents devoted to chores and childcare, with a decrease of 6.91 hours in these activities over the past week and corresponding

10.62-hour decrease in total time spend on labor and chores (q-value = 0.01 and q-value = 0.01, respectively), which are robust to accounting for non-random attrition. These are somewhat puzzling given the lack of meaningful changes on labor supply outside the home (which if anything appear to decline). One explanation for the change in childcare hours, explored in the endline results, is that the HSP increased women's mobility (see Table B.1), perhaps by allowing them to leave children at home due to improved home physical security (as windows and doors were repaired, for instance).

Motivated by this hypothesis, the study tested for impacts on women's mobility in the follow-up round, which shows, as discussed below, that women in treated households in fact experienced increased mobility (Table B.1).

Second, there was a decrease in self-reported respondent happiness, which fell by 0.24 points on a scale of 1-3, from an average of 1.76 (q-value = 0.02). Third, and on a more positive note, there was a significant increase in the proportion of households that reported savings of 30 Jordanian Dinars (roughly 95 USD PPP), likely due to the financial benefits of the HSP in reducing household rent payments. This appears related to the midline finding that treatment households had a lower debt burden and did not take out as many loans. Fourth, program recipients received on average 128 USD PPP in assistance from the implementing partner in the past 12 months (q-value $<$ 0.10) which likely reflects the small proportion of recipients still some receiving HSP assistance.[16]

Other results suggest that the program may have had negative impacts on labor supply and earnings but the estimated effects are less precise and not as robust to changes in measurement or regression specifications.[17]

[16] While the estimated reductions in time spent on chores and in happiness are robust to accounting for non-random attrition, the other two dimensions of effects are not, and are therefore more suggestive.

[17] The detailed results are in Table C.1.

5.4 Impacts in the follow-up survey round

The follow-up survey round was collected in 2022, roughly 1.5 years after all HSP direct rental assistance had ended. At that point the only statistically significant and robust estimated effect of the program is related to child socio-emotional well-being: the standardized Strengths and Difficulties Questionnaire (SDQ) scale decreased by 0.56 standard deviation units in treated households (q-value=0.01). This is a large magnitude and indicates that there was a meaningful decline in child socio-emotional well-being (as reported by parents), an unintended adverse consequence. There are multiple potential explanations for the worsening of child socio-emotional well-being, including short-run (midline) food insecurity, greater COVID-19 exposure, household composition impacts, and changes to respondent time use, though the persistence of the negative effect on child well-being after those effects have dissipated is perhaps surprising. A further channel relates to the nature of social interactions with neighbors, discussed in the next section. Besides that effect, across the wide range of household outcome measures gathered (including in all pre-specified primary outcomes) there were no other statistically significant results that survived a multiple hypothesis testing adjustment.

6 Impacts on Neighbor Attitudes

The 2022 survey of a representative sample of Jordanian neighbors captures the spillover impact of the refugee-targeted assistance program on Jordanians attitudes towards refugees at least one year after program implementation ended (detailed in Figure 2).

We focus first on the pre-specified primary outcomes (Table 7, Panel A). Perhaps unexpectedly, neighbors who live near treatment households experience a significant decline in the social attitude and perceptions index (effect -0.33 standard deviation units, q-value$<$0.08).

A closer examination reveals that this negative effect is primarily due to diminished social ties between Syrian refugees and their Jordanian neighbors (Table 7, Panel B): There is a notable de-

crease in the number of Syrian refugees from whom Jordanian adults seek advice, and a reduction in the number of Syrian friends among Jordanian children. This latter result could be a partial explanation for the reduced well-being of treated Syrian children. A leading explanation for this pattern is that there was a backlash effect among Jordanians living in the working-class neighborhoods that Syrian refugees tend to inhabit, driven perhaps by resentment at the assistance received through HSP.

We find no evidence that the program increased Jordanians' housing costs or had other adverse economic effects on Jordanian neighbors. There are no significant impacts on Jordanian neighbors' total consumption nor housing expenditures (see Appendix).

Regarding the other pre-specified primary outcomes, there are positive but small estimated program impacts on neighbors' policy preferences and altruism towards Syrian refugees but these are not statistically significant.

There are several other notable patterns among neighbors (Table 7, Panel C). First, neighbors of treatment-assigned refugee households report significantly worse subjective assessments of their own well-being, specifically, a 0.30 standard deviation unit decrease in their self-assessed health (q-value=0.06), and a 0.32 standard deviation unit decrease in life satisfaction (q-value=0.06). There are several potential drivers of these impacts, including the possibility that observing refugee neighbors receive generous assistance led them to feel worse about their own living situation and quality of life, which is perhaps related to the resentment and backlash effects hypothesized above.

Second, neighbors of treated households state that they believe refugees typically receive significantly *less* aid than that stated by control group neighbors (effect -347 USD PPP, q-value$<$0.01). Interestingly, the neighbors of treated households have more accurate beliefs about aid levels, with fewer of them believing that refugees receive extremely high levels of aid that are rarely observed. There thus appears to have been some learning of objectively true information regarding the reality of refugees' assistance levels among their Jordanian neighbors. This is consistent with greater information exchange and learning as a result of the highly observable HSP.

Third, neighbors of treated refugee households claim to have -1.08 fewer days of media consumption per week (q-value<0.08), relative to a control mean of (4.78). It is difficult to explain this effect on media consumption but one hypothesis is it reflects a decrease in neighborly socializing, which sometimes resolves around watching or listening to the news on the television or radio; but we are unable to offer more decisive survey-based evidence on the underlying mechanisms.

There was no differential attrition when surveying neighbors of treatment and control communities, and so most of the significant results discussed are robust to bounding for non-random attrition; the only exceptions are the effects on life satisfaction and subjective health, which should therefore be viewed as more suggestive.

Heterogeneity analysis lends insight into potential mechanisms underlying these effects (Table B.2). First, variation in physical distance to the treated refugee household allows us to bolster the premise that the worsening in social attitudes is related to direct exposure to HSP and its recipients: the decrease in the social attitudes index is concentrated among those neighbors who live physically closer to the refugee households (based on GPS-measured distance). This suggests that more direct and frequent observation of treated refugees' circumstances led to more negative effects.

A second informative dimension of heterogeneity is whether the respondent has non-Jordanian grandparents. In this setting, 17 percent of the neighbor sample has a grandparent (or spouse's grandparent) born outside of Jordan, with 72 percent of those born in Palestine. Notably, the large negative program impact on the neighbor social attitudes and perceptions index is driven by those with Jordanian-born grandparents, while those with non-Jordanian born grandparents have an effect close to zero (Table B.2). While the results are noisy due to the relatively small share of neighbors with non-Jordanian born grandparents, this finding suggests that a family history of immigration or displacement — for instance, which is common among those of Palestinian descent — may enhance openness to Syrian refugees.[18]

The pre-analysis plan additionally specified heterogeneity analysis by neighbor gender, age,

[18]This finding is in contrast to the correlational evidence in Ghosn et al. (2019) that a history of displacement did not improve openness towards refugees.

education and socioeconomic status (see Appendix). The final aspect of pre-specified heterogeneity was social desirability, measured using a normalized continuous score from the widely used Marlowe-Crowne scale (Crowne and Marlowe, 1960). The large negative program impact on the neighbor social attitude and perceptions index are predominantly driven by individuals with lower scores on the social desirability scale, indicating that they are less likely to suffer from experimenter demand effects. In our view, this lends additional credibility to the negative social cohesion result, and also implies that the study may underestimate the extent of the negative effect on social attitudes, if Jordanian neighbors with higher social desirability tendencies do not truthfully report their possibly even more negative views of Syrian refugees.

7 Comparing Forecasts to Estimated Program Impacts

Comparing experts' forecasts to estimated the program effects allows us to highlight areas where the study results advanced learning relative to prior expectations. The forecasts focus on the midline and endline measures of the primary outcomes, namely: i) the well-being measures among refugees, and ii) the attitudes and perceptions of Jordanian neighbors towards refugees.

In terms of impacts among refugees, the findings reveal that the actual impacts of the intervention on refugee well-being were generally smaller than the modest positive impacts that experts had anticipated, with the exception of housing quality (Figure 3, Panel A). Notably, statistically significant discrepancies (at over 95% confidence) between the estimated and predicted effects were observed in two areas: housing expenditures and child socio-emotional health. Housing expenditures in the endline survey were notably lower for treatment households compared to control, while forecasts anticipated no such effect. One limitation of the predictions we collected, however, was the fact that we failed to inform the forecasters that the program had mandated that landlords had to maintain stable rent levels for one year following the end of subsidy distribution, and this appears to be the main driver of the non-effect on forecasted housing spending. Thus, some caution is needed in interpreting these effects.

Experts also forecasted a marginally significant improvement in child socio-emotional well-being at endline, yet if anything there are reductions in this outcome at endline though they are not significant; as noted above, there are significant reductions in the child well-being measure at the 1.5-year follow-up but forecasts were not collected for that round. There is thus a meaningful and significant (at 95% confidence) difference between the predicted and actual effects on the child outcome, indicating that the study generated new insights that diverge from the experts' priors.

There is also a meaningful and significant gap between experts' forecasts and the observed effects of the program on neighbors' social attitudes (Figure 3, Panel B). Experts had anticipated an average null effect of the program on Jordanian neighbors across all three primary pre-specified outcomes (namely, policy support, economic perceptions, and social attitudes), and for the first two the predictions and actual estimates are all close to zero. The program's significant negative impact on neighbors' social attitudes is significantly different from the predicted null effect at 95% confidence, once again a meaningful update relative to priors.

8 Conclusion

This paper investigates the short- and medium-term effects of a substantial housing subsidy program on Syrian refugee households and their Jordanian neighbors. The program provided on average one year of free rent, along with improvements in shelter quality and continued rent stability even after the subsidies ended. The study examines the impact on the living standards and well-being of Syrian refugees, as well as the social attitudes and interactions of their Jordanian neighbors. The study's novel design, which restricts the use of subsidies to existing rental relationships, largely prevents the migration responses typically associated with rental subsidies. Moreover, the saturation design allows the study to test for, and rule out, local housing price changes.

The main analysis documents largely null effects on the primary pre-specified living standards measures, including household consumption and respondent well-being. This is perhaps surprising

given that both the humanitarian implementing organization and expert forecasters including the research team, believed that the program would generate moderate positive impacts.

Two striking and unexpected results are the significant deterioration in measured child socio-emotional well-being following the program and the increased strain in relations between neighbors and refugees. In terms of the child outcomes, there are multiple potential channels that could explain the results, including the finding that food security worsened among treated households (as they appear to have been partially cut off from formal or informal food aid), that treated households experienced higher incidence of COVID-19, and treated households' changes in living arrangements (at least in the short run, such as the increase in adolescents).

The negative child well-being and social cohesion outcomes may in fact be interrelated: there are decreased interactions between treated refugee children and Jordanian children. This deterioration in refugee-host community relations may have also jeopardized the informal support in terms of food or other forms of assistance that some refugees received from their Jordanian neighbors.

Taken together, the results indicate that a meaningful housing subsidy did not lead to transformative positive changes for recipient households. The findings thus offer a word of caution when designing assistance programs in settings with strong social ties and the potential for both formal and informal redistribution of assistance. More broadly, the lack of sustained impacts for treatment households may reflect the numerous other constraints that refugees face in gaining access to livelihood opportunities, credit, and quality housing. The deterioration in social cohesion offers novel evidence that assistance targeted exclusively to refugees can prompt host community backlash.

We view these as unexpectedly discouraging results of the housing subsidy program, and they are less optimistic than predicted by experts (or by the research team). Although different results might hold in populations other than Syrian refugees in Jordan, on some level this setting would seem to be almost a global "best-case" scenario for the potential success of such a program given the high degree of linguistic, cultural and religious similarity between the Syrian refugees and their Jordanian hosts; Jordan's status as an upper middle-income country; and the fact that the program

was launched roughly a decade after displacement had begun from the Syrian Civil War, allowing ample time for construction of additional housing units and social integration.

The results also speak to the active ongoing policy debate on how best to support refugees and host states at the same time (Ash and Huang 2018; Baseler et al. 2023). When study participants were asked directly about what form of assistance they preferred (during the 1.5 year follow-up survey), respondents overwhelmingly stated that they preferred direct cash transfers to landlord subsidies (at 70%). Cash transfers to refugees in other contexts have been shown to provide only short-run benefits rather sustained gains (Hidrobo et al. 2014; Özler et al. 2021; Quattrochi et al. 2022; Moussa et al. 2022; Aygün et al. 2024), but the findings of this study suggest that delivering benefits more discreetly (such as via cash or mobile money) may at a minimum reduce the risk of host community backlash. An alternative approach, supported by Baseler et al. 2023, would explicitly pair refugee assistance with enhanced host community assistance, which in this case might have led to some housing investments among host community neighbors as well. There is promising evidence from Colombia that labor market reforms may be an effective way to improve refugee living standards without leading to crowd-out of host community jobs (Bahar et al. 2021; Rozo et al. 2023; Ibáñez et al. 2024). These remain critical areas for future research and policy innovation, and present important opportunities to improve program design and ultimately refugees' lives.

References

Agness, D. (2023). Housing and human capital: Condominiums in ethiopia. *3ie Registry*.

Anderson, M. L. (2012). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of American Statistical Association*.

Ash, N. and C. Huang (2018). Using the compact model to support host states and refugee self-reliance. *World Refugee Council Research Paper Series*.

Aygün, A. H., M. G. Kırdar, M. Koyuncu, and Q. Stoeffler (2024). Keeping refugee children in

school and out of work: Evidence from the world's largest humanitarian cash transfer program. *Journal of Development Economics 168*, 103266.

Bahar, D., A. M. Ibáñez, and S. V. Rozo (2021). Give me your tired and your poor: Impact of a large-scale amnesty program for undocumented refugees. *Journal of Development Economics 151*, 102652.

Baird, S., J. De Hoop, and B. Özler (2013). Income shocks and adolescent mental health. *Journal of Human Resources 48*(2), 370–403.

Baseler, T., T. Ginn, R. Hakiza, H. Ogude-Chambert, and O. Woldemikael (2023). Can redistribution change policy views?: Aid and attitudes toward refugees in uganda. *Center for Global Development, Working Paper 645*.

Beltramo, T., F. Nimoh, and O. Matthew (2024). Financial security, climate shocks and social cohesion. *Center for Economic Policy Research, Working Paper*.

Burke, M., S. M. Hsiang, and E. Edward Miguel (2015). Global non-linear effect of temperature on economic production. *Nature*.

Chetty, R., N. Hendren, and L. F. Katz (2016, April). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review 106*(4), 855–902.

Crowne, D. P. and D. Marlowe (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology 24*(4).

DellaVigna, S., N. Otis, and E. Vivalt (2020). Forecasting the results of experiments: Piloting an elicitation strategy. *National Bureau of Economic Research*.

DellaVigna, S. and D. Pope (2017). What motivates effort? evidence and expert forecasts. *The Review of Economic Studies*.

Egger, D., J. Haushofer, E. Miguel, P. Niehaus, and M. Walker (2022). General equilibrium effects of cash transfers: experimental evidence from kenya. *Econometrica 90*(6), 2603–2643.

Ghosn, F., A. Braithwaite, and T. S. Chu (2019). Violence, displacement, contact, and attitudes toward hosting refugees. *Journal of Peace Research 56*(1), 118–133.

Hagen-Zanker, J., M. Ulrichs, and R. Holmes (2018). What are the effects of cash transfers for refugees in the context of protracted displacement? findings from jordan. *International Social Security Review 71*(2), 57–77.

Haushofer, J., J. Reisinger, and J. Shapiro (2015). Your gain is my pain: Negative psychological externalities of cash transfers. *Working Paper, retrieved on May 13*, 2016.

Haushofer, J. and J. Shapiro (2016). The short-term impact of unconditional cash transfers to the poor: experimental evidence from kenya. *The Quarterly Journal of Economics 131*(4), 1973–2042.

Haushofer, J. and J. Shapiro (2018). The long-term impact of unconditional cash transfers: experimental evidence from kenya. *Busara Center for Behavioral Economics Working Paper, Nairobi, Kenya*.

Hidrobo, M., J. Hoddinott, A. Peterman, A. Margolies, and V. Moreira (2014). Cash, food, or vouchers? evidence from a randomized experiment in northern ecuador. *Journal of Development Economics 107*, 144–156.

Ibáñez, A., A. Moya, M. A. Ortega, S. Rozo, and M. J. Urbina Florez (2024). Life out of the shadows: The impacts of regularization program on the lives of forced migrants. *Journal of European Economic Association*.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Kumar, T. (2021). Home-price subsidies increase local-level political participation in urban india. *Forthcoming, Journal of Politics*.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies 76*(3), 1071–1102.

Lehmann, C. and D. Masterson (2014). Emergency Economies: the Impact of Cash Assistance in Lebanon. Technical report, International Rescue Committee.

Ludwig, J., G. J. Duncan, L. A. Gennetian, L. F. Katz, R. C. Kessler, J. R. Kling, and L. Sanbonmatsu (2013, May). Long-term neighborhood effects on low-income families: Evidence from moving to opportunity. *American Economic Review 103*(3), 226–31.

Moussa, W., N. Salti, A. Irani, R. Al Mokdad, Z. Jamaluddine, J. Chaaban, and H. Ghattas (2022). The impact of cash transfers on syrian refugee children in lebanon. *World Development 150*, 105711.

NRC (2015). In search of a home: Access to adequate housing in Jordan. Technical report, Norwegian Refugee Council.

Özler, B., Ç. Çelik, S. Cunningham, P. F. Cuevas, and L. Parisotto (2021). Children on the move: Progressive redistribution of humanitarian cash transfers among refugees. *Journal of Development Economics 153*, 102733.

Park, S.-H. and H. Y. Yu (2021). How useful is the center for epidemiologic studies depression scale in screening for depression in adults? an updated systematic review and meta-analysis. *Psychiatry Research 302*, 114037.

Quattrochi, J., G. Bisimwa, P. Van Der Windt, and M. Voors (2022). Cash-like vouchers improve psychological well-being of vulnerable and displaced persons fleeing armed conflict. *PNAS Nexus 1*(3), pgac101.

Rozo, S. V., A. Quintana, and M. J. Urbina (2023). *The Electoral Consequences of Easing the Integration of Forced Migrants*.

Tauchmann, H. (2014). Lee (2009) treatment-effect bounds for nonrandom sample selection. *The Stata Journal 14*(4), 884–894.

UNHCR (2023a). Jordan. Technical report, United Nations High Commissioner for Refugees.

UNHCR (2023b). Syria. Technical report, United Nations High Commissioner for Refugees.

UNHCR (2023c). Unhcr global trends. Technical report, United Nations High Commisioner for Refugees.

Woerner, W., B. Fleitlich-Bilyk, R. Martinussen, J. Fletcher, G. Cucchiaro, P. Dalgalarrondo, M. Lui, and R. Tannock (2004). The strengths and difficulties questionnaire overseas: evaluations and applications of the sdq beyond europe. *European child & adolescent psychiatry 13*, ii47–ii54.

Zhou, Y.-Y. (2019). How refugee resentment shapes national identity and citizen participation in africa. In *APSA 2018 Annual Meeting Paper*.

9 Tables and Figures

Figure (1) Geographic Location of Treated and Control Communities

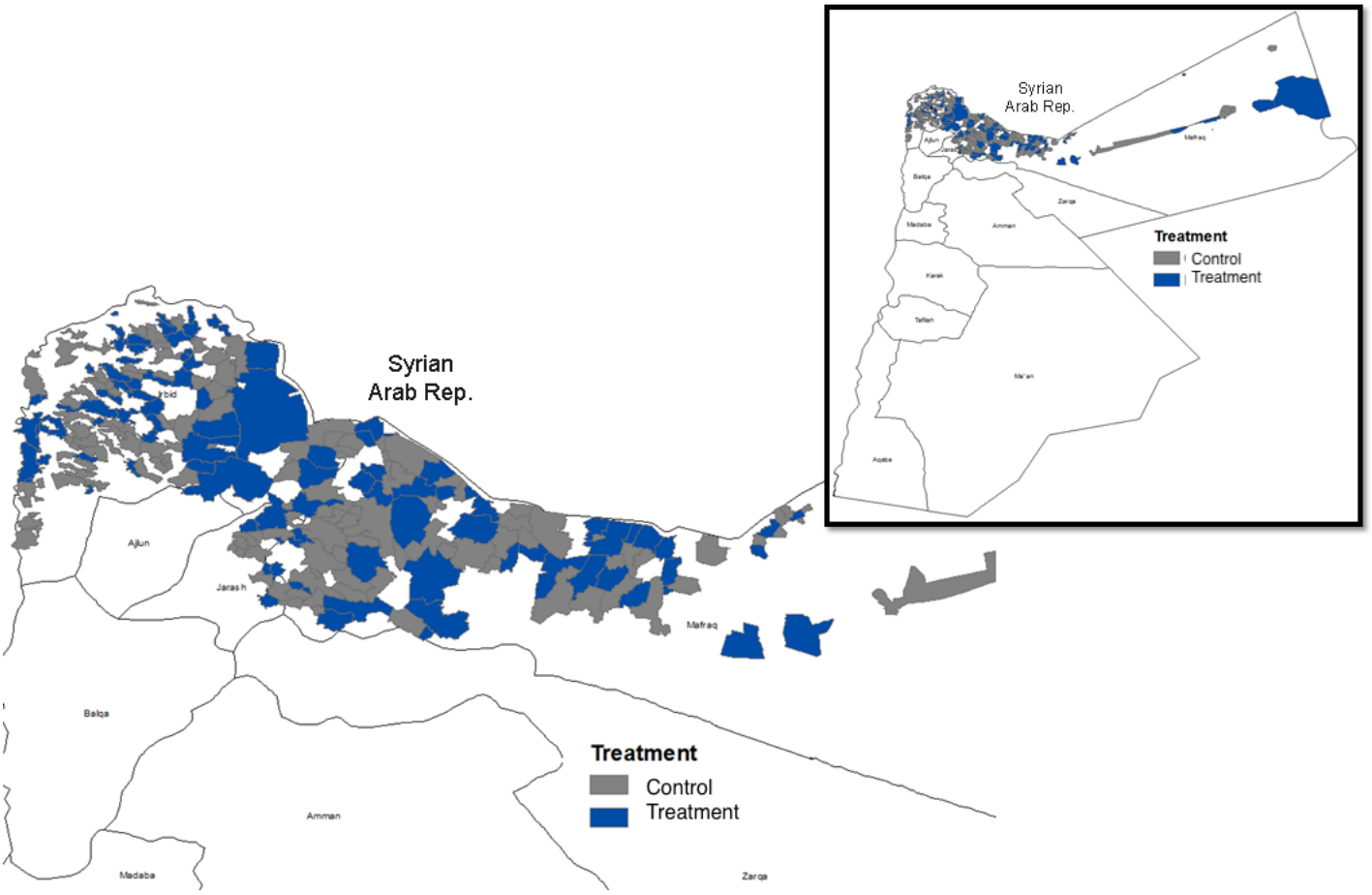


Figure (2) RCT, Data Collection, and Forecasts Timeline

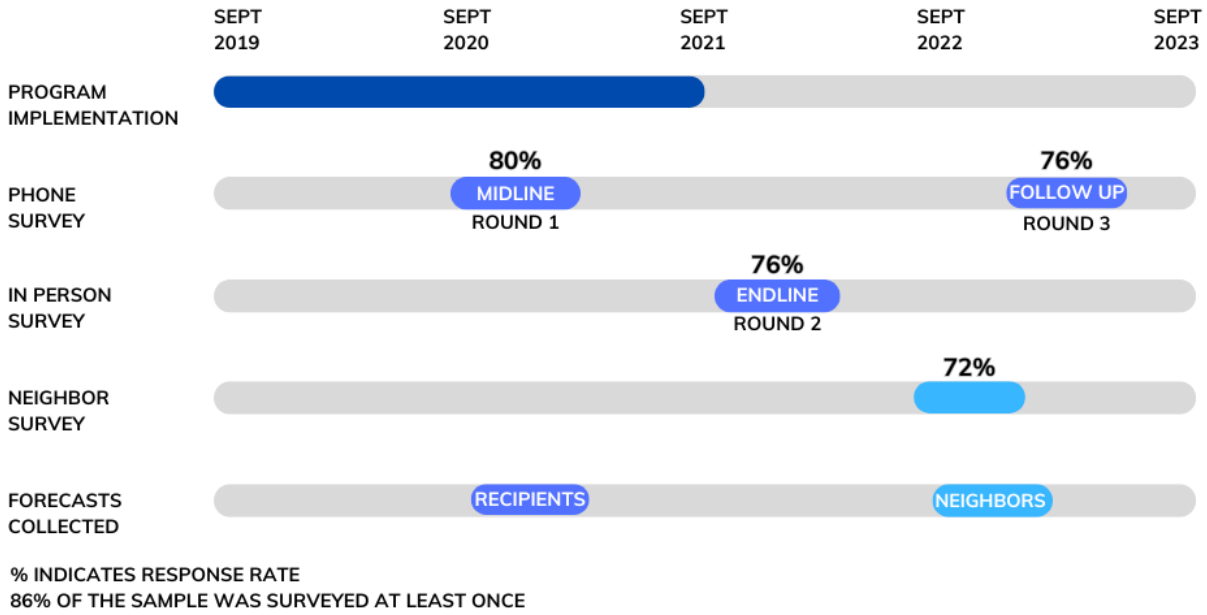


Table (1) Average Respondent, Household, and Shelter Characteristics

		Treatment	Control	Difference	(se)	N
Panel A: Respondent characteristics						
(1)	Female respondent (=1)	0.46	0.52	-0.05**	(0.03)	1,619
(2)	Age (categorical)	33.99	34.19	-0.21	(0.61)	1,619
(3)	Married (=1)	0.84	0.84	0.00	(0.02)	1,616
(4)	Disabled (Washington Group, =1)	0.26	0.25	0.02	(0.03)	1,616
Panel B: Household characteristics						
(5)	Dependency Ratio	1.31	1.27	0.04	(0.07)	1,616
(6)	Household size	5.20	5.14	0.06	(0.15)	1,616
(7)	Number of families in the same house	1.30	1.30	-0.00	(0.04)	1,616
(8)	Number of children	2.96	2.89	0.07	(0.12)	1,619
Panel C: Shelter characteristics						
(9)	Access to piped water (=1)	0.66	0.67	-0.01	(0.07)	1,616
(10)	Fully constructed roof (=1)	0.12	0.08	0.04**	(0.02)	1,616
(11)	Functional windows (=1)	0.22	0.22	-0.01	(0.03)	1,616
(12)	Completed floor (=1)	0.44	0.40	0.04	(0.06)	1,616
(13)	Toilet (=1)	0.93	0.92	0.01	(0.04)	1,616
(14)	Plan to stay in shelter (=1)	0.92	0.92	0.00	(0.02)	1,537
(15)	Monthly rent (USD PPP)	328.02	335.28	-7.26	(17.57)	1,431
(16)	Lease contract (=1)	0.77	0.78	-0.01	(0.04)	1,431
(17)	Number of times moved shelter	0.47	0.44	0.03	(0.08)	1,616
(18)	Permanent shelter (=1)	0.91	0.86	0.05	(0.06)	1,616
	Joint significance p-val: 0.102					

Notes: This table reports treatment balance results for the 1,619 household interviewed at midline from the 2,017 households sampled for the study (see Table 3). We have admin data from the implementing partner's integrated assessment but only for the 1,619 households that we were able to reach at midline. The first two columns report averages by group, while the third column reports results of estimating $y_i = \beta_0 + \beta_1 T_i + \epsilon_i$. Robust standard errors in parentheses in the forth column and are clustered at the locality level. The last column reports the total sample size. *** p<0.01, ** p<0.05, * p<0.1. All outcomes are taken from the implementing organization's data, except for gender and age, which are taken from the phone survey. "USD" denotes United States Dollars and "PPP" stands for Purchasing Power Parity, which is used to compare the absolute purchasing power of countries' currencies. Rent is winsorized at the top 1% of values, in order to limit the influence of outliers. Disability is calculated using the Washington Group Short Set on Disability and uses the recommended cut-off (level 3) to classify respondents as being disabled or not based on six domains: seeing, hearing, walking, cognition, self-care, and communication. For household size, we drop one observation in which the number of individuals at baseline is reported to be 300. For times changed shelter, we drop one observation in which the respondent is reported to have changed shelter 48 times.

