I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17622

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

Abdulrazzak Tamim ®
Emma Smith ®
I. Bailey Palmer ®
Edward Miguel ®
Samuel Leone ®
Sandra V. Rozo ®
Sarah Stillman

JANUARY 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17622

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

Abdulrazzak Tamim
University of California Berkeley

Emma Smith
Georgetown University

I. Bailey Palmer
University of California Berkeley

Edward Miguel
University of California Berkeley and NBER

Samuel Leone
McKinsey & Company

Sandra V. Rozo
World Bank and IZA

Sarah Stillman
London School of Economics

JANUARY 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

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

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

Corresponding author:
Sandra V. Rozo
Development Research Group
The World Bank
1818 H Street
N.W. Washington
DC 20433
USA

E-mail: sandrarozo@worldbank.org

* Author order is Certified Random ® (AEA Confirmation Code gEv7QgDTbi7f). We are grateful to 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.

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

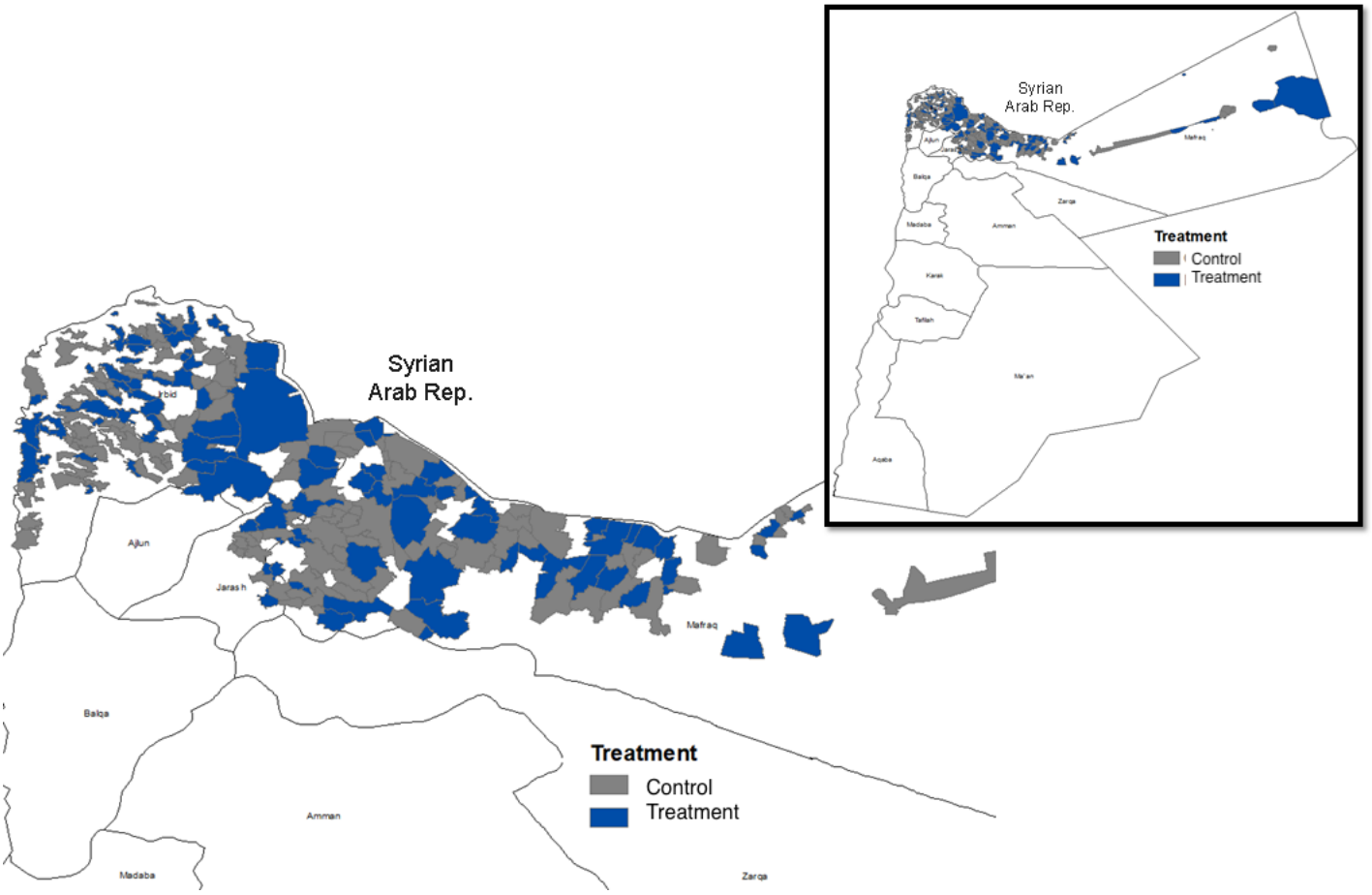


Figure (2) RCT, Data Collection, and Forecasts Timeline

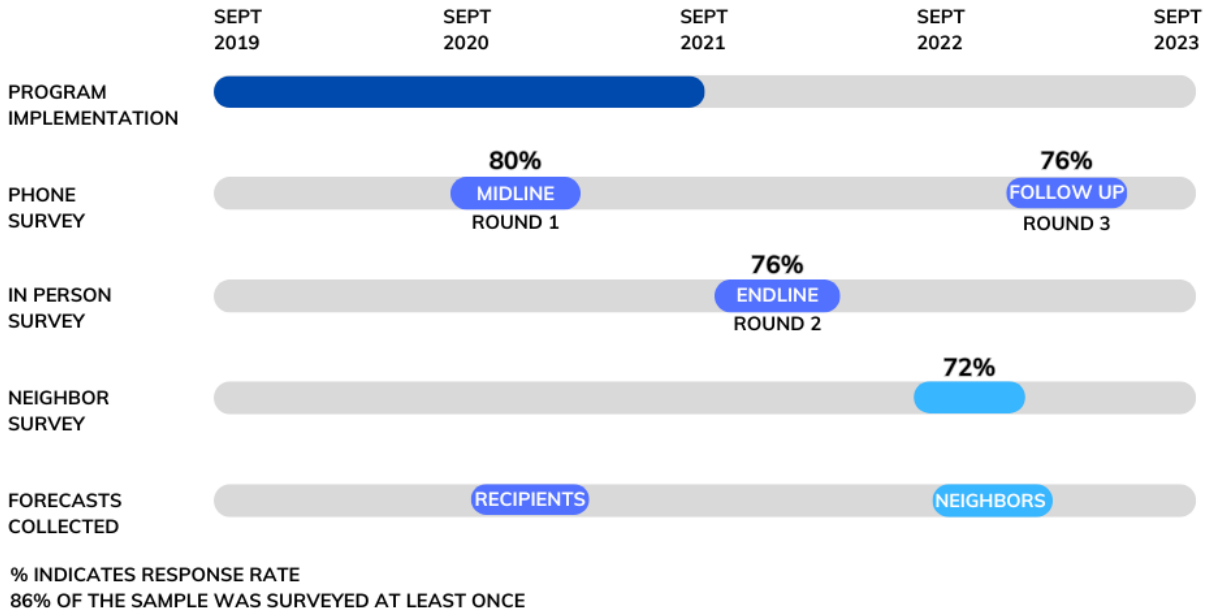


Table (1) Average Respondent, Household, and Shelter Characteristics

		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

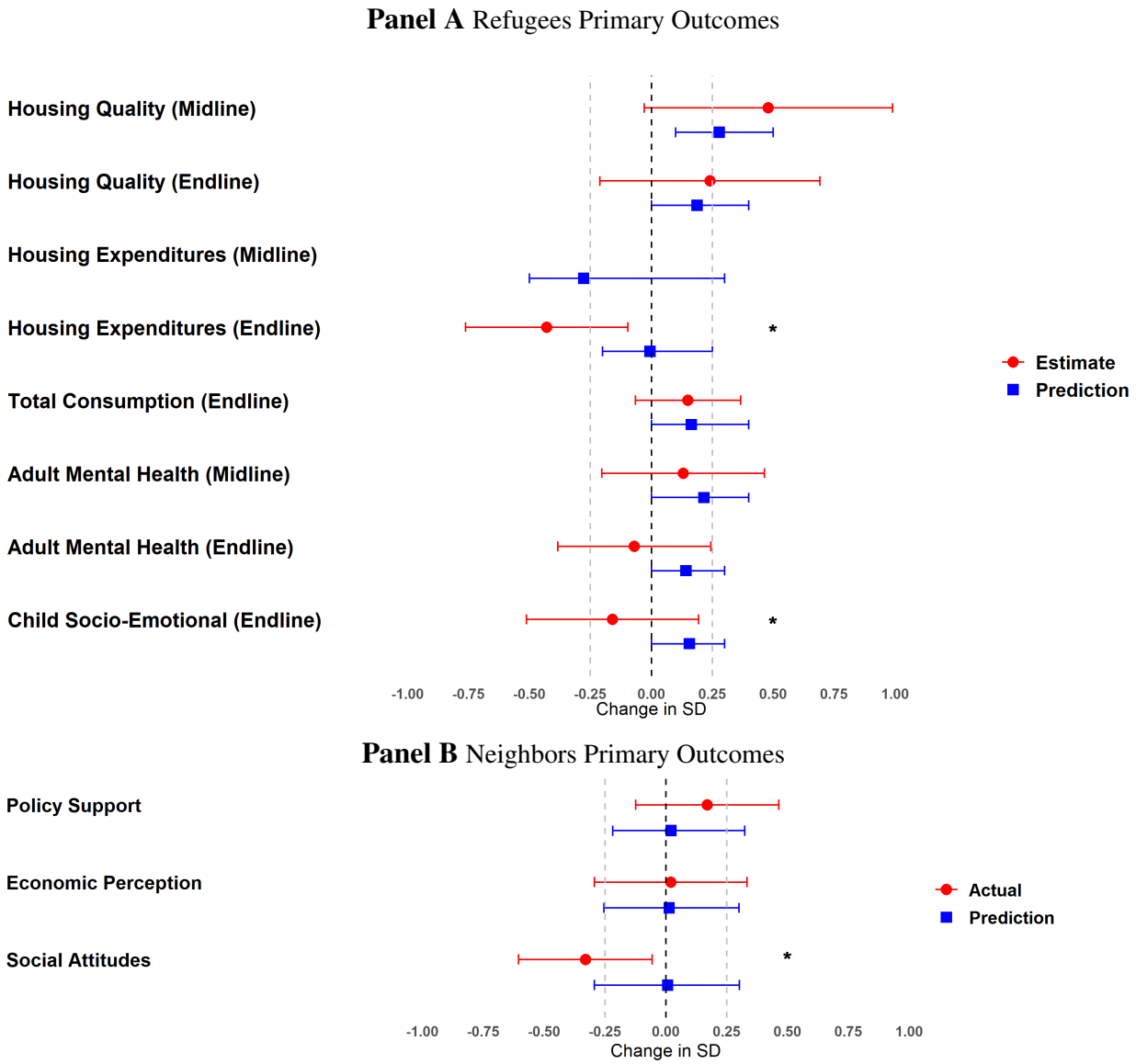
Notes: This table reports summary statistics of the distribution of predictions of the experimental treatment effects. Treatment effect predictions were elicited in terms of standard deviations. p10 and p90 refer to the 10th and 90th percentile of the distribution of predictions. We report here only 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
NRC Assistance in Last 12 Months (=1)	0.08**	(0.03)	[0.14]	-0.11***	0.09***	0.04	(0.19)	1,422
NRC 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.