Table (2) Forecasts of Treatment Effects

Predicted Treatment Impact (SD)	Mean	SE of Mean	p10	p90	N
Panel A: Refugee Impacts					
Housing Expenditures (Midline)	-0.28	0.07	-0.50	0.30	61
Housing Expenditures (Endline)	-0.01	0.03	-0.20	0.25	61
Housing Quality (Midline)	0.28	0.02	0.10	0.50	61
Housing Quality (Endline)	0.19	0.02	0.00	0.40	61
Household Consumption (Endline)	0.16	0.02	0.00	0.40	61
Adult Mental Health (Midline)	0.22	0.02	0.00	0.40	61
Adult Mental Health (Endline)	0.14	0.02	0.00	0.30	61
Child Socio-Emotional (Endline)	0.16	0.03	0.00	0.30	61
Panel B: Neighbor Impacts					
Social Attitudes	0.01	0.03	-0.29	0.30	63
Economic Perception	0.01	0.03	-0.25	0.30	63
Policy Support	0.02	0.03	-0.22	0.32	63

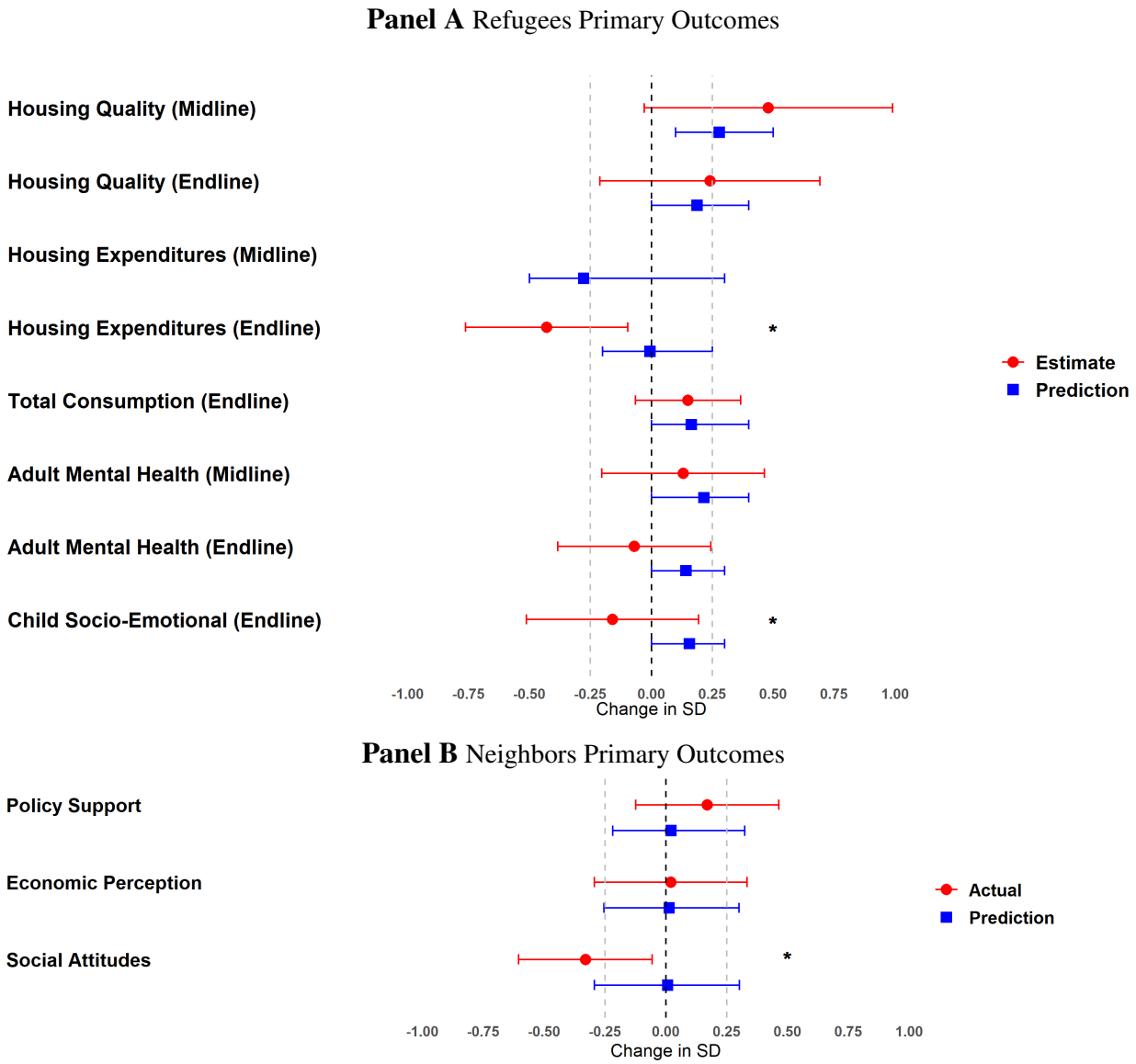
Notes: This table reports summary statistics of the distribution of predictions of the experimental treatment effects. Treatment effect predictions were elicited in terms of standard deviations. p10 and p90 refer to the 10th and 90th percentile of the distribution of predictions. We report here only observations who made predictions for all outcomes.

Table (3) Panel Retention and Compliance

	Ever found	Midline	Endline	Follow-up
Panel A: Refugee sample retention				
All	86.4%	80.3%	76.1%	75.7%
Treatment	90.5%	83.3%	80.4%	80.1%
Control	82.4%	77.3%	71.9%	71.5%
p-value (T-C)	0.000	0.014	0.000	0.000
N	2,017	2,017	2,017	2,017
Panel B: Treatment take-up among the surveyed				
All	17.8%	16.8%	17.6%	17.9%
Treatment	32.8%	33.0%	33.5%	34.0%
Control	0.0%	0.0%	0.0%	0.0%
p-value (T-C)	0.000	0.000	0.000	0.000
N	1,729	1,619	1,466	1,460
Panel C: Neighbors sample retention				
All	72.1%	–	–	–
Treatment	74.7%	–	–	–
Control	69.5%	–	–	–
p-value (T-C)	0.421	–	–	–
N	2,017	–	–	–

Notes: This table shows panel retention and compliance with treatment assignment. The p-values are obtained by estimating $y_{ic} = \beta_0 + \beta_1 \text{Treatment}_c + \varepsilon_{ic}$. Standard errors are clustered at the locality-level. Retention is defined as survey retention, meaning that the participant was located and surveyed in that round. These numbers only include people who were successfully surveyed. Panel A shows retention of the 2,017 households that were sampled for the study. Panel B shows treatment take-up for the sub-sample of households that were interviewed at midline (1,619), endline (1,466), and follow-up (1,460), and Panel C shows retention for the neighbors of the 2,017 refugees sampled. In this case, retention is an indicator for whether a Jordanian neighbor was located near the pre-randomization home location of the refugee household. The neighbor was selected using a randomized algorithm starting from the original home of the refugee participant, with a median distance between the selected neighbor and the original refugee participant's home of 63 meters. Cases of non-retention could result from an inability to identify the refugees' home location based on midline data, or the inability to locate Jordanian neighbors during field visits to the location.

Figure (3) Program Impacts on Primary Outcomes - Estimates vs. Predictions



Notes for Panel A: 10-90th Percentile of Predictions (n = 61), 95% CI of Actual Outcomes, 4-12 months (n = 1610), 12-18 months (n $\sim$ 1395), Housing Expenditures (Midline) was not measured in a comparable fashion (see text for details) and is not shown here, * Denotes Estimate = Prediction rejected at p<0.05.
Notes for Panel B: 10-90th Percentile of Predictions (n = 63), 95% CI of Actual Outcomes (n = 1102), * Denotes Estimate = Prediction rejected at p<0.05.

Table (4) Pooled Treatment Effects on Primary Outcomes

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N	Rounds
Panel A: Primary Outcomes							
Overall Housing Quality (Z-Score)	0.31	(0.20)	[0.19]	-0.01	(1.00)	4,313	1, 2, 3
Total Monthly Housing Expenditures (USD PPP)	-82.05***	(32.07)	[0.07]	221.63	(202.26)	1,421	2
Food Consumption (Log USD PPP)	-0.02	(0.07)	[0.71]	4.39	(0.72)	2,780	2, 3
Log Total Consumption (Log USD PPP)	-0.03	(0.08)	[0.68]	8.73	(0.55)	1,422	2
CESD Score (Higher: Less Depression)	-0.10	(0.13)	[0.52]	0.01	(1.00)	4,302	1, 2, 3
SDQ Score (Higher: Better Child Wellbelling)	-0.34**	(0.15)	[0.07]	-0.01	(1.00)	1,782	2, 3
Panel B: Other Outcomes							
Respondent Hunger Last Week (=1)	0.18***	(0.05)	[0.01]	0.39	(0.49)	4,261	1, 2, 3
Adult Hunger Last Week (=1)	0.13***	(0.05)	[0.03]	0.36	(0.48)	4,177	1, 2, 3
Child Hunger Last Week (=1)	0.15***	(0.05)	[0.01]	0.25	(0.43)	3,813	1, 2, 3
At least 30 JD (95 USD PPP) in savings (=1)	0.08**	(0.04)	[0.05]	0.10	(0.30)	2,797	2, 3

Notes: This table reports 5 primary outcomes plus per capita food consumption (not pre-specified), pooled across as many rounds as were collected for each outcome. The table shows that treatment had no impact on housing quality on average over all 3 rounds, reduced housing expenditures at endline, had no impact on food consumption over endline and followup, had no impact on depression over all 3 rounds, and reduced child socio-emotional wellbeing over rounds 2 and 3. Overall Housing Quality is defined as a normalized housing quality index that includes indicators for quality floors, roofs, and walls, indicators for access to grid electricity and piped water, and the number of people per room. Total monthly housing expenditures is only reported for endline in the pooled table due to variations in measurement of housing expenditures across rounds and due to when treatment ended. Housing expenditure was measured including assistance payments at midline, and excluding assistance payments at endline since most recipients had stopped receiving payments but were still subject to rent freezes at part of the program. Results on housing expenditures in each round are reported in the appendix. Total consumption, including food and non-food expenses, was only measured at endline due to the length of the module. Reported in this table is food consumption, which was measured at endline and followup. Total food consumption is the log of the sum of food consumed in last seven days (Cereals and cereal products, Live animals, meat, and other parts of slaughtered land animals; Fish and other seafood; Milk, other dairy products, and eggs; Oils and fats; Fruits and nuts; Vegetables, tubers, pulses; Sugar and desserts; Ready-made food and other food products (baby food, spices)), plus home-produced foods produced in the last 7 days, plus assistance received in the last 30 days divided by 4.33. CES-D was measured in all rounds with exactly the same questionnaire. It includes 10 questions on depressive symptoms. Individuals who skipped more than 2 questions were marked as missing. The continuous scores were standardized with respect to the control group. The Child Strengths and Difficulties questionnaire (SDQ) was administered regarding a randomly selected child aged 3-8 at endline. At follow-up, the SDQ was administered regarding the same child. If not surveyed previously or if that child is no longer in the household, the survey was administered with respect to a randomly selected child age 4-9. The continuous scores were randomized with respect to the control group. At least 30 JD (95 USD PPP) in savings (=1) is an indicator equal to one if the respondent answered yes to "Do you currently have at least 30 JD (95 USD PPP) in personal savings you can draw from in an emergency? Whether or not it is in a bank?". Respondent, adult, and child hunger last week are indicators equal to 1 if the respondent, other adults in the household or children in the household (respectively) went to bed hungry at least once in the last 7 days, zero otherwise. In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to various families in the pre-analysis plan. Because not all outcomes are available in multiple rounds, some families include fewer outcomes in the pooled estimates than the round-by-round estimates.

Table (5) Statistically Significant Impacts of the Program at Midline

Outcome	TOT	(se)	FDR q-values	Control mean	(sd)	N
Panel A: Housing Improvements						
Quality Roof (=1)	0.24***	(0.09)	[0.08]	0.71	(0.45)	1,610
Clean water (=1)	0.21***	(0.06)	[0.01]	0.17	(0.38)	1,610
Occupied Rooms	0.51***	(0.17)	[0.04]	2.85	(1.05)	1,610
Panel B: Food Aid and Insecurity						
Food consumption (aid) USD PPP	-53.05**	(22.40)	[0.02]	218.51	(201.68)	1,569
Food consumption (aid) USD PPP (*Per capita*)	-9.73***	(3.44)	[0.01]	35.88	(29.19)	1,569
Respondent hunger last week (=1)	0.23***	(0.07)	[0.01]	0.35	(0.48)	1,604
Adults hunger last week (=1)	0.17***	(0.06)	[0.01]	0.38	(0.49)	1,610
Child hunger last week (=1)	0.24***	(0.07)	[0.00]	0.23	(0.42)	1,469
Panel C: Household Recomposition						
Number of children under 18 years	0.30**	(0.15)	[0.03]	3.27	(1.97)	1,610
Number of boys aged 13-17 years in household	0.16**	(0.07)	[0.33]	0.40	(0.66)	1,610
Panel D: COVID-19 Outcomes						
Weekly adult income (during first lockdown)	-20.07***	(7.19)	[0.02]	30.22	(66.86)	1,609
Total COVID-19 symptoms	0.84***	(0.19)	[0.00]	0.47	(1.22)	1,610
Number of people who are symptomatic	0.46***	(0.09)	[0.00]	0.29	(0.61)	1,581
Number of children who are symptomatic	0.17***	(0.05)	[0.01]	0.06	(0.32)	1,472
Number of visits to healthcare institutions	-0.56***	(0.23)	[0.04]	1.14	(2.03)	1,608
Did not keep distance (=1)	0.16***	(0.06)	[0.08]	0.51	(0.50)	1,609
Panel E: Loans						
Loans taken pre-first-lockdown (=1)	-0.20***	(0.07)	[0.01]	0.61	(0.49)	1,608
Loans taken during first lockdown (=1)	-0.34***	(0.07)	[0.00]	0.78	(0.41)	1,608
Loans taken after first lockdown (=1)	-0.33***	(0.07)	[0.00]	0.42	(0.49)	1,607
Panel F: Other effects						
Total number of school days attended per child	-0.43*	(0.24)	[0.04]	2.33	(2.49)	1,031
Child alertness (s.d. units)	-0.73***	(0.20)	[0.00]	0.00	(1.00)	975
Difficulty with self-care (z-score)	0.38***	(0.14)	[0.08]	-0.00	(1.00)	1,610
Number of learning activities	0.19**	(0.09)	[0.08]	0.22	(0.56)	1,385

Notes: The table shows the regression results on statistically significant results from the midline survey. Each row is its own dependent variable. **Panel A** outcomes are Clean Water, defined as an indicator for households having treated drinking water (such as by a filter). Rooms and rooms occupied by family members are reported by the respondents. **Panel B** outcomes are food consumption, Respondent (and adult) food insecurity equals 1 if the respondent went to bed hungry on at least one day in the last week and 0 otherwise. **Panel C** outcomes are COVID symptoms, which is the sum of indicators that increase the likelihood of COVID contraction by the respondent including leaving the house, attending social gatherings, not keeping distance from others, going to mosque or other religious institutions, going to grocery store or market, and leaving village/neighborhood. The other outcomes are as explained in the labels. **Panel D** The periods of the loans referred to are before lockdown: (January 15 - March 15, 2020); during lockdown: (March 15 - May 15, 2020); and after lockdown: (May 15 - interview date). **Panel E** outcomes are self-explanatory. **Panel F** Difficulty with self-scare with higher values indicating more disability. This is an item of the The Washington Group Short Set on Functioning; learning activities are done in the last 24 hours and include homework, e-learning, educational programs and videos, and reading.
Monetary values are in USD PPP and are winsorized at the top 1% of values in order to limit the influence of outliers. Health visits is also winsorized at the top 1% due to outliers in that measure. The regressions also have month-by-year fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to various families in the pre-analysis plan.

Table (6) Statistically Significant Impacts of the Program at Endline and Follow-up

Outcome	TOT	(se)	FDR q-values	Control mean	(sd)	N
Panel A: Endline						
Focus Respondent Childcare & Chores (Hours)	-6.91***	(2.02)	[0.01]	17.76	(18.39)	1,418
Total Labor and Chores Hours Last Week	-10.62***	(3.24)	[0.01]	25.54	(23.39)	1,422
General Happiness (Scale 1-3)	-0.22***	(0.08)	[0.02]	1.76	(0.61)	1,422
At least 30 JD (95 USD PPP) in savings (=1)	0.14***	(0.05)	[0.02]	0.12	(0.32)	1,422
Applied but Did Not Take Loan (=1)	0.06***	(0.02)	[0.00]	0.01	(0.10)	1,422
IP Assistance, Last 12 Months (=1)	0.08**	(0.03)	[0.14]	0.04	(0.19)	1,422
IP Assistance, Last 12 Months (USD PPP)	128.25***	(46.50)	[0.10]	28.74	(198.87)	1,422
Panel B: Follow-up						
Child SDQ score, std	-0.56***	(0.17)	[0.01]	-0.01	(1.00)	859

Notes: The table shows the regression results on statistically significant results from the endline survey. Each row is its own dependent variable. IP refers to the implementing partner. Monetary values are in USD PPP and are winsorized at the top 1% of values in order to limit the influence of outliers. The regressions also have month-by-year fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to various families in the pre-analysis plan. Focus Respondent Childcare & Chores (Hours) was pre-specified in two outcome families; the table reports the lower of the two q-values; the higher q-value is 0.02 .

Table (7) Impacts of the Program on Neighbor's Attitudes Toward Refugees

Outcome	TOT	(se)	FDR q-values	Control mean	(sd)	N
Panel A: Primary Outcomes						
Social attitudes & perceptions (SD)	-0.33**	(0.14)	[0.08]	-0.00	(1.00)	1,102
Economic attitudes & perceptions (SD)	0.02	(0.16)	[0.56]	0.02	(1.02)	1,102
Policy preferences (SD)	0.17	(0.15)	[0.37]	-0.02	(1.00)	1,102
Altruism to Syrians	0.21	(0.18)	[0.37]	0.88	(1.18)	1,102
Panel B: Selected Social Attitudes Index Components						
Of the 3 people you exchange advice with, how many are Syrian refugees?	-0.24**	(0.11)	[0.32]	0.34	(0.72)	1,102
Do the children in this household have any Syrian refugee friends?	-0.19*	(0.11)	[0.37]	0.40	(0.49)	704
Panel C: Other Related Outcomes						
Days of media consumption (last week)	-1.08***	(0.40)	[0.08]	4.78	(2.78)	1,102
Neighbor perceptions of average refugee aid receipt (PPP)	-347.17***	(109.69)	[0.04]	630.84	(657.41)	685
Life Satisfaction (SD)	-0.32**	(0.15)	[0.06]	0.01	(0.98)	1,097
Subjective Health (SD)	-0.30*	(0.16)	[0.06]	0.02	(0.99)	1,102

Notes: This table reports impacts of treatment on Jordanian neighbors of Syrian refugees in the experimental sample. **Panel A** reports pre-specified treatment effects on primary neighbor outcomes. All indices are standardized so that a positive point estimate reflects more "pro-refugee" sentiments. The only statistically significant outcome is a negative treatment effect on social attitudes and perceptions. This outcome is a standardized index of the following outcomes: out of 3 closest friends, how many are Syrian refugees; out of 3 people you exchange advice with, how many are Syrian refugees; do the children in this household have Syrian refugee friends (=1); do the children in this household share recreational spaces with Syrian refugee children (=1); how comfortable would you be accepting the marriage of a friend or loved one to a Syrian refugee (1-5); how comfortable would you be accepting a Syrian refugee as a neighbor (1-5); what is the net effect of Syrian refugees on Jordanian society (1-3); are Syrian refugees hardworking or lazy (1-7). Economic attitudes and perceptions is a standardized index of the following outcomes: an indicator =1 if the respondent listed "hosting Syrian refugees" as one of the most important challenges facing Jordan (options not read); an outcome with higher values for responses associated with beliefs that Syrian refugees pay more in taxes than Jordanians; an outcome with higher values for responses associated with refugees having a positive effect on the economy.Policy preferences in a standardized index of the following outcomes: less belief that Syrian refugees should be forced to live in camps; more belief that refugees should have the right to work outside camps; more belief that refugees should be able to attain Jordanian citizenship; more support for unrestricted work permits for refugees; more support for integrated classrooms with Syrians and Jordanians; more support for refugee right to enter/exit camps freely; support for housing assistance for Syrian refugees; belief that the international community should spend more money supporting refugees.Altruism to Syrians reports how many JD out of 5 the respondent allocated to a charity supporting syrian refugees. The other two options were allocating to a charity supporting low-income Jordanians or keeping for self. **Panel B** reports the two significant subcomponents of the social attitudes and perceptions index. **Panel C** reports other statistically significant treatment effects on neighbors. Subjective health is the response to "Would you describe your general health as good, fair, poor, or very poor? " Life satisfaction is the response to "All things considered, how satisfied are you with your life as a whole these days on a scale of 1 to 10? " Days of media consumption is the response to "In the past 7 days, how many days did you read or listen to the news from any source, including newspapers, online, WhatsApp, etc.?" Neighborhood perceptions of refugee aid reciept is the response to a question asking "Of the refugee households in your neighborhood who receive assistance, what do you think is the average value in Dinar of the assistance (in cash or in kind) that they receive from organizations in a typical month?" The sample size for this question is smaller due to the large number of "don't know" responses. There is no treatment effect on the probability of reporting "don't know" to this question.

In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values in panel A and C are calculated per Anderson (2008) using the outcome families in the pre-analysis plan. Panel B was not prespecified and is reported to aid in interpretation of the negative treatment effect on social attitudes and perceptions. The q-values are estimated using using an outcome family comprised of all the components of the social attitudes and perceptions index.

A Estimates Correcting for Non-Random Attrition

Table (A.1) Pooled Treatment Effects on Primary Outcomes (with Lee Bounds)

Outcome	TOT	(se)	FDR q-values	Lower Bound	Upper Bound	Control mean	(sd)	N	Rounds
Panel A: Primary Outcomes									
Overall Housing Quality (Z-Score)	0.31	(0.20)	[0.19]	0.16	0.80***	-0.01	(1.00)	4,313	1, 2, 3
Total Monthly Housing Expenditures (USD PPP)	-82.05***	(32.07)	[0.07]	-166.08***	-37.20	221.63	(202.26)	1,421	2
Food Consumption (Log USD PPP)	-0.02	(0.07)	[0.71]	-0.32***	0.42***	4.39	(0.72)	2,780	2, 3
Log Total Consumption (Log USD PPP)	-0.03	(0.08)	[0.68]	-0.26***	0.27***	8.73	(0.55)	1,422	2
CESD Score (Higher: Less Depression)	-0.10	(0.13)	[0.52]	-0.68***	0.61***	0.01	(1.00)	4,302	1, 2, 3
SDQ Score (Higher: Better Child Wellbelling)	-0.34**	(0.15)	[0.07]	-0.55***	-0.04	-0.01	(1.00)	1,782	2, 3
Panel B: Other Outcomes									
At least 30 JD (95 USD PPP) in savings (=1)	0.08**	(0.04)	[0.05]	-0.15***	0.11***	0.10	(0.30)	2,797	2, 3
Respondent Hunger Last Week (=1)	0.18***	(0.05)	[0.01]	-0.05	0.38***	0.39	(0.49)	4,261	1, 2, 3
Adult Hunger Last Week (=1)	0.13***	(0.05)	[0.03]	-0.10**	0.29***	0.36	(0.48)	4,177	1, 2, 3
Child Hunger Last Week (=1)	0.15***	(0.05)	[0.01]	-0.13***	0.27***	0.25	(0.43)	3,813	1, 2, 3

Notes: This table reports 5 primary outcomes plus per capita food consumption (not pre-specified), pooled across as many rounds as were collected for each outcome. The table shows that treatment had no impact on housing quality on average over all 3 rounds, reduced housing expenditures at endline, had no impact on food consumption over endline and followup, had no impact on depression over all 3 rounds, and reduced child socio-emotional wellbeing over rounds 2 and 3. Overall Housing Quality is defined as a normalized housing quality index that includes indicators for quality floors, roofs, and walls, indicators for access to grid electricity and piped water, and the number of people per room. Total monthly housing expenditures is only reported for endline in the pooled table due to variations in measurement of housing expenditures across rounds and due to when treatment ended. Housing expenditure was measured including assistance payments at midline, and excluding assistance payments at endline since most recipients had stopped receiving payments but were still subject to rent freezes at part of the program. Results on housing expenditures in each round are reported in the appendix. Total consumption, including food and non-food expenses, was only measured at endline due to the length of the module. Reported in this table is food consumption, which was measured at endline and followup. Total food consumption is the log of the sum of food consumed in last seven days (Cereals and cereal products, Live animals, meat, and other parts of slaughtered land animals; Fish and other seafood; Milk, other dairy products, and eggs; Oils and fats; Fruits and nuts; Vegetables, tubers, pulses; Sugar and desserts; Ready-made food and other food products (baby food, spices)), plus home-produced foods produced in the last 7 days, plus assistance received in the last 30 days divided by 4.33. CES-D was measured in all rounds with exactly the same questionnaire. It includes 10 questions on depressive symptoms. Individuals who skipped more than 2 questions were marked as missing. The continuous scores were standardized with respect to the control group. The Child Strengths and Difficulties questionnaire (SDQ) was administered regarding a randomly selected child aged 3-8 at endline. At follow-up, the SDQ was administered regarding the same child. If not surveyed previously or if that child is no longer in the household, the survey was administered with respect to a randomly selected child age 4-9. The continuous scores were randomized with respect to the control group. At least 30 JD in savings (=1) is an indicator equal to one if the respondent answered yes to "Do you currently have at least 30 JDs in personal savings you can draw from in an emergency? Whether or not it is in a bank?". Respondent, adult, and child hunger last week are indicators equal to 1 if the respondent, other adults in the household or children in the household (respectively) went to bed hungry at least once in the last 7 days, zero otherwise. The lower and upper bounds refer to the ones proposed in Lee (2009) to correct for attrition. In round 3, there are only 70 treated households that answered "yes" to having at least 30 JD, but 84 were needed for the Lee bounds. To overcome this issue, all 70 who answered by "yes" are replaced by "no" in round 3 when finding the lower bound. Round 2 does not suffer from this problem, which is why we still have some variation to run the regression.

Table (A.2) Statistically Significant Impacts of the Program at Midline (with Lee Bounds)

Outcome	TOT	(se)	FDR q-values	Lower Bound	Upper Bound	Control mean	(sd)	N
Panel A: Housing Improvements								
Quality Roof (=1)	0.24***	(0.09)	[0.08]	0.23**	0.33***	0.71	(0.45)	1,610
Clean water (=1)	0.21***	(0.06)	[0.01]	0.07	0.25***	0.17	(0.38)	1,610
Occupied Rooms	0.51***	(0.17)	[0.04]	0.14	0.79***	2.85	(1.05)	1,610
Panel B: Food Aid and Insecurity								
Food consumption (aid) USD PPP	-53.05**	(22.40)	[0.02]	-72.09***	-46.97**	218.51	(201.68)	1,569
Food consumption (aid) USD PPP (*Per capita*)	-9.73***	(3.44)	[0.01]	-13.03***	-8.73***	35.88	(29.19)	1,569
Respondent hunger last week (=1)	0.23***	(0.07)	[0.01]	0.12*	0.29***	0.35	(0.48)	1,604
Adults hunger last week (=1)	0.17***	(0.06)	[0.01]	0.07	0.24***	0.38	(0.49)	1,610
Child hunger last week (=1)	0.24***	(0.07)	[0.00]	0.03	0.30***	0.23	(0.42)	1,469
Panel C: Household Recomposition								
Number of children under 18 years	0.30**	(0.15)	[0.03]	-0.14	0.56***	3.27	(1.97)	1,610
Number of boys aged 13-17 years in household	0.16**	(0.07)	[0.33]	-0.20***	0.24***	0.40	(0.66)	1,610
Panel D: COVID-19 Outcomes								
Weekly adult income (during first lockdown)	-20.07***	(7.19)	[0.02]	-52.48***	-17.86***	30.22	(66.86)	1,609
Total COVID-19 symptoms	0.84***	(0.19)	[0.00]	-0.03	0.94***	0.47	(1.22)	1,610
Number of people who are symptomatic	0.46***	(0.09)	[0.00]	0.03	0.52***	0.29	(0.61)	1,581
Number of children who are symptomatic	0.17***	(0.05)	[0.01]	-0.14***	0.19***	0.06	(0.32)	1,472
Number of visits to healthcare institutions	-0.56***	(0.23)	[0.04]	-1.66***	-0.40*	1.14	(2.03)	1,608
Did not keep distance (=1)	0.16***	(0.06)	[0.08]	0.09	0.26***	0.51	(0.50)	1,609
Panel E: Loans								
Loans taken pre-first-lockdown (=1)	-0.20***	(0.07)	[0.01]	-0.27***	-0.12*	0.61	(0.49)	1,608
Loans taken during first lockdown (=1)	-0.34***	(0.07)	[0.00]	-0.38***	-0.24***	0.78	(0.41)	1,608
Loans taken after first lockdown (=1)	-0.33***	(0.07)	[0.00]	-0.45***	-0.28***	0.42	(0.49)	1,607
Panel F: Other effects								
Total number of school days attended per child	-0.43*	(0.24)	[0.04]	-0.79***	-0.22	2.33	(2.49)	1,031
Child alertness (s.d. units)	-0.73***	(0.20)	[0.00]	-0.50***	-1.45***	0.00	(1.00)	975
Difficulty with self-care (z-score)	0.38***	(0.14)	[0.08]	-0.43***	0.43***	-0.00	(1.00)	1,610
Number of learning activities	0.19**	(0.09)	[0.08]	-0.21***	0.27***	0.22	(0.56)	1,385

Notes: The table shows the regression results on statistically significant results from the midline survey. Each row is its own dependent variable. Panel A outcomes are Clean Water, defined as an indicator for households having treated drinking water (such as by a filter). Rooms and rooms occupied by family members are reported by the respondents. Panel B outcomes are food consumption, Respondent (and adult) food insecurity equals 1 if the respondent went to bed hungry on at least one day in the last week and 0 otherwise. Panel C outcomes are self-explanatory. Panel D outcomes are COVID symptoms, which is the sum of indicators that increase the likelihood of COVID contraction by the respondent including leaving the house, attending social gatherings, not keeping distance from others, going to mosque or other religious institutions, going to grocery store or market, and leaving village/neighborhood. The other outcomes are as explained in the labels. Panel E The periods of the loans referred to are before lockdown: (January 15 - March 15, 2020); during lockdown: (March 15 - May 15, 2020); and after lockdown: (May 15 - interview date). Panel F outcomes are Total number of school days attended per child is the total number of school days (from none to 5 days) attended by all school-aged children in ages 9-17, divided by the total number of children who are school-aged; Difficulty with self-scare with higher values indicating more disability. This is an item of the The Washington Group Short Set on Functioning; learning activities are done in the last 24 hours and include homework, e-learning, educational programs and videos, and reading; Respondent networks corresponds to people (excl. household members) that the respondent knew who were already in their current neighborhood when they moved to their current residence and is coded as 1 if "0", 2 if "1-5", 3 if "6-10", 4 if "11-20", and 5 if "More than 20". The other outcomes are as explained in the labels. Monetary values are in USD PPP and are winsorized at the top 1% of values in order to limit the influence of outliers. Health visits is also winsorized at the top 1% due to outliers in that measure. The regressions also have month-by-year fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to various families in the pre-analysis plan. The lower and upper bounds refer to the ones proposed in Lee (2009) to correct for attrition. The lower bounds marked by NA for number of children is not possible to report because trimming the lower tail would result in no variation in the variable, making the regression impossible to run as it becomes a column of zeros.

Table (A.3) Statistically Significant Impacts of the Program at Endline and Follow-up (with Lee bounds)

Outcome	TOT	(se)	FDR q-values	Lower Bound	Upper Bound	Control mean	(sd)	N
Panel A: Endline								
Focus Respondent Childcare & Chores (Hours)	-6.91***	(2.02)	[0.01]	-13.27***	-5.98***	17.76	(18.39)	1,418
At least 30 JD (95 USD PPP) in savings (=1)	0.14***	(0.05)	[0.02]	-0.07	0.18***	0.12	(0.32)	1,422
Applied but Did Not Take Loan (=1)	0.06***	(0.02)	[0.00]	-0.02***	0.06***	0.01	(0.10)	1,422
Total Labor and Chores Hours Last Week	-10.62***	(3.24)	[0.01]	-22.95***	-6.61**	25.54	(23.39)	1,422
IP Assistance in Last 12 Months (=1)	0.08**	(0.03)	[0.14]	-0.11***	0.09***	0.04	(0.19)	1,422
IP Assistance in Last 12 Months (USD PPP)	128.25***	(46.50)	[0.10]	-88.67***	145.45***	28.74	(198.87)	1,422
General Happiness (Scale 1-3)	-0.22***	(0.08)	[0.02]	-0.51***	-0.04	1.76	(0.61)	1,422
Panel B: Follow-up								
Child SDQ score, std	-0.56***	(0.17)	[0.01]	-0.72***	-0.33*	-0.01	(1.00)	859

Notes: The table shows the regression results on statistically significant results from the endline survey. Each row is its own dependent variable. Monetary values are in USD PPP and are winsorized at the top 1% of values in order to limit the influence of outliers. The regressions also have month-by-year fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to various families in the pre-analysis plan. The lower and upper bounds refer to the ones proposed in Lee (2009) to correct for attrition. The lower bounds marked by NA are not possible to report because trimming the lower tail would result in no variation in the variable, making the regression impossible to run as it becomes a column of zeros. The outcome "Focus Respondent Childcare Chores" was pre-specified in two outcome families; the table reports the lower of the two q-values; the higher q-value is 0.02

Table (A.4) Impacts of the Program on Neighbor's Attitudes Toward Refugees (with Lee Bounds)

Outcome	TOT	(se)	FDR q-values	Lower Bound	Upper Bound	Control mean	(sd)	N
Panel A: Primary Outcomes								
Social attitudes & perceptions (SD)	-0.33**	(0.14)	[0.08]	-0.75***	-0.01	-0.00	(1.00)	1,102
Economic attitudes & perceptions (SD)	0.02	(0.16)	[0.56]	-0.55***	0.22	0.02	(1.02)	1,102
Policy preferences (SD)	0.17	(0.15)	[0.37]	-0.28*	0.49***	-0.02	(1.00)	1,102
Altruism to Syrians	0.21	(0.18)	[0.37]	-0.49***	0.47**	0.88	(1.18)	1,102
Panel B: Selected Social Attitudes Index Components								
Of the 3 people you exchange advice with, how many are Syrian refugees?	-0.24**	(0.11)	[0.32]	-0.56***	-0.19*	0.34	(0.72)	1,102
Do the children in this household have any Syrian refugee friends?	-0.19*	(0.11)	[0.37]	-0.33***	-0.11	0.40	(0.49)	704
Panel C: Other Related Outcomes								
Days of media consumption (last week)	-1.08***	(0.40)	[0.08]	-1.69***	-0.04	4.78	(2.78)	1,102
Neighbor perceptions of average refugee aid receipt (PPP)	-347.17***	(109.69)	[0.04]	-636.19***	-239.58**	630.84	(657.41)	685
Life Satisfaction (SD)	-0.32**	(0.15)	[0.06]	-0.62***	0.16	0.01	(0.98)	1,097
Subjective Health (SD)	-0.30*	(0.16)	[0.06]	-0.51***	0.25*	0.02	(0.99)	1,102

Notes: This table reports impacts of treatment on Jordanian neighbors of Syrian refugees in the experimental sample. Each row is its own dependent variable. Panel A reports pre-specified treatment effects on primary neighbor outcomes. All indices are standardized so that a positive point estimate reflects more "pro-refugee" sentiments. The only statistically significant outcome is a negative treatment effect on social attitudes and perceptions. This outcome is a standardized index of the following outcomes: out of 3 closest friends, how many are Syrian refugees; out of 3 people you exchange advice with, how many are Syrian refugees; do the children in this household have Syrian refugee friends (=1); do the children in this household share recreational spaces with Syrian refugee children (=1); how comfortable would you be accepting the marriage of a friend or loved one to a Syrian refugee (1-5); how comfortable would you be accepting a Syrian refugee as a neighbor (1-5); what is the net effect of Syrian refugees on Jordanian society (1-3); are Syrian refugees hardworking or lazy (1-7). Economic attitudes and perceptions is a standardized index of the following outcomes: an indicator =1 if the respondent listed "hosting Syrian refugees" as one of the most important challenges facing Jordan (options not read); an outcome with higher values for responses associated with beliefs that Syrian refugees pay more in taxes than Jordanians; an outcome with higher values for responses associated with refugees having a positive effect on the economy.Policy preferences in a standardized index of the following outcomes: less belief that Syrian refugees should be forced to live in camps; more belief that refugees should have the right to work outside camps; more belief that refugees should be able to attain Jordanian citizenship; more support for unrestricted work permits for refugees; more support for integrated classrooms with Syrians and Jordanians; more support for refugee right to enter/exit camps freely; support for housing assistance for Syrian refugees; belief that the international community should spend more money supporting refugees.Altruism to Syrians reports how many JD out of 5 the respondent allocated to a charity supporting syrian refugees. The other two options were allocating to a charity supporting low-income Jordanians or keeping for self. Panel B reports the two significant subcomponents of the social attitudes and perceptions index. Panel C reports other statistically significant treatment effects on neighbors. Subjective health is the response to "Would you describe your general health as good, fair, poor, or very poor? " Life satisfaction is the response to "All things considered, how satisfied are you with your life as a whole these days on a scale of 1 to 10? " Days of media consumption is the response to "In the past 7 days, how many days did you read or listen to the news from any source, including newspapers, online, WhatsApp, etc.?" Neighborhood perceptions of refugee aid reciept is the response to a question asking "Of the refugee households in your neighborhood who receive assistance, what do you think is the average value in Dinar of the assistance (in cash or in kind) that they receive from organizations in a typical month?" The sample size for this question is smaller due to the large number of "don't know" responses. There is no treatment effect on the probability of reporting "don't know" to this question. The regressions include month-by-year fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to various families in the pre-analysis plan. The lower and upper bounds refer to those proposed in Lee (2009) to correct for attrition.

B Heterogeneous Effects Analysis

Table (B.1) Gender Heterogeneity at Endline

	Market Mobility	Relatives Freedom	Safety Perceptions	Gender Equality
Treat	-0.33*	-0.22	-0.11	-0.37**
	(0.19)	(0.17)	(0.21)	(0.19)
Female	-1.11***	-0.89***	-0.15**	0.30***
	(0.07)	(0.06)	(0.07)	(0.06)
Treat*Female	1.02***	0.89***	-0.12	0.47*
	(0.28)	(0.27)	(0.31)	(0.24)
p-val: T + T*female=0	0.00	0.00	0.32	0.53
Control mean	2.65	2.79	-0.00	0.00
Control sd	1.04	0.95	1.00	1.00
N	1,415	1,414	1,420	1,415

Notes: The table shows the regression results on mobility and gender outcomes. Each column is its own dependent variable. Market Mobility is defined as a z-score from the FR's response to whether they can go to the local market without permission, after informing someone, after being granted permission, or cannot go alone, with higher values indicating better mobility. Relatives Freedom is defined in the same manner as the market mobility outcome, but it asks about visiting the home of relatives, friends, or neighbors. Safety Perceptions is defined as a z-score from the FR's responses about how safe (very safe, safe, neither safe nor unsafe, unsafe, very unsafe) he or she thinks it is to walk outside during the day in the area where you live for women, with higher values indicating more safety. Gender Equality is defined as a normalized index (z-score) constructed from four items in which the FR strongly agrees, agrees, disagrees, or strongly disagrees. The four items are "A married woman can work outside the home if she wishes", "Husbands should have final say in all decisions concerning the family", "A woman can be a president or prime minister of a Muslim country", and "Women and men should have equal rights in making the decision to divorce." Each item is normalized and then the resulting index is normalized again where higher values indicate views in accordance with more gender equality. The regressions also have month-by-year fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%).

Table (B.2) Neighbor Impacts on Social Attitudes: Heterogeneity

	Social Attitudes			
Treat	-0.53***	-0.39***	-0.39***	-0.43**
	(0.22)	(0.14)	(0.14)	(0.20)
Above Median Marlow-Crowne	-0.01			
	(0.08)			
Treat*Above Median Marlow-Crowne	0.41			
	(0.30)			
Palestinian Grandparents		-0.03		
		(0.10)		
Treat*Palestinian Grandparents		0.54		
		(0.45)		
Non-Jordanian Grandparents			0.16	
			(0.11)	
Treat*Non-Jordanian Grandparents			0.35	
			(0.45)	
Above Median Distance				-0.07
				(0.08)
Treat*Above Median Distance				0.39
				(0.28)
p-val: T + T*Het = 0	0.52	0.73	0.93	0.85
Control Mean	-0.00	-0.00	-0.00	-0.00
Control SD	1.00	1.00	1.00	1.00
N	1,102	1,102	1,102	1,044

Notes: Each column reports a TOT regression of the social attitudes index on treatment fully interacted with one of 4 aspects of heterogeneity. Palestinian is an indicator equal to one if the Jordanian respondent reported one or more of their grandparents was born in Palestine, or answered yes to "Are you or your spouse of Palestinian descent?" Non-Jordanian grandparent is equal to one if the respondent indicated at least one of their grandparents or their spouse's grandparents was not born in Jordan. Above median Marlow-Crowne is equal to one if their Marlow-Crowne score was above median, indicating a higher propensity to give socially desirable responses. Above median distance is equal to one if the neighbor was above median distance away from the Syrian refugee in the experiment. Median distance was 63 meters. Distance is measured in meters between the refugee house and the neighbor house.

C Additional results

Table (C.1) Impacts of the Program on Earnings, Labor, & Occupational Choice

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Total Earnings in Last 30 Days (PPP, IHS)	-0.51	(0.38)	[0.33]	1.59	(2.83)	1,422
Total Labor Hours Last Week	-3.29	(2.35)	[0.33]	7.72	(18.73)	1,422
Total Labor Hours [Monthly Average]	-0.65	(8.20)	[0.78]	24.19	(67.80)	1,422
Net Wage-Employment Income Last Month (PPP, IHS)	-0.46	(0.37)	[0.40]	1.56	(2.81)	1,417

Notes: The table shows the regression results on pre-specified labor market outcomes of the Focus Respondent using the 2021 in-person data. Each row is its own dependent variable. The outcomes that require definitions are: Total Earnings is defined as the total of business profits and net wage salary in all jobs and is transformed using the inverse hyperbolic sine transformation. Total Labor Hours [Monthly Average] is defined as average weekly wage-employment hours multiplied by the numbers of months worked in last 12 months multiplied by 52/12 to obtain a monthly average of all jobs. The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment. Monetary values are in USD PPP. Total Earnings in Last 30 Days, Taxes Paid in Last 30 Days, Net Wage-Employment Income Last Month, Self-Employment Profits, Expenses, and Revenues Last Month and Year, Self-Employment Profits Last Year, Total Employees Last Month, Self-Employment Rent Last Month, and Total Labor Hours [Average] are winsorized at the top 1% of values in order to limit the influence of outliers. The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 4 in the pre-analysis plan.

Housing Subsidies for Refugees: Experimental Evidence on Life Outcomes and Social Integration in Jordan

Last modified: January 14th, 2025

Contents

1 Appendix D: Midline Results **4**

2 Appendix E: Endline Results **14**

3 Appendix F: Follow-Up Results **31**

4 Appendix G: Social Integration Results **39**
4.1 Social Cohesion Heterogeneity 49

5 Appendix H: Ethics Appendix **53**

6 Appendix I: Pooled Results **54**
6.1 Pooled Results Heterogeneity 54

List of Tables

D1 Impacts of the Program on Dwelling Characteristics and Household Composition . 5
D2 Impacts of the Program on Household Consumption and Expenditure 6
D3 Impacts of the Program on Migration 6
D4 Impacts of the Program on Physical Health 7
D5 Impacts of the Program on Mental Health 8
D6 Impacts of the Program on Household-Level Child Outcomes 8
D7 Impacts of the Program on COVID-19 Health-Related Outcomes 9
D8 Impacts of the Program on COVID-19 Labor-Related Outcomes 10
D9 Impacts of the Program on COVID-19 Credit-Related Outcomes 11
D10 Impacts of the Program on Household Composition 12
D11 Impacts of the Program on Individual Components of Indices 13
E1 Impacts of the Program on Primary Outcomes 14
E2 Impacts of the Program on Dwelling Characteristics and Household Composition . 15
E3 Impacts of the Program on Consumption and Expenditures 16
E4 Impacts of the Program on Financial Participation 17
E5 Impacts of the Program on Transfers 18
E6 Impacts of the Program on Earnings, Labor, & Occupational Choice 19
E7 Impacts of the Program on Migration 20
E8 Impacts of the Program on Physical Health 21
E9 Impacts of the Program on Mental Health 22
E10 Impacts of the Program on Sleep 22
E11 Impacts of the Program on Marriage and Fertility 23
E12 Impacts of the Program on Individual Child Outcomes 24
E13 Impacts of the Program on Household-Level Child Outcomes 24
E14 Impacts of the Program on Social Capital 25
E15 Impacts of the Program on Political Attitudes 26
E16 Impacts of the Program on Time Use 26
E17 Impacts of the Program on Education and Cognition 27
E18 Impacts of the Program on Behavioral Games and Preferences 28
E19 Impacts of the Program on Additional Dwelling Characteristics and Household Composition 29
E20 Impacts of the Program on NGO Assistance 30
F1 Impacts of the Program on Primary Outcomes 31
F2 Impacts of the Program on Dwelling Characteristics and Household Composition . 32
F3 Impacts of the Program on Food Consumption and Food Security 33
F4 Impacts of the Program on Financial Participation 34
F5 Impacts of the Program on Earnings, Labor, and Occupational Choice 34
F6 Impacts of the Program on Migration 35
F7 Impacts of the Program on Physical and Mental Health 35
F8 Impacts of the Program on Child Outcomes 36
F9 Impacts of the Program on Time Use 37
F10 Impacts of the Program on Relationships and MacArthur Ladder 38
F11 Impacts of the Program on Preferences towards it 38

G1 Impacts of the Program on Primary Outcomes . . . 39
G2 Impacts of the Program on Assimilation Gap . . . 40
G3 Impacts of the Program on Host Community Relations and Attitudes Towards Refugees . . . 41
G4 Impacts of the Program on Altruism and Trust . . . 42
G5 Impacts of the Program on Social Attitudes and Policy Preferences . . . 43
G6 Impacts of the Program on Dwelling Characteristics . . . 44
G7 Impacts of the Program on Household Consumption and Expenditures . . . 45
G8 Impacts of the Program on Food Security . . . 46
G9 Impacts of the Program on Earnings, Labor, and Occupational Choice . . . 47
G10 Impacts of the Program on Savings and Loans . . . 48
G11 Impacts of the Program on Physical and Mental Health . . . 48
G12 Jordanian Neighbor Primary Outcomes Heterogeneity by: Marlowe-Crowne Score 49
G13 Jordanian Neighbor Primary Outcomes Heterogeneity by: Palestinian Grandparents 49
G14 Jordanian Neighbor Primary Outcomes Heterogeneity by: Grandparents Non-Native Jordanians . . . 50
G15 Jordanian Neighbor Primary Outcomes Heterogeneity by: Proximity to Study Refugee Household . . . 50
G16 Jordanian Neighbor Primary Outcomes Heterogeneity by: Gender . . . 51
G17 Jordanian Neighbor Primary Outcomes Heterogeneity by: Age Group 18-25 . . . 51
G18 Jordanian Neighbor Primary Outcomes Heterogeneity by: Education Level . . . 52
G19 Jordanian Neighbor Primary Outcomes Heterogeneity by: Socioeconomic Status . 52
I1 Pooled Primary Outcomes Effects Heterogeneity by: Respondent Genderl . . . 54

1 Appendix D: Midline Results

Table (D1) Impacts of the Program on Dwelling Characteristics and Household Composition

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Overall housing quality (Z-Score)	0.48*	(0.26)	[0.16]	-0.00	(1.00)	1,610
Housing-material quality (Score 0-2)	0.44**	(0.20)	[0.16]	1.39	(0.83)	1,610
Toilet (=1)	-0.14	(0.14)	[0.26]	0.23	(0.42)	1,610
Piped main water (=1)	0.01	(0.09)	[0.50]	0.24	(0.43)	1,610
Main water access (Scale 1-5)	0.35*	(0.19)	[0.16]	3.42	(1.15)	1,610
Clean water (=1)	0.21***	(0.06)	[0.01]	0.17	(0.38)	1,610
Safe drinking water (Scale 1-5)	0.20	(0.18)	[0.23]	3.46	(1.08)	1,610
Electricity (=1)	0.01**	(0.01)	[0.16]	0.99	(0.07)	1,610
Generator (=1)	-0.01	(0.01)	[0.23]	0.01	(0.08)	1,610
Shared dwelling (=1)	-0.02*	(0.01)	[0.16]	0.01	(0.09)	1,610
# Families per dwelling	-0.02	(0.01)	[0.18]	1.01	(0.11)	1,610
Rooms	0.44***	(0.18)	[0.12]	2.93	(1.08)	1,610
Occupied Rooms	0.51***	(0.17)	[0.04]	2.85	(1.05)	1,610
Total Monthly housing expenditures	41.35	(25.75)	[0.18]	164.59	(177.61)	1,545
Monthly housing expenditures (*per capita*)	8.73**	(4.17)	[0.16]	29.88	(34.75)	1,545
Monthly rent paid (USD PPP)	46.16*	(25.60)	[0.16]	168.43	(177.17)	1,505
Rent amount agreed on (USD PPP)	-25.58	(30.69)	[0.29]	327.76	(115.70)	1,445
Household size	0.29	(0.26)	[0.23]	5.96	(2.39)	1,610
Total ID cards	-0.22	(0.15)	[0.18]	4.20	(0.96)	1,610
Home ownership (=1)	-0.01	(0.02)	[0.45]	0.03	(0.16)	1,610
Respondent chores (Hours)	-1.58	(1.97)	[0.29]	17.40	(18.73)	1,603
Respondent childcare (Hours)	-6.26**	(3.10)	[0.16]	29.90	(27.94)	1,507
Respondent childcare & chores (Hours)	-5.62	(4.08)	[0.18]	45.05	(38.52)	1,607
Other members childcare & chores (Hours)	-0.87	(3.89)	[0.45]	42.05	(29.65)	1,151
Dwelling can be locked (=1)	0.14**	(0.07)	[0.16]	0.79	(0.41)	1,610

Notes:

- The table shows the regression results on pre-specified housing quality and housing-related finances using the phone survey midline data. Each row is its own dependent variable.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The outcomes that require definitions are: Overall housing quality is defined as a normalized housing quality index that includes indicators for quality floors and roofs, indicators for access to grid electricity and piped water, and the number of people per room. Housing-material quality is defined as the summation of three indicators for high-quality floors and roofs. Clean water is defined as an indicator for households having treated drinking water (such as by a filter). *Total* Monthly housing expenditures is defined as the sum of rent paid, mortgage, and upgrade cost. It includes the subsidy payments by the implementing organization. The *Per capita* housing expenditures is divided by household size. Total ID Cards is defined as the sum of the ID cards possessed by the respondent (MOI card, Passport, Residency permit, Work permit, Family book, Syrian ID, and UNHCR file). Hours spent on childcare and chores are measured over the last week.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age).
- In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 1 in the pre-analysis plan.

Table (D2) Impacts of the Program on Household Consumption and Expenditure

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Food consumption (aid) USD PPP	-53.05**	(22.40)	[0.02]	218.51	(201.68)	1,569
Food consumption (aid) USD PPP (*Per capita*)	-9.73***	(3.44)	[0.01]	35.88	(29.19)	1,569
Number of meals consumed the day prior	-0.08	(0.09)	[0.13]	2.05	(0.59)	1,610

Notes:

- The table shows the regression results on pre-specified consumption outcomes using the phone survey midline data. Each row is its own dependent variable.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The outcomes that require definitions are: Food consumption measures are self-reported and are in USD PPP and are winsorized at the top 1% of values in order to limit the influence of outliers.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age).
- In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 2 in the pre-analysis plan.

Table (D3) Impacts of the Program on Migration

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Respondent moved since October 2019 (=1)	0.07	(0.06)	[0.28]	0.10	(0.30)	1,609
Number of times the subject moved	0.19**	(0.10)	[0.11]	0.13	(0.48)	1,609
Respondent networks (s.d. units)	0.35**	(0.15)	[0.10]	0.00	(1.00)	1,610
Camp resident (=1)	-0.02	(0.02)	[0.30]	0.02	(0.15)	1,609
Moving plans (Scale 1-4)	0.14	(0.16)	[0.32]	3.67	(0.71)	862

Notes:

- The table shows the regression results on pre-specified migration outcomes using the phone survey midline data. Each row is its own dependent variable.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The outcomes that require definitions are: Respondent moved is asked about since October 2019 because October 2019 represents the earliest date of the shelter program implementation. Respondent networks corresponds to people (excl. household members) that the respondent knew who were already in their current neighborhood when they moved to their current residence and is coded as 1 if "0", 2 if "1-5", 3 if "6-10", 4 if "11-20", and 5 if "More than 20". Moving plans is coded as 1 if moving on a specific date, 2 if looked for other residences, 3 if have not started looking for other residences, and 4 if do not have plans to change residences.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age).
- In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 5 in the pre-analysis plan.

Table (D4) Impacts of the Program on Physical Health

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Subjective health (Scale 1-4)	-0.12	(0.11)	[0.32]	3.34	(0.70)	1,610
Disability level 1 (=1 if disabled)	0.10	(0.06)	[0.17]	0.61	(0.49)	1,610
Disability level 2 (=1 if disabled)	0.06	(0.08)	[0.37]	0.41	(0.49)	1,610
Disability level 3 (=1 if disabled)	-0.00	(0.06)	[0.71]	0.26	(0.44)	1,610
Disability level 4 (=1 if disabled)	0.04***	(0.01)	[0.01]	0.01	(0.08)	1,610
Respondent hunger last week (=1)	0.23***	(0.07)	[0.01]	0.35	(0.48)	1,604
Adults hunger last week (=1)	0.17***	(0.06)	[0.01]	0.38	(0.49)	1,610
Respondent hunger last week (Days)	0.65***	(0.22)	[0.01]	0.91	(1.46)	1,604
Adult hunger last week (Days)	0.55***	(0.21)	[0.01]	0.79	(1.36)	1,444

Notes:

- The table shows the regression results on pre-specified physical health outcomes using the phone survey midline data. Each row is its own dependent variable.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The outcomes that require definitions are: Disability is defined using the Washington Group Short Set on Functioning and equals 1 if the respondent is classified as disabled and 0 otherwise. Disability 3 is the recommended cutoff and as we moce from Disabiloty 1 to Disability 4, disability level becomes more extreme. Respondent (and adult) hunger equals 1 if the respondent went to bed hungry on at least one day in the last week and 0 otherwise.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age).
- In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 6 in the pre-analysis plan.

Table (D5) Impacts of the Program on Mental Health

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
CESD score (higher is better)	0.13	(0.17)	[0.37]	0.00	(1.00)	1,607
Perceived stress scale (higher is better)	0.10	(0.15)	[0.38]	-0.00	(1.00)	1,607
Depression (=1 if depressed)	-0.09	(0.07)	[0.29]	0.56	(0.50)	1,607

Notes:

- The table shows the regression results on pre-specified mental health outcomes using the phone survey midline data. Each row is its own dependent variable.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The outcomes that require definitions are: Center for Epidemiologic Studies Depression Scale (CES-D) is a measure for depressive symptoms and includes 10 items asking about the past week. It is defined as a mean-effect index of 10 items scored 0-3 in the past week (I was bothered by things that usually don't bother me, I had a problem in concentration on what I was doing, I felt depressed and troubled in my mind, I felt that everything that I did took up all my energy, I felt hopeful about the future (reversed), I felt afraid, I had difficulty in sleeping peacefully, I was happy (reversed), I felt lonely, I lacked the motivation to do anything). It is then normalized by subtracting the control group's mean and dividing by its standard deviation, and reversing again such that higher values correspond to better mental health. The Perceived Stress Scale (PSS-4) is a mean-effect index of four items measured over 1-5 in the last 30 days (how often have you felt that you were unable to control the important things in your life?, how often have you felt certain in your ability to overcome your own personal problems?, how often have you felt that things were going your way?, how often did you feel that the problems were too much for you to manage?). It is then normalized by subtracting the control group's mean and dividing by its standard deviation. Higher scores for CES-D and PSS indicate better outcomes. Depression equals 1 if the CESD score is at least 10 and 0 otherwise.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age).
- In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 6 in the pre-analysis plan.

Table (D6) Impacts of the Program on Household-Level Child Outcomes

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Number of children under 18 years	0.30**	(0.15)	[0.03]	3.27	(1.97)	1,610
Total number of school days attended per child	-0.43*	(0.24)	[0.04]	2.33	(2.49)	1,031
Days any child had to sleep hungry	0.60***	(0.20)	[0.01]	0.57	(1.22)	1,469
Child hunger last week (=1)	0.24***	(0.07)	[0.00]	0.23	(0.42)	1,469

Notes:

- The table shows the regression results on pre-specified and not pre-specified child outcomes using the phone survey midline data. Each row is its own dependent variable.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The outcomes that require definitions are: Child hunger last week is defined as an indicator variable that equals 1 if any child in household had to go to bed hungry in the past week and 0 otherwise.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age).
- In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 8 in the pre-analysis plan.

Table (D7) Impacts of the Program on COVID-19 Health-Related Outcomes

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Adults COVID-19 testing (=1 if any adult tested)	0.00	(0.04)	[0.83]	0.07	(0.25)	1,609
Number of adults tested for COVID-19	0.04	(0.05)	[0.54]	0.06	(0.37)	1,558
Children COVID-19 testing (=1)	0.00	(0.02)	[0.77]	0.01	(0.11)	1,604
Number of children tested for COVID-19	0.03	(0.03)	[0.54]	0.01	(0.13)	1,519
Unable to access testing (=1)	0.01	(0.02)	[0.78]	0.02	(0.14)	1,603
Total COVID-19 symptoms	0.84***	(0.19)	[0.00]	0.47	(1.22)	1,610
COVID-19 symptoms index (z-score)	0.69***	(0.14)	[0.00]	0.00	(1.00)	1,610
Number of people who are symptomatic	0.46***	(0.09)	[0.00]	0.29	(0.61)	1,581
Number of children who are symptomatic	0.17***	(0.05)	[0.01]	0.06	(0.32)	1,472
Days prior that these symptoms have appeared	1.10	(2.08)	[0.68]	7.62	(10.62)	372
Recovery (=1 if respondent recovered)	-0.10	(0.16)	[0.65]	0.74	(0.44)	403
Recovery (=1 if any member recovered)	-0.06	(0.15)	[0.77]	0.76	(0.43)	438
COVID-19 riskiness index	0.19	(0.13)	[0.31]	-0.00	(1.00)	1,609
Sum of six COVID-19-spreading activities	0.39	(0.27)	[0.31]	2.42	(2.06)	1,609
Number of visits to healthcare institutions	-0.56***	(0.23)	[0.04]	1.14	(2.03)	1,608
Seen by a health professional (=1)	-0.08	(0.05)	[0.31]	0.37	(0.48)	1,610
Healthcare expenditures since COVID-19 (USD PPP)	-77.66	(58.11)	[0.33]	130.23	(275.54)	583
Medicine expenditures since COVID-19 (USD PPP)	-30.75*	(18.72)	[0.24]	84.04	(141.52)	1,590
Number of foregone health visits	-1.19	(1.10)	[0.47]	2.65	(8.05)	1,607
Denied service because of refugee status (=1)	-0.04	(0.05)	[0.60]	0.16	(0.37)	1,609
Major health problems since October 2019 (=1)	0.01	(0.04)	[0.77]	0.07	(0.25)	1,610
Major health problems since October 2019 (#)	0.04	(0.04)	[0.54]	0.07	(0.26)	1,610
Change in overall food consumption (-1,0,1)	0.33***	(0.11)	[0.02]	-0.30	(0.79)	1,606
Change in cereal consumption (-1,0,1)	0.17	(0.11)	[0.31]	-0.17	(0.81)	1,603
Change in meat consumption (-1,0,1)	0.11	(0.09)	[0.40]	-0.61	(0.61)	1,606
Number of learning activities	0.19**	(0.09)	[0.08]	0.22	(0.56)	1,385
Learning activities (=1)	0.15***	(0.06)	[0.03]	0.17	(0.37)	1,385
Hours spent on learning activities (average)	0.07	(0.18)	[0.77]	0.43	(1.17)	1,248
Child happiness (Scale 1-7)	-0.22	(0.38)	[0.65]	5.07	(1.90)	974
Child alertness (s.d. units)	-0.73***	(0.20)	[0.00]	0.00	(1.00)	975

Notes:

- The table shows the regression results on pre-specified COVID-19-related outcomes using the phone survey midline data. Each row is its own dependent variable.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The outcomes that require definitions are: The COVID symptoms index is a mean effect index of an indicator for any household member testing positive for COVID and the summation of total symptoms experienced by the respondent. The COVID Riskiness index is a mean effect index of indicators that increase the likelihood of COVID contraction by the respondent including leaving the house, attending social gatherings, not keeping distance from others, going to mosque or other religious institutions, going to grocery store or market, and leaving village/neighborhood. The change in consumption is obtained from asking "Please tick below the changes in your family's consumption in the past 30 days compared to before the COVID-19 outbreak began: Increase, No change, Decrease." Then, increase is coded as +1, no change as 0, and decrease as -1. Learning activities are done in the last 24 hours and include homework, e-learning, educational programs and videos, and reading. Child alertness is the respondent's answer to "On a scale of 1-7, with 1 being tired and 7 being alert, how does (child name) feel right now? " and is normalized by deducting the control group's mean and then dividing the result by the standard deviation of the control group.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age).
- In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 14 in the pre-analysis plan.

Table (D8) Impacts of the Program on COVID-19 Labor-Related Outcomes

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Weekly adult hours worked (before first lockdown)	1.73	(4.93)	[0.77]	29.99	(31.33)	1,604
Weekly adult hours worked (during first lockdown)	0.46	(0.74)	[0.65]	1.33	(6.39)	1,610
Weekly adult hours worked (after first lockdown)	0.55	(3.67)	[0.81]	24.11	(28.87)	1,606
Weekly adult income (before first lockdown)	-21.60	(15.83)	[0.33]	117.16	(112.22)	1,606
Weekly adult income (during first lockdown)	-20.07***	(7.19)	[0.02]	30.22	(66.86)	1,609
Weekly adult income (after first lockdown)	-3.27	(12.47)	[0.77]	99.37	(100.70)	1,605
Assets sold [before lockdown] (=1)	0.10	(0.08)	[0.34]	0.25	(0.43)	1,600
Assets sold [during lockdown] (=1)	0.04	(0.07)	[0.65]	0.28	(0.45)	1,602
Assets sold [after lockdown] (=1)	-0.02	(0.04)	[0.69]	0.11	(0.32)	1,608
Days of work/housework/school missed	0.47	(1.55)	[0.77]	4.48	(11.67)	1,407
Lost job (=1)	0.00	(0.06)	[0.83]	0.66	(0.48)	1,606

Notes:

- The table shows the regression results on pre-specified COVID-19-related outcomes using the phone survey midline data. Each row is its own dependent variable.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The periods referred to in the tables are before lockdown: (January 15 - March 15, 2020); during lockdown: (March 15 - May 15, 2020); and after lockdown: (May 15 - interview date). Monetary values are in USD PPP and are winsorized at the top 1% of values in order to limit the influence of outliers.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age).
- In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 14 in the pre-analysis plan.

Table (D9) Impacts of the Program on COVID-19 Credit-Related Outcomes

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Loans taken pre-first-lockdown (=1)	-0.20***	(0.07)	[0.01]	0.61	(0.49)	1,608
Log loans taken pre-first-lockdown	-0.39**	(0.16)	[0.04]	5.48	(0.85)	923
Loans taken during first lockdown (=1)	-0.34***	(0.07)	[0.00]	0.78	(0.41)	1,608
Log loans taken during first lockdown	-0.28*	(0.16)	[0.19]	5.55	(0.77)	1,167
Loans taken after first lockdown (=1)	-0.33***	(0.07)	[0.00]	0.42	(0.49)	1,607
Log loans taken after first lockdown	-0.80***	(0.19)	[0.00]	5.08	(0.86)	586
Loans given pre-first-lockdown (=1)	0.02*	(0.01)	[0.21]	0.00	(0.06)	1,610
Loans given during first lockdown (=1)	0.00	(0.00)	[0.40]	0.00	(0.00)	1,610
Loans given after first lockdown (=1)	0.00	(0.00)	[0.77]	0.00	(0.04)	1,610

Notes:

- The table shows the regression results on pre-specified COVID-19-related outcomes using the phone survey midline data. Each row is its own dependent variable.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The periods referred to in the tables are before lockdown: (January 15 - March 15, 2020); during lockdown: (March 15 - May 15, 2020); and after lockdown: (May 15 - interview date). Monetary values are in USD PPP and are winsorized at the top 1% of values in order to limit the influence of outliers.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age).
- In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 14 in the pre-analysis plan.

Table (D10) Impacts of the Program on Household Composition

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Number of adults in household	-0.02	(0.17)	[1.00]	2.69	(1.38)	1,610
Number of girls under 18 years in household	0.05	(0.15)	[1.00]	1.60	(1.40)	1,610
Number of boys under 18 years in household	0.25*	(0.13)	[0.33]	1.68	(1.34)	1,610
Number of children under 12 years in household	0.08	(0.16)	[1.00]	2.49	(1.71)	1,610
Number of girls under 12 years in household	-0.00	(0.12)	[1.00]	1.21	(1.23)	1,610
Number of boys under 12 years in household	0.08	(0.12)	[1.00]	1.28	(1.15)	1,610
Number of children aged 13-17 years in household	0.22*	(0.13)	[0.33]	0.79	(0.98)	1,610
Number of girls aged 13-17 years in household	0.06	(0.10)	[1.00]	0.39	(0.65)	1,610
Number of boys aged 13-17 years in household	0.16**	(0.07)	[0.33]	0.40	(0.66)	1,610

Notes:

- The table shows the regression results on not pre-specified household composition outcomes using the phone survey midline data. Each row is its own dependent variable.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The outcomes that require definitions are: Child hunger last week is defined as an indicator variable that equals 1 if any child in household had to go to bed hungry in the past week and 0 otherwise. Learning activities are done in the last 24 hours and include homework, e-learning, educational programs and videos, and reading. Child alertness is the respondent's answer to "On a scale of 1-7, with 1 being tired and 7 being alert, how does (child name) feel right now? " and is normalized by deducting the control group's mean and then dividing the result by the standard deviation of the control group.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age).
- In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 8 in the pre-analysis plan.

Table (D11) Impacts of the Program on Individual Components of Indices

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Overall Housing Quality (Z-Score):	0.48*	(0.26)	[0.24]	-0.00	(1.00)	1,610
Quality Floor (=1)	0.20*	(0.12)	[0.24]	0.68	(0.47)	1,610
Quality Roof (=1)	0.24***	(0.09)	[0.08]	0.71	(0.45)	1,610
Grid Electricity (=1)	0.15*	(0.09)	[0.24]	0.78	(0.41)	1,610
Piped Drinking Water (=1)	0.01	(0.09)	[0.94]	0.24	(0.43)	1,610
# People per Room	-0.32*	(0.19)	[0.24]	2.37	(1.42)	1,610
Disability level 3 (=1 if disabled):	-0.00	(0.06)	[0.99]	0.26	(0.44)	1,610
Vision (Higher is more disability)	0.06	(0.08)	[0.52]	1.32	(0.59)	1,610
Hearing (Higher is more disability)	-0.04	(0.06)	[0.52]	1.12	(0.37)	1,610
Mobility (Higher is more disability)	0.04	(0.10)	[0.71]	1.60	(0.78)	1,610
Cognition (Higher is more disability)	0.20*	(0.11)	[0.24]	1.39	(0.64)	1,610
Difficulty with self-care (z-score)	0.38***	(0.14)	[0.08]	-0.00	(1.00)	1,610
Communication (Higher is more disability)	0.02	(0.04)	[0.71]	1.05	(0.25)	1,609
CESD score (Higher better):	0.13	(0.17)	[0.52]	0.00	(1.00)	1,607
felt hopeful about the future	0.27	(0.18)	[0.31]	1.47	(1.30)	1,601
was happy	-0.06	(0.13)	[0.71]	1.10	(1.02)	1,606
bothered by things that don't bother me (reversed)	0.21	(0.15)	[0.31]	2.14	(0.97)	1,603
had a problem in concentration (reversed)	0.25*	(0.14)	[0.24]	2.33	(0.92)	1,605
felt depressed and troubled in my mind (reversed)	-0.01	(0.20)	[0.99]	1.76	(1.13)	1,606
felt everything that I did took up my energy (reversed)	0.05	(0.17)	[0.80]	1.89	(1.11)	1,603
felt afraid (reversed)	0.15	(0.17)	[0.52]	2.18	(1.09)	1,605
had difficulty in sleeping peacefully (reversed)	0.14	(0.19)	[0.52]	2.01	(1.08)	1,605
felt lonely (reversed)	-0.30*	(0.18)	[0.24]	1.92	(1.18)	1,606
lacked the motivation to do anything (reversed)	0.04	(0.12)	[0.72]	2.17	(1.04)	1,602
COVID-19 symptopms Index:	0.69***	(0.14)	[0.00]	0.00	(1.00)	1,610
COVID-19-positive (=1)	0.15**	(0.08)	[0.24]	0.02	(0.13)	105
Total COVID-19 symptoms	0.84***	(0.19)	[0.00]	0.47	(1.22)	1,610
COVID-19 Riskiness Index:	0.19	(0.13)	[0.31]	-0.00	(1.00)	1,609
Left home (=1)	0.06	(0.07)	[0.52]	0.62	(0.48)	1,609
Attended social gatherings (=1)	0.05	(0.07)	[0.52]	0.36	(0.48)	1,609
Did not keep distance (=1)	0.16***	(0.06)	[0.08]	0.51	(0.50)	1,609
Attended mosque (=1)	0.09**	(0.04)	[0.19]	0.25	(0.43)	1,609
Went to market (=1)	0.01	(0.06)	[0.93]	0.46	(0.50)	1,609
Left village (=1)	0.03	(0.06)	[0.71]	0.22	(0.41)	1,609

Notes:

- The table shows the regression results on index items using the phone survey midline data. Each row is its own dependent variable.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age).
- In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to all indices components.

2 Appendix E: Endline Results

Table (E1) Impacts of the Program on Primary Outcomes

Outcomes	**Treatment**	**(se)**	**FDR q-values**	**Control mean**	**(sd)**	**N**
Overall Housing Quality (Z-Score)	0.24	(0.23)	[0.99]	0.00	(1.00)	1,422
Total Monthly Housing Expenditures (USD PPP)	-84.71***	(33.10)	[0.06]	227.72	(208.45)	1,421
Total Household Consumption (Log USD PPP)	-0.03	(0.08)	[1.00]	8.73	(0.55)	1,422
CESD Score (Higher: Less Depression)	-0.07	(0.16)	[1.00]	0.00	(1.00)	1,396
SDQ Score (Higher: Better Child Wellbelling)	-0.16	(0.18)	[0.99]	0.00	(1.00)	905

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2021 in-person data. Each row is its own dependent variable.
- Overall Housing Quality is defined as a normalized housing quality index that includes indicators for quality floors, roofs, and walls, indicators for access to grid electricity and piped water, and the number of people per room.
- Total Monthly Housing Expenditures is the sum of rent paid, mortgage, and housing upgrade costs paid in the last month.
- Total Consumption is log of the sum of food consumed in last seven days (Cereals and cereal products, Live animals, meat, and other parts of slaughtered land animals; Fish and other seafood; Milk, other dairy products, and eggs; Oils and fats; Fruits and nuts; Vegetables, tubers, pulses; Sugar and desserts; Ready-made food and other food products (baby food, spices)) multiplied buy number of months consumed in last 12 months, home-produced goods multiplied by months produced, annualized food received as gifts and annualized non-food purchases (utilities; water; infant needs; household and hygiene items; linens; clothing and footwear; basic household items; and school costs).
- Center for Epidemiologic Studies Depression Scale (CES-D) is a measure for depressive symptoms and includes 10 items asking about the past week. It is defined as a mean-effect index of 10 items scored 0-3 in the past week (I was bothered by things that usually don't bother me, I had a problem in concentration on what I was doing, I felt depressed and troubled in my mind, I felt that everything that I did took up all my energy, I felt hopeful about the future (reversed), I felt afraid, I had difficulty in sleeping peacefully, I was happy (reversed), I felt lonely, I lacked the motivation to do anything). It is then normalized by subtracting the control group's mean and dividing by its standard deviation.
- Strengths and Difficulties Questionnaire (SDQ) is a screening questionnaire administered to a randomly selected child aged three to eight years old and covers emotional, conduct, hyperactivity, peer, and prosocial problems. Both CES-D and SDQ are standardized by subtracting the control group's mean and dividing by its standard deviation.
- Higher scores for CES-D and SDQ indicate better outcomes.
- Monetary values are in USD PPP.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 1 in the pre-analysis plan.

Table (E2) Impacts of the Program on Dwelling Characteristics and Household Composition

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Overall Housing Quality (Z-Score)	0.24	(0.23)	[1.00]	0.00	(1.00)	1,422
Housing-Material Quality (Score 0-3)	0.30	(0.25)	[1.00]	2.27	(1.09)	1,422
Toilet (=1)	0.04*	(0.02)	[0.61]	0.97	(0.17)	1,422
Piped Drinking Water (=1)	-0.02	(0.07)	[1.00]	0.27	(0.44)	1,422
Piped Main Water (=1)	-0.11	(0.12)	[1.00]	0.72	(0.45)	1,422
Water Access (1-5)	-0.12	(0.21)	[1.00]	3.62	(1.09)	1,422
Clean Water (=1)	-0.02	(0.06)	[1.00]	0.22	(0.42)	1,422
Safe Drinking Water (1-5)	0.06	(0.14)	[1.00]	3.52	(1.06)	1,422
Electricity (=1)	0.00	(0.01)	[1.00]	0.98	(0.15)	1,422
Grid Electricity (=1)	0.06	(0.06)	[1.00]	0.85	(0.36)	1,422
Generator (=1)	-0.01	(0.02)	[1.00]	0.01	(0.11)	1,422
Shared Dwelling (=1)	0.03	(0.03)	[1.00]	0.93	(0.25)	1,422
Families per Dwelling	-0.01	(0.01)	[1.00]	1.01	(0.13)	1,343
Rooms	0.24	(0.30)	[1.00]	3.90	(1.49)	1,422
Occupied Rooms	0.22	(0.28)	[1.00]	3.78	(1.46)	1,422
Total Monthly Housing Expenditures (USD PPP)	-84.71***	(33.10)	[0.15]	227.72	(208.45)	1,421
Per capita Monthly Housing Expenditures (USD PPP)	-8.01	(6.73)	[1.00]	43.76	(48.22)	1,421
Monthly Rent Paid (USD PPP)	-67.81**	(30.30)	[0.25]	222.31	(178.94)	1,269
Household Size	-0.21	(0.32)	[1.00]	5.94	(2.43)	1,422
2011 Household Size	-0.32	(0.56)	[1.00]	5.88	(3.58)	1,422
Household Size Change	0.11	(0.54)	[1.00]	0.07	(3.73)	1,422
Total ID Cards	-0.09	(0.12)	[1.00]	3.61	(1.02)	1,422
Home Ownership (=1)	-0.00	(0.05)	[1.00]	0.06	(0.24)	1,421
Focus Respondent Childcare & Chores (Hours)	-6.91***	(2.02)	[0.02]	17.76	(18.39)	1,418
Other Members Childcare & Chores (Hours)	-0.50	(2.15)	[1.00]	18.54	(18.49)	1,390
Total Land Owned (Donum)	4.21	(9.55)	[1.00]	14.36	(80.11)	1,422

Notes:

- The table shows the regression results on pre-specified housing quality and housing-related finances using the 2021 in-person data. Each row is its own dependent variable.
- The outcomes that require definitions are: Overall Housing Quality is defined as a normalized housing quality index that includes indicators for quality floors, roofs, and walls, indicators for access to grid electricity and piped water, and the number of people per room. Housing-Material Quality is defined as the summation of three indicators for high-quality floors, roofs, and walls. Water Access is defined on a scale from 1-5 where 1 is very inaccessible while 5 is very accessible. Clean Water is defined as an indicator for households having treated drinking water (such as by a filter). *Total* Monthly Housing Expenditures is defined as the total of Rent paid, mortgage, and upgrade cost, and *Per capita* Housing Expenditures is divided by household size. They do not include the subsidy payments by the implementing organization. Total ID Cards is defined as the sum of the ID cards possessed by the respondent (MOI card, Passport, Residency permit, Work permit, Family book, Syrian ID, and UNHCR file). Hours spent on childcare and chores are measured over the last week and are winsorized at the top 1% of values in order to limit the influence of outliers.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- Monetary values are in USD PPP.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age).

  In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 2 in the pre-analysis plan.

Table (E3) Impacts of the Program on Consumption and Expenditures

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Total Consumption (Log USD PPP)	-0.03	(0.08)	[1.00]	8.73	(0.55)	1,422
Cereals Consumption (Months)	-0.27	(0.20)	[1.00]	11.72	(1.40)	1,410
Meat Consumption (Months)	-0.67	(0.45)	[1.00]	10.53	(3.05)	1,413
Fish and Seafood Consumption (Months)	0.04	(0.45)	[1.00]	1.88	(3.93)	1,419
Dairy Products Consumption (Months)	-0.27	(0.31)	[1.00]	11.31	(2.49)	1,417
Oils and Fats Consumption (Months)	-0.20	(0.15)	[1.00]	11.89	(0.84)	1,417
Fruits and Nuts Consumption (Months)	0.26	(0.66)	[1.00]	5.13	(5.17)	1,415
Vegetables Consumption(Months)	-0.19	(0.19)	[1.00]	11.72	(1.58)	1,415
Sugar and Desserts Consumption (Months)	0.09	(0.18)	[1.00]	11.76	(1.46)	1,417
Ready-Made Food & Other Consumption (Months)	-0.34	(0.71)	[1.00]	6.07	(5.77)	1,420
Cereals Consumption in Typical Week (USD PPP)	-0.95	(2.30)	[1.00]	23.52	(21.29)	1,385
Meat Consumption in Typical Week (USD PPP)	2.15	(1.90)	[1.00]	15.86	(13.55)	1,390
Fish and Seafood Consumption in Typical Week (USD PPP)	1.50	(0.95)	[1.00]	3.16	(7.50)	1,403
Dairy Products Consumption in Typical Week (USD PPP)	2.42	(1.67)	[1.00]	17.47	(14.11)	1,405
Oils and Fats Consumption in Typical Week (USD PPP)	1.62	(1.44)	[1.00]	10.99	(10.40)	1,398
Fruits and Nuts Consumption in Typical Week (USD PPP)	1.21	(1.11)	[1.00]	6.07	(8.48)	1,371
Vegetables Consumption in Typical Week (USD PPP)	3.23	(2.96)	[1.00]	24.81	(19.67)	1,407
Sugar and Desserts Consumption in Typical Week (USD PPP)	0.40	(0.87)	[1.00]	7.32	(6.93)	1,405
Ready-Made Food & Other Consumption in Typical Week (USD PPP)	0.38	(1.67)	[1.00]	7.00	(12.44)	1,410
Annualized Consumed Food Gifts and Assistance (USD PPP)	-9.16	(20.83)	[1.00]	20.57	(119.90)	1,417
Annual Own Food Production (USD PPP)	1.56	(1.20)	[1.00]	0.37	(4.50)	1,420
Overall Food Consumption Index (Last Week)	0.38	(0.24)	[1.00]	0.00	(1.01)	1,422
Overall Food Consumption Index (Typical Week)	0.10	(0.12)	[1.00]	0.02	(0.99)	1,422
# Meals Eaten by Respondent Yesterday	-0.09	(0.08)	[1.00]	2.04	(0.61)	1,421
Food Diversity	-0.04	(0.11)	[1.00]	7.34	(1.11)	1,422
Reduced Coping Strategy Index	-0.72	(1.52)	[1.00]	13.51	(11.98)	1,422
Infant Needs Expenditures in Last 30 Days (USD PPP)	-9.67	(6.54)	[1.00]	26.44	(42.40)	1,421
Water Expenditures in Last 30 Days (USD PPP)	6.29	(7.25)	[1.00]	42.50	(41.40)	1,415
Hygiene Items Expenditures in Last 30 Days (USD PPP)	-2.74	(3.72)	[1.00]	39.12	(26.11)	1,410
Linens Expenditures in Last 30 Days (USD PPP)	-0.29	(0.86)	[1.00]	0.72	(4.92)	1,418
Clothing Expenditures in Last 30 Days (USD PPP)	13.77*	(7.79)	[1.00]	25.16	(55.91)	1,414
Basic Items Expenditures in Last 30 Days (USD PPP)	-2.97*	(1.66)	[1.00]	4.01	(11.38)	1,417
School Expenditures in Last 30 Days (USD PPP)	-3.71	(5.87)	[1.00]	37.48	(39.93)	1,421
Winter Utilities Expenditures (USD PPP)	11.38	(13.91)	[1.00]	145.52	(82.94)	1,411
Summer Utilities Expenditures (USD PPP)	8.34	(10.19)	[1.00]	82.42	(64.10)	1,412
Non-Food Gifts (USD PPP)	0.32	(2.33)	[1.00]	3.20	(18.07)	1,422
Total Asset Value (USD PPP)	20.06	(311.88)	[1.00]	2,883.37	(1,831.70)	1,422
Durables (USD PPP)	160.01	(117.10)	[1.00]	703.61	(688.86)	1,422
Durable Gifts (USD PPP)	12.58**	(6.08)	[1.00]	5.61	(37.30)	1,422

Notes:

- The table shows the regression results on pre-specified consumption and expenditures using the 2021 in-person data. Each row is its own dependent variable.
- The outcomes that require definitions are: Total Consumption is log of the sum of food consumed in last seven days (Cereals and cereal products, Live animals, meat, and other parts of slaughtered land animals; Fish and other seafood; Milk, other dairy products, and eggs; Oils and fats; Fruits and nuts; Vegetables, tubers, pulses; Sugar and desserts; Ready-made food and other food products (baby food, spices)) multiplied by number of months consumed in last 12 months, home-produced goods multiplied by months produced, annualized food received as gifts and annualized non-food purchases (utilities; water; infant needs; household and hygiene items; linens; clothing and footwear; basic household items; and school costs). Food Gifts and Assistance are annualized by multiplying the monthly estimate by 12 since total value of food consumed that was received as gift (from friends, neighbors) or in-kind assistance is asked in the past 30 days. Annual Own Food Production is defined as the food that the household grew or produced and consume in a typical week multiplied by number of months in which the household consumed food that the household grew or produced. Overall Food Consumption Index is defined a mean effects index of the sum of the aforementioned nine food types, home-produced goods, and assistance last week and then in a typical week. The index is standardized by subtracting the control group's mean and dividing by its standard deviation. Food Diversity is defined as the number of food categories consumed during the past 12 months. Reduced Coping Strategy Index is a measure of food insecurity and is defined following WFP's guidelines from the five coping strategies that the household may have used in the seven days prior to the interview (Rely on less preferred foods [weight 1]; Borrow food, or reply on help from a friend or relative [weight 2]; Limit portion size at meal-times [weight 1]; Restrict consumption by adults in order for small children to eat [weight 3]; Reduce number of meals eaten in a day [weight 1]) and its values vary between 0 and 56. Total Asset Value is defined as the sum of non-food expenses in last 12 months and expenses on durables, which is defined as money spent on cellular phones, televisions, motorized vehicles, air conditioners, refrigerators, propane heaters, water tanks, computers and tablets, livestock, and other durables in the last 12 months.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- Monetary values are in USD PPP and are winsorized at the top 1% of values in order to limit the influence of outliers; durables are winsorized at the top 10% as pre-specified.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 3 in the pre-analysis plan.

Table (E4) Impacts of the Program on Financial Participation

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
At least 30 JD (95 USD PPP) in savings (=1)	0.14***	(0.05)	[0.02]	0.12	(0.32)	1,422
Loans Received (Annual, USD PPP)	30.76	(187.32)	[0.94]	364.33	(1,186.27)	1,422
Interest Rate on Last Loan Taken (percent/month)	82.01	(61.82)	[0.69]	14.13	(86.95)	242
Weighted Interest Rate on Last Loan Given (percent/month)	72.94	(59.93)	[0.69]	14.42	(71.21)	242
Loans Given (Annual, USD PPP)	5.71	(4.46)	[0.69]	1.24	(23.93)	1,422
Last Loan in Default	-0.07	(0.13)	[0.94]	0.70	(0.46)	254
Applied for Loan (=1)	0.11*	(0.07)	[0.56]	0.17	(0.37)	1,422
Applied and Took Loan (=1)	0.06	(0.07)	[0.94]	0.16	(0.36)	1,422
Applied but Did Not Take Loan (=1)	0.06***	(0.02)	[0.00]	0.01	(0.10)	1,422

Notes:

- The table shows the regression results on pre-specified financial participation outcomes using the 2021 in-person data. Each row is its own dependent variable.
- The outcomes that require definitions are: 30 JD Saved? reports an indicator for whether the focus respondent had at least 30 JD in personal savings to draw from in case of an emergency, regardless of whether it is in a bank. Loans received reports the total quantity of formal and informal loans received by the FR in the past year. Interest rate on last loan taken reflects the self reported interest rate on the FR's most recent loan in the past 12 months. FRs were allowed to report interest as dinars/[day/week/month/year] or percent/[day/week/month/year] or as a flat rate (common in Islamic finance). Interest reported in dinars was translated to percent using the most recent loan principal. Interest reported as a flat rate was translated to percent by calculating $\frac{flatinterestrate}{principal}$/loan period. The loan period is calculated as repayment date of the loan minus the survey date. (Note that this is not ideal since we do not have the start date of the loan.) The loan period for individuals in default or those with a negative implied loan period is imputed using the median period among non-zero loan periods corresponding to non-defaulted, flat interest rate loans. Loan had flat rate interest (=1) is an indicator for taking a loan with a flat rate interest amount, and all else (including taking no loans) is zero. Loans Given reports the total quantity of formal and informal loans loaned to others by the FR in the past year. Last Loan in Default =1 the last loan the FR took is in default, and 0 is the FR has taken a loan in the past 12 months and the most recent loan is not in default. Applied for Loan = 1 if the FR applied for a loan in the past 12 months, and 0 otherwise. Applied and Took Loan = 1 if the FR applied for a loan in the past 12 months and took out a loan, and 0 otherwise. Applied but did not Take Loan = 1 if the FR applied for a loan in the past 12 months but did not take a loan, and 0 otherwise. This table departs from the PAP in several ways. First, we removed all questions about community savings groups from the survey since these are not common in this setting. Second, we added a question asking whether respondents have at least 30 JD. Applied for Loan and Applied and Took Loan were also not included in the PAP. Finally, we exclude quantity of bank transfers and interest charged on loans from the focus respondent to other individuals from the table since they are both identically zero in the sample population.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 3 in the pre-analysis plan.

Table (E5) Impacts of the Program on Transfers

Outcomes	**Treatment**	**(se)**	**FDR q-values**	**Control mean**	**(sd)**	**N**
# Senders	0.02	(0.04)	[0.94]	0.06	(0.25)	1,422
Total Received Transfers (Annual, USD PPP)	22.08	(97.98)	[0.94]	75.73	(562.57)	1,421
# Receipients	-0.03	(0.04)	[0.94]	0.06	(0.25)	1,422
Total Transfers Sent (Annual, USD PPP)	51.94	(55.52)	[0.94]	25.61	(219.89)	1,422

Notes:

- The table shows the regression results on pre-specified transfers outcomes using the 2021 in-person data. Each row is its own dependent variable.
- The outcomes that require definitions are: # Senders reports the number of households who have sent the focus respondent a cash or in-kind transfer in the past 12 months. Total received transfers indicates the total cash value of transfers received from these households. # Recipients reports the number of households to whom the focus respondent has sent a cash or in-kind transfer in the past 12 months. Total transfers sent indicates the total cash value of transfers received by these households. Total values are adjusted for 2020 PPP (2021 is not yet available from the world bank).
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 3 in the pre-analysis plan.

Table (E6) Impacts of the Program on Earnings, Labor, & Occupational Choice

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Self-Employment (=1)	-0.02	(0.02)	[0.45]	0.01	(0.12)	1,422
Wage-Employment (=1)	-0.05	(0.06)	[0.54]	0.25	(0.44)	1,422
Employment (=1)	-0.05	(0.06)	[0.54]	0.26	(0.44)	1,422
Total Earnings in Last 30 Days (PPP, IHS)	-0.51	(0.38)	[0.33]	1.59	(2.83)	1,422
Total Labor Hours Last Week	-3.29	(2.35)	[0.33]	7.72	(18.73)	1,422
Total Labor and Chores Hours Last Week	-10.62***	(3.24)	[0.01]	25.54	(23.39)	1,422
Total Labor Hours [Monthly Average]	-0.65	(8.20)	[0.78]	24.19	(67.80)	1,422
Wage-Employment Hours Last Week	-2.48	(2.34)	[0.49]	7.54	(18.45)	1,414
Self-Employment Hours Last Week	-0.64	(0.46)	[0.33]	0.22	(3.30)	1,422
Wage-Employment Adult Labor Hours Last Week	0.01	(4.53)	[0.78]	18.66	(29.85)	1,422
Net Wage-Employment Income Last Month (PPP,IHS)	-0.46	(0.37)	[0.40]	1.56	(2.81)	1,417
Any Seasonal Wage-Employment (=1)	-0.07*	(0.04)	[0.25]	0.07	(0.26)	1,422
Any Manufacturing Employment (=1)	0.05**	(0.02)	[0.21]	0.01	(0.08)	1,137
Any Construction Employment (=1)	-0.12***	(0.05)	[0.11]	0.09	(0.29)	1,150
Any Service Employment (=1)	-0.02	(0.04)	[0.67]	0.09	(0.28)	1,166
Any High-Skill Service Employment (=1)	-0.02	(0.04)	[0.67]	0.08	(0.27)	1,149
Any Low-Skill Service Employment (=1)	0.00	(0.02)	[0.78]	0.01	(0.11)	1,069
Any Trade Employment (=1)	-0.01	(0.01)	[0.52]	0.01	(0.10)	1,062
Any Formal Business (=1)	-0.02**	(0.01)	[0.23]	0.01	(0.08)	1,411
Self-Employment Revenues Last Year (PPP,IHS)	-0.15	(0.11)	[0.33]	0.08	(0.72)	1,419
Any Loss of Business Control (=1)	-0.01	(0.01)	[0.33]	0.00	(0.05)	1,407
Self-Employment Profits Last Year (USD PPP)	-82.56*	(43.83)	[0.25]	28.18	(374.36)	1,419
Job Search (Weeks)	-0.80	(2.69)	[0.67]	5.63	(8.68)	306
Focus Respondent Childcare & Chores (Hours)	-6.91***	(2.02)	[0.01]	17.76	(18.39)	1,418

Notes:

- The table shows the regression results on pre-specified labor market outcomes of the Focus Respondent using the 2021 in-person data. Each row is its own dependent variable.
- The outcomes that require definitions are: Total Earnings is defined as the total of business profits and net wage salary in all jobs and is transformed using the inverse hyperbolic sine transformation. Total Labor Hours [Monthly Average] is defined as average weekly wage-employment hours multiplied by the numbers of months worked in last 12 months multiplied by 52/12 to obtain a monthly average of all jobs. Wage-Employment Adult Labor Hours Last Week is the sum of respondent hours and total hours where adults (other than respondent) in the household were working. Net Wage-Employment Income is defined as gross salary plus benefits minus taxes in all jobs. High-Skill Service Employment refers to employment in the sectors of Utilities, Transportation and warehousing, Health care and social assistance, Arts, entertainment, recreation, and Accommodation and food services. Low-Skill Service Employment refers to employment in the sectors of Wholesale or retail trade, Information, Finance and insurance, Real estate, Professional, scientific, and technical services, Management, Administrative support, and Education.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- Monetary values are in USD PPP. Total Earnings in Last 30 Days, Taxes Paid in Last 30 Days, Net Wage-Employment Income Last Month, Self-Employment Profits, Expenses, and Revenues Last Month and Year, Self-Employment Profits Last Year, Total Employees Last Month, Self-Employment Rent Last Month, and Total Labor Hours [Average] are winsorized at the top 1% of values in order to limit the influence of outliers.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 4 in the pre-analysis plan.

Table (E7) Impacts of the Program on Migration

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
# Moves Since 2011	0.07	(0.20)	[1.00]	2.64	(1.35)	1,417
Stay Permanently in Jordan (=1)	-0.05	(0.06)	[1.00]	0.60	(0.49)	1,375
Plans to Leave MENA (=1)	0.01	(0.08)	[1.00]	0.27	(0.44)	1,415
Better off Now than Syria 2011 (=1)	0.03	(0.06)	[1.00]	0.50	(0.50)	1,419
Better off Now than Last Residence (=1)	0.05	(0.06)	[1.00]	0.78	(0.41)	1,419
Ppl Known in Jordan upon Arrival (SD)	-0.09	(0.17)	[1.00]	0.00	(1.00)	1,420
Ppl Travelled With (SD)	-0.01	(0.14)	[1.00]	-0.00	(1.00)	1,421
Plans to Move? (SD)	-0.16	(0.13)	[1.00]	-0.00	(1.00)	1,416
Prep Time before Flee (SD)	0.21	(0.17)	[1.00]	-0.00	(1.00)	1,416
Plan to Stay in Current Residence (SD)	-0.03	(0.15)	[1.00]	-0.00	(1.00)	808
Conflict End in 2 Years? (SD)	0.29**	(0.15)	[1.00]	0.00	(1.00)	1,209
Return in 2 Years? (SD)	0.12	(0.13)	[1.00]	-0.00	(1.00)	1,403
Return 2 Years After War End? (SD)	-0.02	(0.16)	[1.00]	-0.00	(1.00)	1,393

Notes:

- The table shows the regression results on pre-specified migration outcomes using the 2021 in-person data. Each row is its own dependent variable.
- The outcomes that require definitions are: Variables with (SD) are scaled with respect to the control population and (=1) represents indicators. # moves since 2011 reports the number of times the FR's household moved residences for at least 4 months at a time since January 2011. Stay permanently in Jordan is an indicator where 1 corresponds to the FR responding 'yes' to "Even when the conflict ends, would you like to stay in Jordan permanently?". Plans to leave MENA corresponds to "Have you taken any concrete steps towards moving to the US, Europe, Australia, or somewhere else outside of the region?". People known in Jordan upon arrival is a standardized categorical variable with 4 roughly equally populated bins: 0, 1-3,4-10, 11+. People travelled with is a standardized categorical variable based on "When you first moved to Jordan, how many people travelled with you?" with 4 roughly equally populated bins: 0-3, 4-5,6-10, 11+. Plans to move is a standardized categorical variable based on responses to "Do you have firm plans to change your residence from your current location within the next six months?" Prep time before flee is a standardized categorical variable based on responses to "How long before you left Syria for the first time did you start preparing to move?" Conflict end in 2 years is a standardized variable based on "How likely do you think the conflict will end within the next two years? " Return in 2 years is a standardized variable based on "How likely are you to return to Syria in the next two years if the conflict is unresolved?" Return in 2 years after war ends is a standardized variable based on "How likely are you to return to Syria within two years of the conflict resolving?"
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 5 in the pre-analysis plan.

Table (E8) Impacts of the Program on Physical Health

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
FR Normalized Symptom Index (Last 4 Weeks)	0.26	(0.18)	[0.44]	0.00	(1.00)	1,422
FR Subjective Health (Scale 1-4)	-0.21*	(0.11)	[0.38]	2.98	(0.73)	1,422
FR Major Health Problems since Jan 2011 (=1)	0.14**	(0.06)	[0.29]	0.22	(0.41)	1,422
FR Major Health Problems Persistence (=1)	-0.12	(0.09)	[0.61]	0.74	(0.44)	333
FR Washington Group Disability (=1)	-0.02	(0.06)	[1.00]	0.25	(0.43)	1,422
FR Hospital Visits, Last 4 Weeks	-0.08	(0.21)	[1.00]	0.86	(1.33)	1,422
FR Foregone Hospital Visits, Last 4 Weeks	1.08**	(0.54)	[0.38]	2.74	(3.83)	1,416
FR Health Expenditures (USD PPP), Last 4 Weeks	-10.17	(19.42)	[1.00]	65.77	(131.06)	1,422
FR Food Insecurity Last Week (=1)	0.11*	(0.07)	[0.38]	0.34	(0.47)	1,421
Adult Food Insecurity (Days)	0.05	(0.22)	[1.00]	0.78	(1.47)	1,391

Notes:

- The table shows the regression results on pre-specified physical health outcomes using the 2021 in-person data. Each row is its own dependent variable.
- The outcomes that require definitions are: Normalized Symptom Index is defined as mean-effect index of common symptoms (Fever, Persistent cough, Always feeling tired, Muscle pain (myalgia), Headache/migraine, Stomach pain, Blood in stool, Rapid weight loss, Open sores/boils, Diarrhea/nausea/vomiting, Back pain or other muscle pain, Runny nose, Sore throat, Pneumonia, Loss of sense of smell/not being able to taste food, Frequent and excessive urination, Constant thirst/increased drinking of fluids, Skin rash or irritation, Difficulty swallowing, Fast or irregular heartbeat, Difficulty breathing/chest tightness). FR Subjective Health is based on the FR's describtion of his or her general health and is categorized as good (4), fair (3), poor (2), or very poor (1). FR Major Health Problems Persistence is an indicator variable that equals unity if the major health problems have not been resolved. FR Washington Group (WG) Disability is defined following the WG Short Set of questions to identify functional limitations on the levels of vision, hearing, mobility, cognition, self care, and communication. It uses the third disability level to classify people as having functional limitations or not. FR Health Expenditures is in USD PPP and is winsorized at the top 1% of values in order to limit the influence of outliers. FR Food Insecurity is defined as an indicator variable that equals unity if FR slept hungry because there was not enough food on at least one day the last 7 days and zero otherwise.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 6 in the pre-analysis plan.

Table (E9) Impacts of the Program on Mental Health

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
FR Normalized CESD Scale (Last Week)	-0.07	(0.16)	[1.00]	0.00	(1.00)	1,396
FR Normalized PSS Scale (Last 30 days)	-0.10	(0.14)	[1.00]	-0.00	(1.00)	1,420
FR Life Satisfaction (Scale 1-10)	-0.64*	(0.37)	[0.38]	5.16	(2.81)	1,420
FR Predeterminsim (Scale 1-10)	0.53	(0.47)	[0.68]	4.22	(3.42)	1,412
FR Grit (Scale 1-5)	0.02	(0.13)	[1.00]	3.50	(0.89)	1,410
FR Happiness (Scale 1-7)	-0.12	(0.27)	[1.00]	3.41	(1.93)	1,421
FR Alertness (Scale 1-7)	-0.21	(0.36)	[1.00]	3.08	(1.99)	1,421

Notes:

- The table shows the regression results on pre-specified mental health outcomes using the 2021 in-person data. Each row is its own dependent variable.
- The outcomes that require definitions are: Center for Epidemiologic Studies Depression Scale (CES-D) is a measure for depressive symptoms and includes 10 items asking about the past week. It is defined as a mean-effect index of 10 items scored 0-3 in the past week (I was bothered by things that usually don't bother me, I had a problem in concentration on what I was doing, I felt depressed and troubled in my mind, I felt that everything that I did took up all my energy, I felt hopeful about the future (reversed), I felt afraid, I had difficulty in sleeping peacefully, I was happy (reversed), I felt lonely, I lacked the motivation to do anything). It is then normalized by subtracting the control group's mean and dividing by its standard deviation. The Perceived Stress Scale (PSS-4) is a mean-effect index of four items measured over 1-5 in the last 30 days (how often have you felt that you were unable to control the important things in your life?, how often have you felt certain in your ability to overcome your own personal problems?, how often have you felt that things were going your way?, how often did you feel that the problems were too much for you to manage?). It is then normalized by subtracting the control group's mean and dividing by its standard deviation. Higher scores for CES-D and PSS indicate better outcomes. Predeterminsim is measured over 1-10 where 1 means "everything in life is determined by fate" and 10 means "people shape their fate themselves". Grit is the average of "I am a hard worker" and "I often set a goal but later choose to pursue a different one" where both are measured over 1-5 and the latter is reversed to indicate higher grit.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 6 in the pre-analysis plan.

Table (E10) Impacts of the Program on Sleep

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Sleep Time	-4.08***	(1.49)	[0.15]	15.74	(9.71)	1,414
Wake up Time	0.14	(0.33)	[1.00]	7.32	(2.00)	1,414
Sleep Duration (Hours)	-0.16	(0.30)	[1.00]	7.36	(2.15)	1,414
Sleep Quality (1-5)	-0.03	(0.18)	[1.00]	3.04	(1.03)	1,422

Notes:

- The table shows the regression results on pre-specified physical and mental health outcomes using the 2021 in-person data. Each row is its own dependent variable.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 6 in the pre-analysis plan.

Table (E11) Impacts of the Program on Marriage and Fertility

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Marriage quality	0.28**	(0.13)	[0.29]	-0.23	(1.06)	1,235
Index across household decisions	-0.02	(0.12)	[1.00]	0.08	(0.85)	1,194
Age at first marriage	-1.34**	(0.64)	[0.29]	21.27	(5.22)	1,372
Marriage registered (=1)	-0.00	(0.02)	[1.00]	0.98	(0.13)	1,235
Number of marriages (polyg. separate)	0.02	(0.05)	[1.00]	1.04	(0.36)	1,422
Number of marraiges after first	-0.01	(0.04)	[1.00]	0.07	(0.30)	1,422
Wife received prompt dowry (=1)	0.01	(0.02)	[0.94]	0.98	(0.13)	1,182
Age gap between respondend and spouse	1.63*	(0.86)	[0.34]	0.38	(8.36)	1,225
Age at marriage end if divorced/widowed	5.78	(4.95)	[0.94]	36.81	(13.08)	142
Current marriage polygamous (=1)	0.03	(0.04)	[0.94]	0.07	(0.26)	1,245
Respondent selected current spouse (=1)	-0.01	(0.07)	[1.00]	0.78	(0.42)	1,226
Number of pregnancies	-0.45	(0.53)	[0.94]	5.29	(2.89)	813
Number of any partner's pregnancies	1.16***	(0.38)	[0.05]	5.16	(2.90)	606
Used modern birth control with current spouse (=1)	-0.09	(0.09)	[0.94]	0.60	(0.49)	1,231
Number of childred died age 0-5	0.01	(0.06)	[1.00]	0.09	(0.35)	1,422
Age at menarche	-0.46	(0.30)	[0.55]	13.62	(1.22)	815
Number of miscarriages and stillbirths	-0.12	(0.15)	[0.94]	0.46	(1.00)	1,422
Number of terminated pregnancies	0.09	(0.12)	[0.94]	0.33	(0.81)	1,422
Number of births in hospital	0.37	(0.25)	[0.55]	3.49	(2.24)	1,422
Number of babies born preterm	0.05	(0.09)	[1.00]	0.19	(0.59)	1,422

Notes:

- The table shows the regression results on pre-specified financial participation outcomes using the 2021 in-person data. Each row is its own dependent variable.
- The outcomes that require definitions are: Marriage quality is a index including spousal education and age gap between respondent and spouse. Index across household decisions is an index across various types of household decisions equalling 1 if the respondent has some decision making power. Age at first marriage is the age at which the respondent was first married. Marriage registered is an indicator equal to 1 if the marriage is registered in a court. Number of marriages reports the number of times the respondent has been married, where polygamous marriages count separately. Number of marriages after first reports the number of time the respondent has been remarried. Wife received prompt dowry is an indicator equal to 1 if the wife received a mahr muqaddam amount greater than zero. Number of pregnancies reports the number of pregnancies experienced by female respondents. Number of any partner's pregnancies reports the number of pregnancies experienced by partners of male respondents. Number of children who died reports the number of biological children of the respondent who passed away between the ages of 0-5. Age at menarche reports the age at which female respondents began menstruating. Number of miscarriages and stillbirths reports the number of pregnancies ending in stillbirth or miscarriage for both male and female respondents. Number of terminated pregnancies reports the number of pregnancies ending in abortion for both male and female respondents. Number of births in hospital reports the number of children delivered in a hospital (opposed to at home). Number of babies born preterm reports the number of biological children born preterm for both male and female respondents. Value of dowry paid is included in the pre-analysis plan but excluded here since the number of new marriages since 2019 was very small and adjusting Syrian Lira for purchasing power parity during this time is unreliable.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 3 in the pre-analysis plan.

Table (E12) Impacts of the Program on Individual Child Outcomes

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Strengths and Difficulties Questionnaire (Z-Score)	-0.16	(0.18)	[1.00]	0.00	(1.00)	905
Child Happiness (Scale 1-7)	-0.16	(0.31)	[1.00]	5.73	(1.54)	951
Child Alertness (Scale 1-7)	0.04	(0.25)	[1.00]	5.72	(1.64)	951
Child Sleep Duration (Hours)	-0.13	(0.24)	[1.00]	9.79	(1.49)	939
Child Bed Time (Military)	2.02	(1.70)	[1.00]	17.26	(8.65)	948
Child Wake Up Time (Military)	-0.06	(0.33)	[1.00]	8.20	(1.61)	943

Notes:

- The table shows the regression results on pre-specified child outcomes using the 2021 in-person data. Each row is its own dependent variable.
- The outcomes that require definitions are: Strengths and Difficulties Questionnaire (SDQ) is a screening questionnaire administered to a randomly selected child aged three to eight years old and covers emotional, conduct, hyperactivity, peer, and prosocial problems. SDQ is standardized by subtracting the control group's mean and dividing by its standard deviation.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 8 in the pre-analysis plan.

Table (E13) Impacts of the Program on Household-Level Child Outcomes

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Count of Non-Adult Dependents in Household	-0.07	(0.18)	[1.00]	3.30	(2.02)	1,422
Student Child School Attendance in Household (=1)	0.18	(0.12)	[1.00]	0.51	(0.50)	760
Student Child School Attendance in Household (0-5)	-0.13	(0.30)	[1.00]	3.80	(1.29)	760
Child Hunger in Household (Days Last Week)	0.12	(0.18)	[1.00]	0.47	(1.15)	1,269
Any Child Vaccinated (=1)	0.09	(0.12)	[1.00]	0.84	(0.37)	509

Notes:

- The table shows the regression results on pre-specified household-level child outcomes using the 2021 in-person data. Each row is its own dependent variable.
- The outcomes that require definitions are: Child Hunger is defined as the number of days in the last 7 days on which any children in the household had to go to sleep hungry because there was not enough food.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 8 in the pre-analysis plan.

Table (E14) Impacts of the Program on Social Capital

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Crime Index (Extensive Margin)	0.05	(0.16)	[1.00]	-0.00	(1.00)	1,422
Crime Index (Intensive Margin)	-0.06	(0.12)	[1.00]	0.00	(1.00)	1,422
Security Compared to 2011 (Z-Score)	-0.03	(0.13)	[1.00]	0.00	(1.00)	1,419
Arrested in Syria (=1)	-0.03	(0.04)	[1.00]	0.06	(0.24)	1,421
Imprisoned in Syria (=1)	-0.06**	(0.03)	[0.38]	0.05	(0.21)	1,422
Children Share Spaces with Jordanians (=1)	-0.07	(0.07)	[1.00]	0.50	(0.50)	1,345
Languages Spoken	0.04	(0.03)	[0.74]	1.01	(0.12)	1,422
Languages Learned	-0.01	(0.02)	[1.00]	0.08	(0.29)	1,422
Gender Equality Index (Z-Score)	-0.11	(0.12)	[1.00]	0.00	(1.00)	1,415
Religiosity Change in Community (Z-Score)	-0.13	(0.15)	[1.00]	0.00	(1.00)	1,378
Religion Importance (Z-Score)	0.11	(0.14)	[1.00]	0.00	(1.00)	1,422
Mosque Donations Last Month (=1)	-0.12**	(0.05)	[0.38]	0.12	(0.32)	1,418
Religious Identity Salience (=1)	-0.06	(0.05)	[0.74]	0.84	(0.37)	1,419
Attended Mosque Last Week (=1)	-0.06	(0.04)	[0.74]	0.34	(0.47)	1,422
Donated Time to Charities Last Month (=1)	-0.01	(0.02)	[1.00]	0.02	(0.13)	1,422
Economic Optimism (Z-Score)	-0.38**	(0.19)	[0.38]	0.00	(1.00)	995

Notes:

- The table shows the regression results on pre-specified social capital outcomes using the 2021 in-person data. Each row is its own dependent variable.
- The outcomes that require definitions are: Crime Index (Extensive Margin) is defined as a normalized index from two indicators for having been stolen from or gone through an attempt to steal any cash, household items, or livestock from the FR and another indicator if someone physically assaulted the FR last 12 months. Crime Index (Intensive Margin) is defined the same as Crime Index (Extensive Margin) but the number of times if robbery and assault is used instead of the indicators. Security Compared to 2011 (Z-Score) is defined as a normalized index from FR's answer to whether the security better or same or worse for him/her now compared to January 2011 (referring to the place and circumstances of his/her life at that time). Gender Equality Index is defined as a normalized index constructed from four items in which the FR strongly agrees, agrees, disagrees, or strongly disagrees. The four items are "A married woman can work outside the home if she wishes", "Husbands should have final say in all decisions concerning the family", "A woman can be a president or prime minister of a Muslim country", and "Women and men should have equal rights in making the decision to divorce." Each item is normalized and then the resulting index is normalized again where higher values indicate views in accordance with more gender equality. Religiosity Change in Community (Z-Score) is defined as a normalized index from the FR's responses to whether he or she would say their community has become more religious, stayed the same or become less religious in the last 12 months, where higher values indicate increasing religiosity. Religion Importance (Z-Score) is defined as a normalized index from the FR's response to whether religion is very important, somewhat important or not very important to the lives of most of their neighbors, where higher values indicate more importance of religion. Religious Identity Salience is defined as an indicator variables that equals unity if the FR chooses "Above all I am a Muslim" among other identities that include Syrian, Arab, or Christian. Economic Optimism (Z-Score) is defined as an index of whether in two years from now, the FR things that their own personal economic situation will be the same, better, or worse, with higher values indicate better economic situation.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 9 in the pre-analysis plan.

Table (E15) Impacts of the Program on Political Attitudes

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Support for Democracy (Z-Score)	0.15	(0.17)	[0.84]	0.00	(1.00)	837
Support for Government (Z-Score)	0.16	(0.20)	[0.84]	-0.00	(1.00)	1,056
Justify Human Rights Violations (Z-Score)	0.19	(0.13)	[0.84]	0.00	(1.00)	1,084
Interest in Politics (Z-Score)	0.16	(0.18)	[0.84]	-0.00	(1.00)	1,413
Followed News Last Week (Days)	0.17	(0.50)	[0.84]	2.26	(2.87)	1,422
Missing Support for Democracy (=1)	0.09	(0.06)	[0.84]	0.39	(0.49)	1,422
Missing Support for Government (=1)	-0.06	(0.05)	[0.84]	0.26	(0.44)	1,422
Missing Justify Human Rights Violations (=1)	-0.08	(0.07)	[0.84]	0.25	(0.44)	1,422

Notes:

- The table shows the regression results on pre-specified political attitudes outcomes using the 2021 in-person data. Each row is its own dependent variable.
- The outcomes that require definitions are: Support for Democracy (Z-Score) is defined as a z-score from whether the FR strongly agrees, agrees, disagrees, or strongly disagrees with "Democratic systems may have problems, yet they are better than other systems." Higher values indicate more support for democracy. Support for Government (Z-Score) is defined as a z-score from whether the FR strongly agrees, agrees, disagrees, or strongly disagrees with "Citizens must support the government's decisions, even if they disagree with them." Higher values indicate more support for the government. Justify Human Rights Violations (Z-Score) is defined as a z-score from whether the FR strongly agrees, agrees, disagrees, or strongly disagrees with "The violation of human rights in Syria is justifiable in the name of promoting security and stability." Higher values indicate more justification for human rights violations. Interest in Politics (Z-Score) is defined as a z-score from the extent to which the FR is interested in politics (very uninterested, uninterested, interested, very interested), with higher values indicating more interest in politics.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 10 in the pre-analysis plan.

Table (E16) Impacts of the Program on Time Use

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
FR Childcare & Chores (Hours Last Week)	-6.91***	(2.02)	[0.00]	17.76	(18.39)	1,418
Others Childcare & Chores (Hours Last Week)	-0.50	(2.15)	[0.69]	18.54	(18.49)	1,390

Notes:

- The table shows the regression results on pre-specified time use outcomes using the 2021 in-person data. Each row is its own dependent variable.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 11 in the pre-analysis plan.

Table (E17) Impacts of the Program on Education and Cognition

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
FR Years of Schooling	0.39	(0.71)	[1.00]	6.49	(3.97)	1,422
FR Private Education (=1)	0.02	(0.02)	[1.00]	0.01	(0.11)	1,422
FR Religious Education (=1)	0.00	(0.01)	[1.00]	0.01	(0.09)	1,422
FR Fluid Intelligence (Z-Score)	-0.09	(0.37)	[1.00]	7.42	(2.94)	1,422

Notes:

- The table shows the regression results on pre-specified education and cognition outcomes using the 2021 in-person data. Each row is its own dependent variable.
- The outcomes that require definitions are: FR Years of Schooling is calculated as follows: first, because few FRs reported total years of schooling rather than the marginal years at their latest stage, each stage is capped at its correct upper bound (10 years for primary in Jordan; 9 years for primary in other countries; 2 years for secondary in Jordan; 3 years for secondary in other countries; 3 years for college; 8 years for university). After that, those who never attended school are assigned zero years. As for those who indicated attending a religious institution, they are assigned zero years as well. FRs with Primary educational attainment are assigned the total years in which they attended primary school. FRs with secondary educational attainment in Jordan (non-Jordan) are assigned 10 (9) plus the total years in which they attended secondary school. FRs with a college degree are assigned 12 plus the total years in which they attended college. FRs with vocational training are assigned 10 plus the total years in which they attended vocational training. FRs with a university degree are assigned 12 plus the total years in which they attended university. FR Private Education is defined an indicator variable that equals unity if the FR attended a private school/university in his or her last year of study. FR Religious Education is defined an indicator variable that equals unity if the FR attended a religious school/university at any point in the past. FR Fluid Intelligence is defined as the sum of correct answers to 14 questions to Raven's Short form Standard Progressive Matrices (RSPM-SF). Then, the total score is normalized by the control group's mean and standard deviation such that the final score has a mean of zero and standard deviation of one in the control group.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 12 in the pre-analysis plan.

Table (E18) Impacts of the Program on Behavioral Games and Preferences

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Risk Taking Preferences (Z-Score)	-0.14	(0.11)	[1.00]	0.00	(1.00)	1,401
Impatience Preferences (Z-Score)	-0.06	(0.16)	[1.00]	-0.00	(1.00)	1,415
Ambiguity Aversion (=1)	-0.01	(0.04)	[1.00]	0.87	(0.33)	1,403
$ shared with a Syrian refugee living in Jordan	0.18	(0.40)	[1.00]	4.26	(2.84)	1,411
$ shared with a non-Syrian refugee living in Jordan	0.40	(0.36)	[1.00]	3.73	(2.80)	1,404
$ shared with a Jordanian citizen from same ethnicity	-0.25	(0.39)	[1.00]	4.06	(2.83)	1,411
$ shared with a Jordanian citizen from different ethnicity	-0.20	(0.43)	[1.00]	3.70	(2.84)	1,406
$ shared with a Jordanian soldier/police officer	-0.06	(0.49)	[1.00]	4.32	(3.23)	1,403
$ shared with an employee of the Jordanian government	-0.11	(0.46)	[1.00]	3.37	(2.98)	1,406
$ shared with an employee of an international NGO	0.37	(0.40)	[1.00]	3.00	(2.83)	1,402

Notes:

- The table shows the regression results on pre-specified behavioral outcomes using the 2021 in-person data. Each row is its own dependent variable.
- The outcomes that require definitions are: Risk Taking Preferences is defined as in the Jordan version of the Global Preferences Survey. The quantitative measure consists of a series of five binary choices. Choices were between a fixed lottery, in which the individual could win a certain amount or zero, and varying sure payments. Choice of the lottery resulted in an increase of the sure amount being offered in the next question, and vice versa. The total score is normalized by the control group's mean and standard deviation such that the final score has a mean of zero and standard deviation of one in the control group. Higher values indicate a tendency to take more risks. Impatience Preferences is defined as in the Jordan version of the Global Preferences Survey. It includes a series of five interdependent hypothetical binary choices between immediate and delayed financial rewards. The total score is normalized by the control group's mean and standard deviation such that the final score has a mean of zero and standard deviation of one in the control group. Higher values indicate a tendency to be impatient. Ambiguity aversion is defined using a modified, hypothetical version of the standard ambiguity aversion game whereby the respondent selects either a bag with a specified probability of success or an unspecified probability. This measure is an indicator variable that equals unity if respondent selected from the bag of known ball colors and zero if selected from the bag of unknown colors. Trust outcomes are measured using a modified, hypothetical version of the standard trust game where respondent is given 10 JOD to share with members of 7 different social groups (or keep). The amount is doubled and members select how much to share back. The 0 - 10 scale varies between 0 denoting no trust and 10 denoting full trust. Individual outcomes derived for each social group. The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 13 in the pre-analysis plan.

Table (E19) Impacts of the Program on Additional Dwelling Characteristics and Household Composition

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Monthly Rent Agreed (USD PPP)	-28.15*	(16.82)	[0.96]	336.55	(117.73)	1,272
Housing Upgrade Costs Last Month (USD PPP)	-26.88**	(12.09)	[0.62]	31.83	(88.28)	1,417
Total Rental Debt (USD PPP)	-190.59	(129.13)	[1.00]	805.85	(861.00)	1,272
Land Owned in 2011 (Donum)	4.75	(11.96)	[1.00]	14.26	(78.61)	1,401
Land Owned in Jordan (Donum)	-1.07	(1.03)	[1.00]	0.70	(18.45)	1,420
Dwelling Upgraded by Landlord (=1)	0.01	(0.05)	[1.00]	0.09	(0.29)	1,383
Eviction (=1)	-0.00	(0.06)	[1.00]	0.15	(0.36)	1,422
Eviciton (#)	0.07	(0.14)	[1.00]	0.27	(0.76)	1,422
Quality Floor (=1)	0.13**	(0.06)	[0.62]	0.66	(0.47)	1,421
Quality Roof (=1)	0.10*	(0.05)	[0.96]	0.75	(0.44)	1,421
Quality Walls (=1)	0.07*	(0.04)	[0.96]	0.86	(0.35)	1,422
People per Room	-0.22	(0.16)	[1.00]	1.88	(1.29)	1,422
Total Debt Waived| Agreed $>$ Paid (=1)	0.01	(0.09)	[1.00]	0.39	(0.49)	530
Total Debt Waived (=1)	0.04	(0.05)	[1.00]	0.16	(0.37)	1,239
Total Debt Waived $\mid$ Agreed $>$ Paid (USD PPP)	-2.31	(31.50)	[1.00]	84.57	(142.73)	530
Total Debt Waived (USD PPP)	16.61	(15.19)	[1.00]	34.45	(98.95)	1,239
Adults	-0.14	(0.18)	[1.00]	2.65	(1.32)	1,422
Kids	-0.07	(0.18)	[1.00]	3.30	(2.02)	1,422
Men	-0.09	(0.12)	[1.00]	1.24	(0.87)	1,422
Women	-0.05	(0.12)	[1.00]	1.41	(0.81)	1,422
Kids Below 13	-0.05	(0.19)	[1.00]	2.43	(1.71)	1,422
Kids (13-17)	0.03	(0.11)	[1.00]	0.71	(0.90)	1,422
Boys	0.10	(0.17)	[1.00]	1.70	(1.34)	1,422
Boys Below 13	0.05	(0.16)	[1.00]	1.28	(1.13)	1,422
Boys (13-17)	0.04	(0.09)	[1.00]	0.35	(0.63)	1,422
Girls	-0.17	(0.19)	[1.00]	1.60	(1.45)	1,422
Girls Below 13	-0.10	(0.17)	[1.00]	1.16	(1.23)	1,422
Girls (13-17)	-0.02	(0.09)	[1.00]	0.36	(0.61)	1,422

Notes:

- The table shows the regression results on housing quality and housing-related finances that were not pre-specified and on individual items of the indices reported in Table E2 using the 2021 in-person data. Each row is its own dependent variable.
- The outcomes that require definitions are: Quality floor is defined as an indicator variables that equals unity if floors are made of tiles and zero otherwise. Quality roof is defined as an indicator variables that equals unity if the roof is made of finished concrete or tiles/standard bricks on roof and zero otherwise. Quality walls is defined as an indicator variables that equals unity if walls are made of cement or tiles and zero otherwise.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- Monetary values are in USD PPP. Eviction and monetary values are winsorized at the top 1% of values in order to limit the influence of outliers.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 2 in the pre-analysis plan.

Table (E20) Impacts of the Program on NGO Assistance

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Government, Last 12 Months (=1)	-0.03	(0.05)	[0.40]	0.10	(0.30)	1,422
UNHCR, Last 12 Months (=1)	0.14*	(0.08)	[0.33]	0.64	(0.48)	1,422
WFP, Last 12 Months (=1)	-0.06	(0.04)	[0.33]	0.94	(0.24)	1,422
IP Assistance, Last 12 Months (=1)	0.08**	(0.03)	[0.14]	0.04	(0.19)	1,422
Community-Groups, Last 12 Months (=1)	-0.00	(0.03)	[0.51]	0.04	(0.20)	1,422
Other, Last 12 Months (=1)	0.06*	(0.04)	[0.33]	0.06	(0.24)	1,422
Total Cash, Last 12 Months (USD PPP)	706.91	(586.39)	[0.40]	4,431.51	(4,004.01)	1,422
Total In-Kind, Last 12 Months (USD PPP)	92.67	(67.83)	[0.33]	127.02	(481.35)	1,422
Total, Last 12 Months (USD PPP)	788.26	(562.86)	[0.33]	4,578.48	(4,020.73)	1,422
Government, Last 12 Months (USD PPP)	111.04	(96.21)	[0.40]	169.37	(792.44)	1,422
UNHCR, Last 12 Months (USD PPP)	420.89	(305.03)	[0.33]	1,655.25	(2,226.52)	1,422
WFP, Last 12 Monthse (USD PPP)	74.02	(293.95)	[0.51]	2,618.38	(2,404.40)	1,422
IP Assistance, Last 12 Months (USD PPP)	128.25***	(46.50)	[0.10]	28.74	(198.87)	1,422
Community-Groups, Last 12 Months (USD PPP)	-2.58	(10.47)	[0.51]	10.17	(70.04)	1,422
Other, Last 12 Months (USD PPP)	60.20*	(36.76)	[0.33]	44.94	(231.61)	1,422

Notes:

- The table shows the regression results on financial assistance using the 2021 in-person data. Each row is its own dependent variable. These outcomes were not pre-specified as they were added after the first-wave phone survey data were collected. UNHCR refers to the United Nations High Commissioner for Refugees; WFP refers to the World Food Programme; IP refers to the implementing partner.
- The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment.
- All values are in USD PPP and are winsorized at the top 1% of values in order to limit the influence of outliers.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to the outcomes in the table.

3 Appendix F: Follow-Up Results

Table (F1) Impacts of the Program on Primary Outcomes

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Housing Quality Index	0.09	(0.19)	[0.47]	-0.01	(0.99)	1,321
Log total consumption (USD PPP)	-0.05	(0.10)	[0.47]	4.54	(0.78)	1,316
Depression (CESD, std)	0.26*	(0.15)	[0.13]	-0.01	(1.00)	1,285
Child SDQ score, std	-0.56***	(0.17)	[0.01]	-0.01	(1.00)	859

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022-2023 phone survey data. Each row is its own dependent variable.
- Housing quality index is defined in table 2. Total Consumption is defined in table 3. Depression is defined in table 7 (higher here reflects more depression). Child SDQ score is defined in table 8 (higher here reflects more socioemotional difficulties).
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 1 in the pre-analysis plan.

Table (F2) Impacts of the Program on Dwelling Characteristics and Household Composition

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Housing quality index	0.09	(0.19)	[1.00]	-0.01	(0.99)	1,321
Housing Material quality index	0.22	(0.22)	[1.00]	-0.14	(1.08)	1,321
Has access to electricity {0,1}	-0.02	(0.02)	[1.00]	0.94	(0.24)	1,321
Has access to grid (if electricity=1) {0,1}	0.02	(0.07)	[1.00]	0.77	(0.42)	1,321
Has access to generator (if electricity=1) {0,1}	0.01	(0.01)	[1.00]	0.01	(0.09)	1,321
Number of rooms in dwelling	-0.26	(0.16)	[1.00]	3.47	(1.06)	1,321
Number of occupied rooms	-0.19	(0.17)	[1.00]	3.37	(1.08)	1,321
Household size	-0.38	(0.29)	[1.00]	5.95	(2.30)	1,321
Owns house {0,1}	-0.00	(0.03)	[1.00]	0.04	(0.20)	1,321
Change in rental debt (JD, 30 days)	13.31	(14.15)	[1.00]	44.35	(89.54)	1,187
Outstanding rental debt (JD, 12 months)	-77.44*	(41.23)	[1.00]	315.20	(274.57)	1,208
Moved between dwellings ({0,1}, 12 months)	-0.03	(0.06)	[1.00]	0.19	(0.39)	1,321
Number of moves	-0.02	(0.10)	[1.00]	0.25	(0.62)	1,321
Evicted {0,1}	0.03	(0.04)	[1.00]	0.13	(0.34)	1,321
Number of evictions	0.21**	(0.10)	[1.00]	0.19	(0.57)	1,321
Shares dwelling {0,1}	0.00	(0.01)	[1.00]	0.00	(0.05)	1,321
Number of families (if shares dwelling=1)	0.00	(0.01)	[1.00]	0.00	(0.05)	1,321

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022-2023 phone survey data. Each row is its own dependent variable.
- Housing quality index is an standardized index of the following: Indicator for high-quality floors, Indicator for high-quality roof, Indicator for high-quality walls, Indicator for reliable electricity, people per room. Housing material quality index is a standardized index of indicators for high quality floors, roof and walls. Has access to grid is an indicator for having access to grid electricity, zero if no electricity or from a generator, etc. Generator is an indicator equal to one if the household uses a generator for electricity, zero otherwise. Number of rooms in dwelling is the total number of separate rooms excluding toilets and store rooms, including rooms shared with other households if applicable. Number of occupied rooms only includes rooms occupied by the household. Household size is the number of individuals of any age in the household. Owns house is an indicator equal to one if the household owns their dwelling. Change in rental debt is the total amount of rental debt accumulated minus total amount of rental debt repaid in the past month. Outstanding rental debt is the total amount of unpaid rent from all months that the landlord expect to be repaid. Moved between dwellings is an indicator if the household has moved from one dwelling to another in the past 12 months. Evicted is an indicator if the household has been evicted in the past 12 months. Number of evictions is the number of evictions in the past 12 months, including zeroes. Shared dwelling is an indicator if the household shares the dwelling with another household. Number of families reports the number of extra families living in the household.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family xxx in the pre-analysis plan.

Table (F3) Impacts of the Program on Food Consumption and Food Security

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Total Consumption (Log USD PPP)	-0.05	(0.10)	[1.00]	4.54	(0.78)	1,316
Out-of-pocket Consumption (USD PPP)	-6.18	(11.63)	[1.00]	120.53	(87.30)	1,320
Assistance Consumption (USD PPP)	0.15	(0.53)	[1.00]	0.76	(2.85)	1,308
Home Production Consumption (USD PPP)	-0.65	(1.65)	[1.00]	2.34	(11.36)	1,321
Meat Consumption	0.08	(0.83)	[1.00]	3.38	(5.13)	1,286
# Meals Eaten by Respondent Yesterday	-0.18**	(0.08)	[0.12]	2.04	(0.61)	1,321
Respondent Hunger Last Week (Days)	0.59**	(0.28)	[0.12]	1.30	(1.72)	1,277
Adult Hunger Last Week (Days)	0.56**	(0.26)	[0.12]	1.06	(1.61)	1,233
Food Diversity	-0.03	(0.26)	[1.00]	5.35	(1.82)	1,313
Reduced Coping Strategy Index	0.04	(0.14)	[1.00]	-0.00	(0.99)	1,283

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022-2023 phone survey data. Each row is its own dependent variable.
- Log total consumption is the log of the sum of the following weekly variables: total food consumption from purchases, total food consumption from home production, and total food consumption from assistance. The individual components are listed in the 3 rows below. Meat consumption reports the number of JD spent on meat, fish and other seafood in the last 7 days. Meal consumption reports the number of meals FR ate yesterday. Respondent hunger (days) reports the number of days in the past 7 days that the FR went to bed hungry because there wasn't enough food to eat. Adult hunger (days) reports the number of days in the past 7 days that other adults in the household went to bed hungry because there wasn't enough food to eat. Food diversity is the total number of food categories (out of 7) that the household spent any money on in the past 7 days. Reduced coping strategies index is the world food programme's reduced coping strategies index, defined as follows: 1*(days in last week relied on less preferred foods) + 2*(borrowed food) + 1*(limited portion size) + 3*(restrict adult consumption so children can eat) + 1*(reduce number of meals eat in a day)
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family xxx in the pre-analysis plan.

Table (F4) Impacts of the Program on Financial Participation

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Has savings $\geq$ 30 JD	0.01	(0.04)	[1.00]	0.09	(0.28)	1,321
Lent informally last month {0,1}	0.03	(0.02)	[1.00]	0.02	(0.14)	1,321
Lent informally last month (JD)	0.16	(0.53)	[1.00]	0.43	(3.48)	1,321
Borrowed informally last month {0,1}	-0.00	(0.06)	[1.00]	0.72	(0.45)	1,321
Borrowed informally last month (JD)	-34.49	(46.00)	[1.00]	229.83	(359.22)	1,318
Borrowed or lent formally last year {0,1}	0.02	(0.04)	[1.00]	0.07	(0.25)	1,321

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022-2023 phone survey data. Each row is its own dependent variable.
- Has savings $\geq$ 30 JD is an indicator for whether the household has more than 30 JD saved (in a bank or otherwise). Lent informally last month is an indicator for whether FR lent assistance to someone outside the household in the form of money or goods with the expectation of being paid back in money, goods or favors. Lent informally (JD) is the amount of this informal lenting in JD, including zeros. Borrowed informally in the past month is defined as above for borrowing, not lending. Borrowed or lent formally is an indicator for whether or not FR took out a formal loan (e.g. commercial bank, commercial lender mobile service provider, etc.) in the past 12 months.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family xxx in the pre-analysis plan.

Table (F5) Impacts of the Program on Earnings, Labor, and Occupational Choice

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Earnings (JD, 30 days, IHS transform)	-0.52	(0.33)	[0.30]	1.81	(2.38)	1,321
Earnings (JD, 30 days, log transform)	0.01	(0.25)	[0.64]	3.90	(1.25)	527
Had any income generating work ({0,1}, 30 days)	-0.14*	(0.07)	[0.30]	0.40	(0.49)	1,321
Hours worked in typical week (30 days)	-6.11*	(3.41)	[0.30]	12.84	(24.02)	1,320
Hours worked last week	-2.45	(2.06)	[0.30]	6.40	(14.23)	1,253
Was self employed ({0,1}, 30 days)	-0.06	(0.07)	[0.32]	0.25	(0.44)	1,075
Was informally emplyed ({0,1}, 30 days)	-0.09	(0.05)	[0.30]	0.19	(0.39)	1,321
Was formally emplyed ({0,1}, 30 days)	-0.01	(0.01)	[0.32]	0.01	(0.08)	1,321
Hours worked incl. chores last week	-4.73*	(2.63)	[0.30]	20.93	(18.07)	1,300
Hours spent on job search last week	0.52	(1.19)	[0.58]	4.67	(8.31)	1,259

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022-2023 phone survey data. Each row is its own dependent variable.
- Earnings include all earnings in the past 30 days from any income-generating activity including wage work and self employment. IHS transform is the inverse hyperbolic sine transform which includes zeros; log drops zeros. Any income generating work is an indicator for whether the FR did any income generating work, even for an hour, in the past 30 days. Hours worked in a typical week in the past 30 days refers to the hours of any income generating activity in a typical week in the past month. Hours worked last week refers to the last 7 days. Was ... employed are indicators referring to the following question "Was this work primarily self-employment, informal employment without a written contract, or formal employment with a written contract?" People not working have zeros. Hours worked including chores is "hours worked last week" + hours spent on household chores. Hours spent on job search last week on "actively searching for jobs, applying for jobs, or in interviews?"
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family xxx in the pre-analysis plan.

Table (F6) Impacts of the Program on Migration

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Probability of conflict resolving in two years (index)	-0.00	(0.16)	[1.00]	0.02	(1.01)	1,130
Probability of returning to Syria w/ conflict unresolved (index)	0.10	(0.16)	[1.00]	-0.02	(0.96)	1,300
Probability of returning to Syria w/ conflict resolved (index)	0.08	(0.18)	[1.00]	0.01	(1.00)	1,294
Conflict more likely to increase in next three months {0,1}	0.07	(0.09)	[1.00]	0.54	(0.50)	1,073

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022-2023 phone survey data. Each row is its own dependent variable.
- Probability of conflict resolving in 2 years is a standardized likert scale (1-4) from very unlikely to very likely. Probability of returning to Syria with conflict unresolved/resolved are standardized likert scales (1-4) from very unlikely to very likely. Conflict more likely to increase in the next 3 months is an indicator.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family xxx in the pre-analysis plan.

Table (F7) Impacts of the Program on Physical and Mental Health

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Recent symptoms (index)	0.11	(0.17)	[0.43]	-0.01	(1.01)	1,321
Current health conditions (index)	0.21	(0.16)	[0.25]	0.01	(1.01)	1,321
Diagnosis of (2) after Oct 2019 (index)	0.12	(0.19)	[0.43]	0.02	(1.03)	1,321
Subjective health (index)	-0.17	(0.15)	[0.25]	-0.00	(1.00)	1,321
Subjective happiness (index)	-0.36**	(0.16)	[0.19]	0.01	(1.00)	1,319
Depression (CESD, std)	0.26*	(0.15)	[0.21]	-0.01	(1.00)	1,285
Depressed {0,1}	0.11*	(0.07)	[0.21]	0.75	(0.43)	1,321

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022-2023 phone survey data. Each row is its own dependent variable.
- Recent symptoms index includes the following: rapid weight loss, fast and irregular heart beat, difficulty breathing, stomach pain, headaches. Current health conditions is an index including diabetes, hypertension and cancer. Diagnosis of (2) after 2019 is an index of indicators equalling one if FR was diagnosed with one of the above conditions after Oct 2019, zero if before or if does not have diagnosis. Subjective health is a standardized likert scale (1-4) of self-reported health from good to very poor. Subjective happiness is a standardized likert scale (1-3) of self reported happiness. Depression reports the standardized CESD score where higher values report higher risk of depression. Center for Epidemiologic Studies Depression Scale (CES-D) is a measure for depressive symptoms and includes 10 items asking about the past week. It is defined as a mean-effect index of 10 items scored 0-3 in the past week (I was bothered by things that usually don't bother me, I had a problem in concentration on what I was doing, I felt depressed and troubled in my mind, I felt that everything that I did took up all my energy, I felt hopeful about the future (reversed), I felt afraid, I had difficulty in sleeping peacefully, I was happy (reversed), I felt lonely, I lacked the motivation to do anything). It is then normalized by subtracting the control group's mean and dividing by its standard deviation. Depressed is an indicator for being above the CESD threshold of 10 out of 30.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family xxx in the pre-analysis plan.

Table (F8) Impacts of the Program on Child Outcomes

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Child SDQ score, std	0.56***	(0.17)	[0.01]	0.01	(1.00)	859
Count of non-adult dependents (under 18)	-0.22	(0.20)	[0.37]	3.44	(2.02)	1,321
Share of non-adults attended school all of last week	-0.02	(0.07)	[0.55]	0.78	(0.35)	981
Days child slept hungry last week	0.32	(0.23)	[0.31]	0.89	(1.50)	1,118

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022-2023 phone survey data. Each row is its own dependent variable.
- Child SDQ score, std reports the standardized SDQ score where higher scores reports worse difficulties. Strengths and Difficulties Questionnaire (SDQ) is a screening questionnaire administered to a randomly selected child aged three to eight years old and covers emotional, conduct, hyperactivity, peer, and prosocial problems. Count of non-adult dependents under 18 reports household members under 18. Share of non-adults attended school all of last week reports the share of children aged 6-17 who attended 5 days of school last week. Days child slept hungry last week reports the number of days any child under 18 went to bed hungry because there wasn't enough food to eat.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family xxx in the pre-analysis plan.

Table (F9) Impacts of the Program on Time Use

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Domestic work hours (Last week)	-2.82*	(1.72)	[0.46]	14.86	(14.85)	1,296
Childcare hours (Last week)	-5.76**	(2.81)	[0.46]	22.63	(20.75)	1,274
Hour spent outside the home (Last week)	-3.17	(2.56)	[0.46]	13.00	(17.69)	1,268
Hours children alone at home (Last week)	0.51	(0.60)	[0.51]	0.66	(2.75)	1,123
Hours childcare provided by friends/family (Last week)	-0.61	(0.49)	[0.46]	0.92	(4.68)	1,275
Hours w friends/family outside the home (Last week)	0.15	(1.08)	[0.51]	3.32	(6.44)	1,276
Non-family childcare by Jordanians (Last week)	-0.12	(0.10)	[0.46]	0.07	(0.85)	1,274
Non-family childcare by Syrians (Last week)	-0.44*	(0.24)	[0.46]	0.20	(1.68)	1,274
Any paid childcare in 2020	0.00	(0.02)	[0.51]	0.01	(0.10)	1,318
Months of paid childcare	-0.06	(0.11)	[0.51]	0.06	(0.72)	1,318
Hours spent outside the home w/o children (Last week)	-2.00	(2.46)	[0.51]	9.92	(15.95)	1,268
Childcare hours by other HH members (Last week)	-17.50*	(10.03)	[0.46]	21.06	(22.46)	243
Chore hours by other HH members (Last week)	-11.97	(8.42)	[0.46]	18.78	(20.30)	267
Paid childcare hours (Last week)	-0.25	(0.16)	[0.46]	0.06	(0.83)	1,123

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022-2023 phone survey data. Each row is its own dependent variable.
- Domestic work hours reports the number of hours FR spent in the past 7 days on household chores like cooking and cleaning. Childcare hours including time spent taking care of children while not doing other chores. Hours spent outside the home is the sum of hours spent running errands or shopping, visiting friends or family, or working for pay. Hours children alone at home reports hours in the past 7 days children were at home without adults present. Hours of children by friends/family outside the home is the total hours cared by family or friends/neighbors (without pay). Non-family childcare by Jordanians/Syrians is the total non-family childcare weighted by whether they reported with was provided by a) jordanians, b) syrians or c) both. If both, the hours are split 50/50 between syrian and jordanian. Any paid childcare in 2020 is an indicator. (2020 is when the program was active). Months of paid childcare reports the number of months in 2020 paid childcare was used, including zeros. Hours spent outside the home without children report the number of hours spent running errands or visiting friends/family without bringing children. Childcare hours and chore hours report the number of hours spent by all other household members on these tasks, defined above. Paid childcare hours reports the number of hours of paid childcare the household used in the past 7 days.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family xxx in the pre-analysis plan.

Table (F10) Impacts of the Program on Relationships and MacArthur Ladder

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
No. of close Jordanian friends	-0.47	(0.51)	[0.56]	2.98	(3.66)	1,321
Met close Jordanian friend at child's school	-0.01	(0.01)	[0.56]	0.00	(0.07)	1,249
Met close Jordanian friend as neighbor	-0.13*	(0.07)	[0.56]	0.52	(0.50)	1,321
Overwhelmed by financial needs of others (index)	0.26	(0.17)	[0.56]	-0.02	(1.01)	936
Current Macarthur rung (index)	-0.01	(0.12)	[0.84]	-0.01	(0.98)	1,321
Aspired Macarthur rung (index)	0.03	(0.13)	[0.84]	0.00	(1.00)	1,321

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022-2023 phone survey data. Each row is its own dependent variable.
- N. of close Jordanian friends reports the treatment effect on the number of close Jordanian friends the respondent reports. Met close Jordanian friend at ... reports an indicator for where the respondent met the most recent close Jordanian friend they made. Overwhelmed by financial needs of others is a standardized likert response to "Over the past week, I have felt overwhelmed or burdened by the financial needs of people outside my household". Current Macarthur rung is a standardized version of the following question: "Imagine a ladder with 10 rungs representing society. At the top of the ladder are the people who are the best off, those who have the most money, most education, and best jobs. At the bottom are the people who are the worst off, those who have the least money, least education, worst jobs, or no job. If the top of the ladder is the 10th rung and the bottom is the 1st. Which rung best represents where you think you stand?" Aspired macarthur rung is a standardized version of the 1-10 response on which rung they would like to acheive in their life.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family xxx in the pre-analysis plan.

Table (F11) Impacts of the Program on Preferences towards it

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Prefers transfer to landlord vs. cash	0.03	(0.07)	[1.00]	0.53	(0.50)	1,321
Prefers transfer to landlord vs. food vouchers	0.07	(0.07)	[1.00]	0.27	(0.44)	1,321
Expected Benefit from Program Index	-0.05	(0.12)	[1.00]	-0.01	(1.00)	1,297

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022-2023 phone survey data. Each row is its own dependent variable.
- Prefers transfer to landlord vs cash is the treatment effect on an indicator for preferring a program that pays 75 JD to the landlord towards rent each month for 1 year over a program that gives cash worth [75, 65 or 55] JD, each month. (75, 65 and 55 were randomized with equal probability)
- Prefers transfer to landlord vs cash is the treatment effect on an indicator for preferring a program that pays 75 JD to the landlord towards rent each month for 1 year over a program that gives food voucher worth [75, 65 or 55] JD, each month (averaged). (75, 65 and 55 were randomized with equal probability)
- Expected benefit program index is a standardized index of the following outcomes: 1) would spend more on food consumption if they were receiving the program 2) would spend more on healthcare if they were receiving the program and 3) would be happier if they were receiving the program.
- The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family xxx in the pre-analysis plan.

4 Appendix G: Social Integration Results

Table (G1) Impacts of the Program on Primary Outcomes

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Social attitudes & perceptions (SD)	-0.33**	(0.14)	[0.08]	-0.00	(1.00)	1,102
Economic attitudes & perceptions (SD)	0.02	(0.16)	[0.56]	0.02	(1.02)	1,102
Policy preferences (SD)	0.17	(0.15)	[0.37]	-0.02	(1.00)	1,102
Altruism to Syrians	0.21	(0.18)	[0.37]	0.88	(1.18)	1,102

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022 neighbor survey data. Each row is its own dependent variable.
- The outcomes that require definitions are: Social attitudes and perceptions is an index of the following outcomes: of the 3 people the respondent socializes most with, how many are Syrian; of the 3 people the respondent most often shares advice with, how many are Syrian; do children in the household have Syrian friends (binary); do children in the household share recreational space with Syrian children (binary); what is the net effect of Syrian refugees on Jordan's society (positive-neutral-negative); do Syrian refugees tend to be lazy or hardworking? (1-7); To what degree would people feel comfortable accepting the marriage of their family member to a Syrian refugee? (1-5); To what degree would people feel comfortable accepting a Syrian refugee as a neighbor? (1-5, standardized). These outcomes are standardized with respect to the control group, averaged, then standardized with respect to the control group again.

  Economic attitudes and perceptions is an index of the following outcomes: a binary variable equal to one if the respondent listed 'hosting syrian refugees' as one of the most important challenges facing Jordan; a 5-point likert scale on whether syrians or jordanians pay more in taxes; the net effect of refugees on the Jordanian economy (positive-neutral-negative). These outcomes are standardized with respect to the control group, averaged, then standardized with respect to the control group again.

  Policy preferences is an index of outcomes which report the degree to which respondents agree with the following statements, each of which were asked using a 5-point likert scale: refugees should be relocated to refugee camps; refugees should have the right to work outside refugee camps; refugees should be allowed to become full citizens if they have lived in Jordan for a long time and would like to become a Jordanian; Syrian refugees should be given unrestricted work permits; Syrian refugee children should be allowed to be in classes with Jordanian children; Syrian refugees should be allowed to enter and leave camps freely; Syrian refugees should be given housing assistance through shelter programs that subsidize their rent; the international community should spend more money to support refugees. All outcomes are standardized with respect to the control group, averaged, then standardized with respect to the control group again.

  Altruism to Syrians ranges from 0-5 JD and reports how much of a 5JD transfer was allocated to a charity serving Syrian refugees vs a charity serving vulnerable Jordanians vs kept for self. The altruism exercise was incentivized for one third of the sample, who were actually given 5 JD to allocate between the 3 choices. The rest of the sample was asked to make a hypothetical choice.
- The regressions also have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 1 in the pre-analysis plan.

Table (G2) Impacts of the Program on Assimilation Gap

Outcomes	Treatment	(se)	Treatment *Refugee	(se)	Refugee	(se)	ctrl. mean	(sd)	N
Housing Expenditure (PC) PPP	30.02	(97.82)	-38.97	(93.38)	-31.25*	(18.05)	57.04	(274.59)	2,491
Log total consumption (PC) PPP	-0.13	(0.22)	0.17	(0.25)	-0.83***	(0.07)	7.76	(0.68)	1,482
CESD Score	-0.24	(0.16)	0.41	(0.30)	-0.05	(0.05)	0.02	(0.99)	2,471
Child Strengths and Difficulties	-0.24	(0.32)	0.59	(0.39)	-0.01	(0.11)	0.01	(1.00)	1,321

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022 neighbor survey data *and* 2021 refugee survey data. Each row is its own dependent variable.
- This table reports estimates from the assimilation gap specification described in the pre-analysis plan amendment, which reports the degree to which treatment reduced the gap between Syrian refugees on 4/5 of the primary outcomes. Regrettably, housing quality is excluded due to data collection errors in the field.
- Housing expenditure includes total monthly rent payment, mortgage payment, and home upgrade costs, converted to USD PPP, then divided by household size. Log total consumption is the sum of the monetary value of goods consumed by households through purchase, gift and barter, excluding housing costs, converted to annual USD PPP, then divided by household size. Goods include food consumption, non-food consumption and durables purchases. CESD score is a standardized version of the Center for Epidemiological Studies Depression scores which measures depressive symptoms. 10 questions are scaled and summed according to the scoring guidelines, then standardized with respect to the control group. An increase corresponds to an increase in the CESD score which indicates increased symptoms of depression. Child strengths and difficulties is scored and summed according to the guidelines, then standardized with respect to the control group. An increase corresponds to an increase in the SDQ score, which indicates increased difficulties.
- The regressions also have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 2 in the pre-analysis plan.

Table (G3) Impacts of the Program on Host Community Relations and Attitudes Towards Refugees

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
List experiment; Syrian as neighbor	0.13	(0.11)	[0.40]	1.23	(0.86)	1,102
Share Friends Syrian	-0.15	(0.13)	[0.40]	0.51	(0.92)	1,102
Share Advice Syrian	-0.24**	(0.11)	[0.32]	0.34	(0.72)	1,102
Kids have Syrian Friends	-0.19*	(0.11)	[0.37]	0.40	(0.49)	704
Kids share rec space w Syrians	-0.10	(0.08)	[0.40]	0.48	(0.50)	707
Accept Syrian marriage	-0.23	(0.15)	[0.40]	0.01	(1.00)	1,099
Accept Syrian Neighbor	-0.15	(0.15)	[0.44]	-0.00	(1.01)	1,101
Syrians pay more in taxes	0.06	(0.14)	[0.49]	0.03	(1.01)	1,078

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022 neighbor survey data. Each row is its own dependent variable.
- The first outcome titled "list experiment" uses a different specification specific to the list experiment. All respondents were asked how many of the listed groups they'd accept as a neighbor. Half saw a list that included Syrian refugees, and the other half saw a list that did not include Syrian refugees. (The other 3 groups were Persons with disabilities, People in poverty, and Someone who does not follow the law). The reported coefficient is the coefficient on an indicator for seeing the list including Syrian refugees.
- The other outcomes that require definitions are: Share friends Syrian is the share of the 3 friends they visit regularly who are Syrian refugees. Share advice Syrian is the share of the 3 people they seek and give advice to who are Syrian refugees. Kids have Syrian friends is an indicator equaling 1 if the children in the household have any Syrian friends, and is missing for households without children.Kids share rec space is an indicator equaling 1 if the children in the houseold share recreational space with Syrian refugee children. Accept Syrian marriage is a standardized 5-point likert scale regarding "To what degree would people in your community feel comfortable accepting the marriage of their son/daughter/sister/brother to a Syrian refugee?" Accept Syrian neighbor is a standardized 5-point likert scale regarding "To what degree would people in your community feel comfortable being neighbors with a Syrian refugee?" Syrians pay more in taxes is a standardized 5-point likert scale where an increase corresponds with a belief that Syrian refugees pay more in taxes than the average Jordanian in response to a short vignette.
- The regressions also have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 3 in the pre-analysis plan.

Table (G4) Impacts of the Program on Altruism and Trust

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Trust in Syrians	-0.45	(0.46)	[0.37]	4.75	(3.37)	1,102
Dictator game, Altruism to Jordanians	-0.51*	(0.28)	[0.37]	1.88	(1.77)	1,102
Dictator game, keep	0.30	(0.31)	[0.37]	2.24	(2.03)	1,102
Incentivized Dictator game, Altruism to Jordanians	-0.43	(0.53)	[0.37]	1.95	(1.98)	375
Incentivized Dictator game, keep	0.97	(0.60)	[0.37]	2.29	(2.26)	375

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022 neighbor survey data. Each row is its own dependent variable.
- The outcomes that require definitions are: trust in Syrians reflects how much of the total endowment the respondent chooses to share with a Syrian refugee game partner in a hypothetical trust game. The respondent is told they (hypothetically) have 10 JD and can share as much of it or as little of it with their game partner as they want. Anything shared will be doubled, and the partner could chose to share or not share the proceeds with the respondent, so the amount shared with the game partner reveals a measure of trust. The subsequent outcomes correspond to a dictator game that was conducted as follows: the respondent was given 5 JD and told to distribute it between the following categories: keep for self, donate to a charity helping needy Jordanians, or donate to a charity helping needy Syrian refugees in Jordan. 2/3 of the sample did this hypothetically, and 1/3 actually received the 5 JD. “Altruism to Jordanians” reports how much of the 5 JD was allocated to the charity for Jordanians, and “keep" reports how much of the 5 JD was kept for self. Both these outcomes include the entire sample. The final two rows report these outcomes only for the incentivized sample. Altruism to Syrians is reported in table G1
- The regressions also have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 4 in the pre-analysis plan.

Table (G5) Impacts of the Program on Social Attitudes and Policy Preferences

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Attitude on refugee relocation	0.01	(0.16)	[1.00]	0.01	(0.99)	1,102
Attitude on refugee work	0.11	(0.15)	[0.91]	-0.01	(0.99)	1,102
Attitude on refugee citizenship	0.05	(0.14)	[1.00]	0.01	(1.01)	1,102
Refugee effect on economy	-0.23	(0.14)	[0.38]	-0.01	(1.00)	1,102
Refugee effect on society	0.02	(0.16)	[1.00]	-0.04	(0.97)	1,102
Positive effects of refugees	-0.13	(0.15)	[0.76]	-0.02	(0.98)	1,102
Negative effects of refugees	-0.23	(0.14)	[0.38]	0.03	(1.01)	1,074
Taxes best to reduce poverty	-0.01	(0.04)	[1.00]	0.06	(0.23)	1,042
Attitude on Syrian refugee work ethic	0.13	(0.30)	[1.00]	5.82	(1.75)	1,086
Support for work permits	-0.03	(0.16)	[1.00]	-0.01	(1.00)	1,102
Support for integrated schooling	0.19	(0.15)	[0.46]	-0.02	(1.01)	1,102
Support for freedom of movement	-0.02	(0.17)	[1.00]	-0.03	(1.00)	1,102
Support for housing assistance	0.24	(0.15)	[0.38]	-0.04	(1.01)	1,102
Support for intl refugee assistance	0.28*	(0.15)	[0.38]	-0.03	(1.02)	1,102
Support for integration index	0.08	(0.16)	[1.00]	-0.01	(0.98)	1,102
Primary identity - religion	-0.11	(0.07)	[0.38]	0.84	(0.36)	1,097
Primary identity - not national	-0.07	(0.06)	[0.50]	0.87	(0.33)	1,097
Days of media consumption	-1.08***	(0.40)	[0.08]	4.78	(2.78)	1,102
Refugees a top challenge	0.04	(0.04)	[0.76]	0.05	(0.22)	1,102
Perception of Jordanian economy	0.38**	(0.17)	[0.25]	-0.01	(0.95)	1,092
Share of Syrians receiving aid	-0.31	(3.87)	[1.00]	62.12	(32.13)	1,075
Avg aid amount (PPP)	-347.17***	(109.69)	[0.04]	630.84	(657.41)	685

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022 neighbor survey data. Each row is its own dependent variable.
- Each of the following outcomes are standardized so that an increase corresponds with a more "pro-refugee" attitude. Refugee relocation is the response to "All refugees in Jordan should be relocated to live in the camps." Refugee work corresponds to " Refugees who live in Jordan right now should be allowed to continue to work outside the camps." Refugee citizenship corresponds to "Refugees should be allowed to become full citizens if they have lived in Jordan for a long time and would like to become a Jordanian. As citizens, they would have the right to vote in Jordan elections." Effect on economy corresponds to " In your opinion, is the net effect of Syrian refugees on Jordan's economy positive, neutral, or negative?" Effect on society corresponds to " In your opinion, is the net effect of Syrian refugees on Jordan's society positive, neutral, or negative? Positive effects corresponds to the number of positive effects on society listed by the respondent (options not read aloud). Negative effects corresponds to the number of negative effects on society listed by the respondent (options not read aloud). Taxes best to reduce poverty indicates support for "Statement 1: The best way to reduce poverty is to increase your taxes so the government can help the poor through social spending." vs "Statement 2: The best way to reduce poverty is to encourage people like yourself to pay more sadaqa for charitable distribution." Attitude on Syrian refugee work ethic corresponds to "Do Syrian refugees tend to be hardworking or lazy? On a scale of 1-7, with 1 being lazy and 7 being hardworking?" Work permits corresponds to "Do you think Syrian refugees should be given unrestricted work permits?" Integrated schooling corresponds to "Do you think Syrian refugee children should be allowed to be in classes with Jordanian children?" Freedom of movement corresponds to " Do you think Syrian refugees should be allowed to enter and leave camps freely?" Housing assistance corresponds to "Do you think Syrian refugees should be given housing assistance through shelter programs that subsidize their rent?" Intl refugee assistance corresponds to " Do you think the international community should spend more money to support refugees?" Support for integration index is an index composed of the the following attitude questions: relocation, work, citizenship, work permits, integrated schooling, freedom of movement, and housing assistance. Primary identity corresponds to the following question: "Which of the following best describes you?" (Read the options aloud.); Above all I am a Jordanian; Above all I am a Muslim; Above all I am an Arab; Above all I am a Christian. Days of media consumption corresponds to "In the past 7 days, how many days did you read or listen to the news from any source, including newspapers, online, WhatsApp, etc.?" Refugees a top challenge is an indicator equalling 1 if the respondent volunteered that refugees were one of the most important challenges facing Jordan (options not read aloud). Perception of jordanian economy corresponds to "How would you evaluate the current economic situation in your country?' Share of Syrians receiving aid and average aid amount correspond to the following questions: "What percent of refugee households in your neighborhood do you think receive any assistance from any organizations in a typical month, in cash or in kind? Enumerator: Do not include assistance received from other households." and "Of the refugee households in your neighborhood who receive assistance, what do you think is the average value in Dinar of the assistance (in cash or in kind) that they receive from organizations in a typical month?" The negative treatment effect stems from the control group being more likely than the treatment group to report outlandishly large numbers.
- The regressions also have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 5 in the pre-analysis plan.

Table (G6) Impacts of the Program on Dwelling Characteristics

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Housing expenditure (PPP PC)	158.54	(203.08)	[1.00]	96.06	(564.32)	1,096
Rent paid (PPP)	101.12	(94.99)	[1.00]	371.63	(169.40)	232
=1 if owns home	-0.10	(0.07)	[1.00]	0.80	(0.40)	1,102
Dwelling value (PPP)	9,320.34	(18,361.52)	[1.00]	46,908.00	(85,074.48)	566
=1 if hh paid for any improvements	0.06	(0.07)	[1.00]	0.40	(0.49)	1,102
=1 if NGO paid for any improvements	-0.00	(0.01)	[1.00]	0.00	(0.06)	1,102
Index of improvements	0.01	(0.04)	[1.00]	0.16	(0.23)	1,102
FR childcare and chore hours	-3.66	(2.86)	[1.00]	19.51	(18.70)	1,070
HH childcare and chore hours	1.25	(5.31)	[1.00]	32.70	(25.24)	681
Any domestic employees	0.01	(0.02)	[1.00]	0.02	(0.15)	1,102
N domestic employees	0.02	(0.03)	[1.00]	0.02	(0.15)	1,102
N Syrian domestic employees	-0.00	(0.01)	[1.00]	0.01	(0.07)	1,102

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022 neighbor survey data. Each row is its own dependent variable.
- The outcomes that require definitions are: Housing expenditure is the sum of monthly rent paid, mortgage paid and money spent on housing upgrades, divided by household size. Rent paid corresponds to the monthly rent paid by households who rent. Dwelling value corresponds to "If you were to sell your house in the next month, how much money do you think it would sell for? In other words, what is the current market value of your home?" =1 if .. paid for any improvements corresponds to " Have you or an NGO made (or paid for) housing improvements where you live since October 2019?" Index of improvements is an index of structural, cosmetic, mold, insulation and utilities upgrades. FR/HH childcare and chores hours corresponds to the hours spent by the focus respondent/household in the last 7 days on childcare and chores. Any domestic employees is an indicator equal to 1 if the household currently employs any domestic employees, and N refers to the number of employees.
- The regressions also have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 6 in the pre-analysis plan.

Table (G7) Impacts of the Program on Household Consumption and Expenditures

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Diversity of food purchases	0.75**	(0.36)	[0.66]	-0.02	(1.03)	277
Food exp, typical week (USD PPP)	-6.56	(44.64)	[1.00]	277.97	(157.96)	287
Food exp, last week (USD PPP)	-14.50	(18.36)	[1.00]	189.65	(112.09)	1,090
Value of gifted food, last month (USD PPP)	9.60	(8.15)	[1.00]	2.09	(14.55)	297
Food consumption, last year	651.14	(2,517.53)	[1.00]	14,409.39	(9,974.92)	257
Non-food exp, last month (USD PPP)	13.49	(156.37)	[1.00]	831.77	(874.02)	1,016
Diversity of non-food purchases	-0.03	(0.15)	[1.00]	0.01	(1.01)	1,054
Value of gifted non-food, last yr (USD PPP)	-18.87	(19.03)	[1.00]	11.68	(186.07)	1,100
Value of all durables acquired, last yr (USD PPP)	752.90	(862.95)	[1.00]	354.88	(1,977.68)	296
Value of all durables gifted, last yr (USD PPP)	146.24	(179.81)	[1.00]	16.05	(174.57)	296

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022 neighbor survey data. Each row is its own dependent variable.
- The outcomes that require definitions are: Diversity of food purchases is a standardized measure of the number of months per year the household consumes food from 9 different food groups. Food exp refers to the PPP adjusted amount of money spent in a typical week or last week. Value of gifted food is the value of food given to the household as formal or informal assistance in the past month. Food consumption, last year is an annualized measure of food consumption calculated from weekly food expenditure * 4.3 * months that food was consumed + annualized weekly food produced + annualized monthly food assistance. Non-food expenditure includes expenditures in the last 30 days from 9 different non-food expenditure categories. Diversity of non-food purchases is calculated in the same way as described above for food purchases. Value of gifted non-food includes all items from 9 non-food categories received as gifts in the past 12 months. Value of all durables acquired includes durables from 9 categories purchased or gifted to the household. Value of durables gifted includes all durables the household received as gifts.
- The regressions also have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family xxx in the pre-analysis plan.

Table (G8) Impacts of the Program on Food Security

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
N meals yesterday	0.08	(0.10)	[0.71]	2.03	(0.71)	1,101
Food diversity	0.65**	(0.31)	[0.30]	8.12	(1.17)	296
Days last week FR went hungry	-0.13	(0.14)	[0.71]	0.28	(0.90)	1,102
Days last week adults went hungry	0.03	(0.11)	[1.00]	0.18	(0.70)	1,018
Days last week children went hungry	-0.16	(0.14)	[0.68]	0.22	(0.75)	707
women go hungry more often than men (SD)	0.09	(0.13)	[0.72]	0.04	(1.02)	962
girls go hungry more often than boys (SD)	-0.37*	(0.21)	[0.30]	0.04	(1.02)	425
elderly go hungry more often than young (SD)	-0.02	(0.28)	[1.00]	0.09	(1.06)	135
Reduced coping	-1.68*	(0.86)	[0.30]	3.10	(6.91)	1,102

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022 neighbor survey data. Each row is its own dependent variable.
- The outcomes that require definitions are: N means yesterday reports the number of meals the respondent ate yesterday. "How many meals did you eat yesterday? Tea alone is not to be considered as a meal." Food diversity reports the number of food categories out of 9 the respondent ate from in the past 12 months. Days last week FR/adults/children went hungry reports the number of nights in the past 7 days the respondent/adults in household/children in household went to bed hungry. Women/girls/elderly go hungry more often than men/boys/young reports agreement with the statement as a standardized 5-point likert scale. Reduced coping is the reduced Coping Strategies Index, where larger numbers correspond to more food insecurity.
- The regressions also have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 8 in the pre-analysis plan.

Table (G9) Impacts of the Program on Earnings, Labor, and Occupational Choice

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Total pre-tax earnings, 30 days IHS	0.08	(0.38)	[1.00]	1.20	(2.83)	1,090
Total labor supply, 7 days	-0.26	(4.34)	[1.00]	29.61	(29.98)	1,073
Total labor supply w/o chores, 7 days	2.53	(3.56)	[1.00]	8.55	(24.10)	1,091
Total labor supply, avg month	-2.22	(2.77)	[1.00]	8.09	(22.75)	1,091
Self-employed labor supply, typical week	-2.84	(2.30)	[1.00]	2.82	(14.13)	1,100
Wage employment labor supply, typical week	1.83	(3.24)	[1.00]	7.14	(22.31)	1,093
Net wage employment earnings, 30 days IHS	0.15	(0.36)	[1.00]	0.94	(2.48)	1,082
Manufacturing sector	0.01	(0.02)	[1.00]	0.01	(0.07)	1,102
Service sector	0.08	(0.05)	[1.00]	0.10	(0.30)	1,102
Retail sector	0.04	(0.03)	[1.00]	0.03	(0.18)	1,102
Agricultural sector	-0.00	(0.01)	[1.00]	0.01	(0.08)	1,102
Business revenue, 30 days	-36.48	(34.79)	[1.00]	25.97	(374.33)	1,101
Business revenue, 12 months	-442.81	(360.12)	[1.00]	278.07	(3,783.38)	1,101
N. employees	-0.07	(0.11)	[1.00]	0.04	(0.65)	1,102
Currently operating business	-0.02	(0.03)	[1.00]	0.03	(0.17)	1,102
Self-employment profit, 30 days	32.56	(41.14)	[1.00]	16.91	(192.12)	1,097
Self-employment profit, 12 months	99.60	(432.52)	[1.00]	176.99	(1,385.86)	1,095
Business expenses, 30 days	-26.01	(22.62)	[1.00]	21.75	(300.90)	1,101
Business expenses, 12 months	-306.58	(249.12)	[1.00]	217.15	(3,118.88)	1,101

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022 neighbor survey data. Each row is its own dependent variable.
- The outcomes that require definitions are: total pre-tax earnings includes all earnings from wage work and profit from self employment in the last 30 days, winsorized then transformed using the inverse hyperbolic sine (prespecified). Total labor supply reports hours in wage work, self employment, and chores/childcare in the past 7 days. Total labor supply w/o chores only includes wage work and self employment. Total labor supply, avg month reports wage work + self employment hours in a typical month. Self-employed labor supply reports hours spent in self employment in a typical week. Wage employment labor supply reports hours spent in wage work in a typical week. Net wage employment earnings include earnings from wage work minus taxes plus value of benefits, winsorized then transformed with the inverse hyperbolic sine. Manufacturing/service/retail/agricultural sector is an indicator equal to one if the respondent has worked in that industry in wage work or self employment in the past 12 months. Business revenue reports total business revenue earned by the respondent in the past 30 days/12 months. N employees reports the number of people the respondent employs in all businesses in the past 30 days. Self-employment reports revenue less expenses for all businesses. Business expenses reports self employment costs.
- The regressions also have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 9 in the pre-analysis plan.

Table (G10) Impacts of the Program on Savings and Loans

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
At least 30 JD in savings	-0.12*	(0.07)	[0.34]	0.32	(0.47)	1,102
Value of loans taken in past year, USD PPP	-144.22	(1,980.97)	[1.00]	4,128.63	(12,492.84)	1,089
Value of loans given in past year, USD PPP	-113.90	(635.12)	[1.00]	613.62	(4,147.31)	1,084

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022 neighbor survey data. Each row is its own dependent variable.
- The outcomes that require definitions are: At least 30 JD in savings is an indicator equal to 1 if the household has at least 30 JD in savings, in a bank or otherwise. Value of loans taken in the past year includes formal and informal loans taken out by the household. Value of loans given includes all formal and informal loans made by the household to other individuals.
- The regressions also have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 10 in the pre-analysis plan.

Table (G11) Impacts of the Program on Physical and Mental Health

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Life Satisfaction (SD)	-0.32**	(0.15)	[0.06]	0.01	(0.98)	1,097
Subjective Health (SD)	-0.30*	(0.16)	[0.06]	0.02	(0.99)	1,102

Notes:

- The table shows the regression results on pre-specified main outcomes using the 2022 neighbor survey data. Each row is its own dependent variable.
- The outcomes that require definitions are: Life satisfaction corresponds to a standardized measure from "All things considered, how satisfied are you with your life as a whole these days on a scale of 1 to 10?" Subjective health is a standardized measure from "Would you describe your general health as good, fair, poor, or very poor?"
- The regressions also have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family xxx in the pre-analysis plan.

4.1 Social Cohesion Heterogeneity

Table (G12) Jordanian Neighbor Primary Outcomes Heterogeneity by: Marlowe-Crowne Score

	(1) Social Attitudes and Perceptions	(2) Economic Attitudes and Perceptions	(3) Policy Preferences	(4) Altruism to Syrians
Treat	-0.530**	0.094	0.215	0.117
	(0.216)	(0.228)	(0.184)	(0.284)
Treat*Above Median Marlow-Crowne	0.411	-0.161	-0.096	0.189
	(0.297)	(0.257)	(0.266)	(0.467)
Above Median Marlow-Crowne	-0.048	0.101	0.015	0.053
	(0.096)	(0.093)	(0.079)	(0.148)
p-value: T + T*Het	0.53	0.72	0.58	0.33
Control Mean	0.48	0.48	0.48	0.48
Control SD	0.50	0.50	0.50	0.50
N	1102.00	1102.00	1102.00	1102.00

This table shows heterogeneity analysis of the primary outcomes for Jordanian neighbors, by whether the respondent's Marlowe-Crowne score is above or below the median score in the control group. The outcomes are as defined in table G1. The regressions have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%).

Table (G13) Jordanian Neighbor Primary Outcomes Heterogeneity by: Palestinian Grandparents

	(1) Social Attitudes and Perceptions	(2) Economic Attitudes and Perceptions	(3) Policy Preferences	(4) Altruism to Syrians
Treat	-0.390***	0.085	0.156	0.233
	(0.143)	(0.184)	(0.161)	(0.199)
Treat*Palestinian Grandparents	0.541	-0.558	0.164	-0.141
	(0.453)	(0.402)	(0.518)	(0.659)
Palestinian Grandparents	-0.031	0.135	0.340**	0.160
	(0.101)	(0.095)	(0.133)	(0.141)
p-value: T + T*Het	0.73	0.18	0.51	0.88
Control Mean	0.12	0.12	0.12	0.12
Control SD	0.32	0.32	0.32	0.32
N	1102.00	1102.00	1102.00	1102.00

This table shows heterogeneity analysis of the primary outcomes for Jordanian neighbors, by whether any of the Jordanian neighbor respondent's grandparents are of Palestinian decent, indicating a likely family history of displacement. The outcomes are as defined in table G1. The regressions have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%).

Table (G14) Jordanian Neighbor Primary Outcomes Heterogeneity by: Grandparents Non-Native Jordanians

	(1) Social Attitudes and Perceptions	(2) Economic Attitudes and Perceptions	(3) Policy Preferences	(4) Altruism to Syrians
Treat	-0.385***	0.122	0.212	0.235
	(0.135)	(0.192)	(0.163)	(0.199)
Treat*Non-Jordanian Grandparents	0.347	-0.627	-0.270	-0.139
	(0.449)	(0.406)	(0.445)	(0.521)
Non-Jordanian Grandparents	0.161	0.135*	0.414***	0.201*
	(0.109)	(0.079)	(0.118)	(0.115)
p-value: T + T*Het	0.93	0.14	0.89	0.84
Control Mean	0.16	0.16	0.16	0.16
Control SD	0.37	0.37	0.37	0.37
N	1102.00	1102.00	1102.00	1102.00

This table shows heterogeneity analysis of the primary outcomes for Jordanian neighbors, by whether any of the Jordanian neighbor respondent's grandparents are of non-Jordanian decent, indicating a likely family history of migration or displacement. The outcomes are as defined in table G1. The regressions have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%).

Table (G15) Jordanian Neighbor Primary Outcomes Heterogeneity by: Proximity to Study Refugee Household

	(1) Social Attitudes and Perceptions	(2) Economic Attitudes and Perceptions	(3) Policy Preferences	(4) Altruism to Syrians
Treat	-0.427**	-0.014	0.190	0.067
	(0.204)	(0.191)	(0.215)	(0.285)
Treat*Above Median Distance	0.392	0.122	-0.018	0.283
	(0.282)	(0.280)	(0.277)	(0.527)
Above Median Distance	-0.069	0.033	0.046	-0.036
	(0.076)	(0.071)	(0.071)	(0.131)
p-value: T + T*Het	0.85	0.67	0.39	0.33
Control Mean	0.50	0.50	0.50	0.50
Control SD	0.50	0.50	0.50	0.50
N	1044.00	1044.00	1044.00	1044.00

This table shows heterogeneity analysis of the primary outcomes for Jordanian neighbors, by whether the distance (in meters) from the Jordanian neighbor respondent's home to the refugee household from the study is above or below the median distance of all neighbors in the control group. The outcomes are as defined in table G1. The regressions have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%).

Table (G16) Jordanian Neighbor Primary Outcomes Heterogeneity by: Gender

	(1) Social Attitudes and Perceptions	(2) Economic Attitudes and Perceptions	(3) Policy Preferences	(4) Altruism to Syrians
Treat	-0.179	-0.105	0.240	0.144
	(0.253)	(0.272)	(0.248)	(0.289)
Treat*Neighbor Female	-0.236	0.194	-0.117	0.110
	(0.360)	(0.371)	(0.296)	(0.397)
Neighbor Female	0.145	-0.077	0.083	0.027
	(0.101)	(0.071)	(0.062)	(0.093)
p-value: T + T*Het	0.04	0.69	0.50	0.31
Control Mean	0.60	0.60	0.60	0.60
Control SD	0.49	0.49	0.49	0.49
N	1102.00	1102.00	1102.00	1102.00

This table shows heterogeneity analysis of the primary outcomes for Jordanian neighbors, by whether the Jordanian neighbor respondent is female. The outcomes are as defined in table G1. The regressions have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%).

Table (G17) Jordanian Neighbor Primary Outcomes Heterogeneity by: Age Group 18-25

	(1) Social Attitudes and Perceptions	(2) Economic Attitudes and Perceptions	(3) Policy Preferences	(4) Altruism to Syrians
Treat	-0.268*	0.119	0.183	0.134
	(0.138)	(0.168)	(0.165)	(0.183)
Treat*Neighbor 18-25	-0.716	-1.280*	-0.208	0.927
	(0.624)	(0.713)	(0.525)	(1.241)
Neighbor 18-25	0.094	0.153	0.181	0.169
	(0.130)	(0.150)	(0.116)	(0.213)
p-value: T + T*Het	0.11	0.09	0.96	0.38
Control Mean	0.09	0.09	0.09	0.09
Control SD	0.29	0.29	0.29	0.29
N	1102.00	1102.00	1102.00	1102.00

This table shows heterogeneity analysis of the primary outcomes for Jordanian neighbors, by whether the Jordanian neighbor respondent is in the age group 18-25. The outcomes are as defined in table G1. The regressions have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%).

Table (G18) Jordanian Neighbor Primary Outcomes Heterogeneity by: Education Level

	(1) Social Attitudes and Perceptions	(2) Economic Attitudes and Perceptions	(3) Policy Preferences	(4) Altruism to Syrians
Treat	-0.222	0.396	0.232	-0.026
	(0.264)	(0.290)	(0.269)	(0.335)
Treat*Neighbor Secondary School or More	-0.163	-0.596*	-0.102	0.373
	(0.340)	(0.353)	(0.353)	(0.417)
Neighbor Secondary School or More	0.184*	0.093	0.142	-0.185
	(0.110)	(0.103)	(0.090)	(0.151)
p-value: T + T*Het	0.03	0.32	0.51	0.13
Control Mean	0.67	0.67	0.67	0.67
Control SD	0.47	0.47	0.47	0.47
N	1102.00	1102.00	1102.00	1102.00

This table shows heterogeneity analysis of the primary outcomes for Jordanian neighbors, by whether the Jordanian neighbor respondent has secondary education or higher. The outcomes are as defined in table G1. The regressions have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%).

Table (G19) Jordanian Neighbor Primary Outcomes Heterogeneity by: Socioeconomic Status

	(1) Social Attitudes and Perceptions	(2) Economic Attitudes and Perceptions	(3) Policy Preferences	(4) Altruism to Syrians
Treat	-0.382*	0.048	-0.320	0.511
	(0.208)	(0.265)	(0.215)	(0.328)
Treat*Above Median Expenditure	0.086	-0.017	0.832***	-0.401
	(0.270)	(0.310)	(0.288)	(0.504)
Above Median Expenditure	-0.191***	-0.054	-0.266***	0.011
	(0.074)	(0.074)	(0.067)	(0.114)
p-value: T + T*Het	0.12	0.87	0.02	0.71
Control Mean	0.50	0.50	0.50	0.50
Control SD	0.50	0.50	0.50	0.50
N	1016.00	1016.00	1016.00	1016.00

This table shows heterogeneity analysis of the primary outcomes for Jordanian neighbors, by whether the Jordanian neighbor respondent has above median non-food expenditure per capita. Non-food expenditure was selected as the measure of socioeconomic status since it was collected in the same way in both the abridged and non-abridged versions of the consumption module. It includes monthly expenditures on utilities, water, basic household items, debt repayment, linens, clothing, school fees, taxes, insurance, and phone bills. The outcomes are as defined in table G1. The regressions have enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%).

5 Appendix H: Ethics Appendix

To review the ethical considerations of this study, please review the comprehensive ethics appendix which is available on the Open Science Framework.

6 Appendix I: Pooled Results

6.1 Pooled Results Heterogeneity

Table (I1) Pooled Primary Outcomes Effects Heterogeneity by: Respondent Genderl

	(1) Overall Housing Quality (Z-Score)	(2) Total Monthly Housing Expenditures (USD PPP)	(3) Food Consumption (Log USD PPP)	(4) Log Total Consumption (Log USD PPP)	(5) CESD Score (Higher: Less Depression)	(6) SDQ Score (Higher: Better Child Wellbeing)
Treat	0.208	-32.083	0.053	0.011	-0.016	-0.294
	(0.217)	(33.821)	(0.122)	(0.125)	(0.171)	(0.202)
Treat*Female	0.208	0.500	-0.145	-0.077	-0.172	-0.091
	(0.210)	(48.234)	(0.194)	(0.212)	(0.256)	(0.302)
Female	0.118***	-5.136	-0.147***	-0.098*	-0.308***	-0.153**
	(0.039)	(11.549)	(0.048)	(0.052)	(0.055)	(0.075)
p-value: T + T*Het	0.07	0.40	0.45	0.62	0.33	0.09
Control Mean	0.96	0.64	0.63	0.33	0.96	0.42
Control SD	0.19	0.48	0.48	0.47	0.20	0.49
N	4218.00	2812.00	2753.00	1422.00	4207.00	1782.00

This table shows heterogeneity analysis of the refugee primary outcomes by respondent gender. The outcomes are as defined in table **??**. The regressions pool across all rounds of available data, and have enumerator fixed effects, round and year-month fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), refugee household-level controls (shelter program, baseline number of children, baseline number of children plus adults), the Marlowe–Crowne scale, and respondent gender, education, and age. In the parentheses are robust standard errors clustered at the locality level. Statistical significance represented by * (10%), ** (5%), and *** (1%).