



Image: A beneficiary who launched a bakery with support of the project shows off a cake she produced.

Impacts on Household Consumption for Cohort 1 of the Refugee Livelihoods Development Impact Bond

Laura Meyer, Evan Borkum, and Gray Collins

November 13, 2024

Estimated household consumption metric: 0.225 standard deviations

Summary of findings

About 23 months after disbursement of grants funded by the Refugee Impact Bond to Cohort 1 program participants, among whom two-thirds received a grant, average annual household consumption was 636 Jordanian Dinars (JOD) higher (\$897 in nominal terms; \$2,366 in purchasing power parity [PPP] terms; 0.22 standard deviations) than that of a matched sample of Cohort 3 participants who had just started the program. Accounting for household size, per-person household consumption was 116 JOD higher (\$164; PPP \$432; 0.13 standard deviations) for Cohort 1. Average annual household income was 674 JOD higher (\$950; PPP \$2,507; 0.24 standard deviations) for Cohort 1, driven by higher rates of business ownership and higher average take-home business incomes, and likely translated into higher expenditure on food and non-food goods and services. Findings were within the range of estimated impacts on consumption or expenditure and income for similar livelihoods plus cash grant interventions in low and middle-income countries and for cash or livelihoods interventions implemented previously in Jordan and Lebanon.

Introduction

The Near East Foundation's (NEF's) micro-enterprise creation and resilience-building program in Jordan includes training in life and business skills, cash grants to finance micro-enterprises, and additional support for these enterprises. The program, which serves vulnerable Jordanians and refugees, is being funded by a Development Impact Bond (DIB) and implemented through five *Siraj* centers across the country. The DIB-funded program included three cohorts that began training in April 2022, January 2023, and April 2024, respectively. Training participants develop a business plan and apply for the program's grants, which are awarded to successful applicants between four and five months after

their cohort starts training, on average. About two-thirds of program participants, or 3,400 in total across the three cohorts, will receive grants of between 400 and 700 JOD to start and/or grow their micro-enterprises.

Mathematica is conducting an evaluation of the impacts of the DIB program on household consumption and other outcomes related to social and economic wellbeing for participants in Cohort 1. The impact evaluation compares the outcomes of participants in Cohort 1 about 23 months after grant disbursement with those of a comparison group of participants in Cohort 3, who were just starting the program at that point in time. Although the timing of outcome measurement is linked to grant disbursement, the impact



evaluation includes both grantee and non-grantee participants from Cohort 1 to fully capture the impacts of the program. (About two-thirds of Cohort 1 participants are grantees.) Moreover, NEF has not finalized selection of Cohort 3 grantees, so including Cohort 1 grantees and non-grantees increases the comparability of the two cohorts.

To estimate these impacts, Mathematica and our local data collection partner Mindset conducted a survey between May and July 2024 with a sample of program participants from Cohorts 1 and 3. The survey measured a variety of outcomes linked to the program logic, including consumption of goods and services, household income, savings, and debt, their subjective sense of well-being, women’s social and economic empowerment, strategies to meet household needs in the face of limited food or financial resources, children’s school enrollment and attendance, and receipt of assistance and social protection.

The remainder of this brief summarizes the estimated impacts from the analysis of these survey data, with a focus on the household consumption metric that will be used to determine DIB payments. A comprehensive set of data tables are in Appendix A and technical details about the survey and the analysis are available in Appendix B. Appendix C contextualizes our findings with existing studies of cash and livelihoods support in





Jordan and other low- and middle-income countries. Impacts on additional outcomes, as well as impact estimates comparing Cohort 1 and Cohort 3 grantees, will be provided in the preliminary final evaluation report, expected in December 2024.

Household consumption metric

The consumption metric is defined as the average monetary value of annual household consumption. Household consumption includes four categories: food, non-food goods and services, durable goods (such as appliances, vehicles, and electronics), and housing (Figure 1). We calculated the value of each of these four consumption categories for each household on an annual basis, as summarized in Figure 1 and described in more detail in Appendix B. We then added these values for each household to estimate the household consumption metric.

The consumption metric described in this brief reflects two changes from the definition presented in the evaluation framework report (Borkum et al. 2022). First, we focus on total household consumption rather than consumption per household member. While the latter is a standard measure of overall household wellbeing in the literature on consumption measurement, the former better aligns with outcomes reported in the studies that were used to benchmark the payment thresholds for the DIB and with the fact that the

Figure 1. The household consumption metric incorporates four categories of annual household consumption

Food 	Non-food 
<p>The total value of food consumed by all household members, including food that was purchased, prepared at home, received as a gift, or as in-kind payment.</p>	<p>Household expenditures on transportation, personal care products, utilities, tobacco products, clothing, maintenance of home and vehicles, health care, education, and other miscellaneous items.</p>
Durable goods 	Housing 
<p>The estimated value a household derives from using durable goods like appliances, vehicles, and personal electronic devices.</p>	<p>The rent paid or market value of rent for a household’s current housing.</p>

same level of support was provided to all households regardless of size. Second, the framework report definition of household consumption included the annual value of debt repayments for debt accrued prior to the end of 2021, when Cohort 1 started the program. Debt repayments are not included in standard measures of consumption but were added to the DIB consumption metric at the DIB design stage because program participants were expected to be heavily indebted, and repayment of this debt was identified as a potentially important channel for improvements in wellbeing. However, we ultimately omitted debt repayments from the consumption metric because: (1) it was challenging to measure these repayments accurately given challenges with respondents recalling pre-2021 debt and distinguishing it from present day debt; (2) we were not confident that pre-2021 debt levels were similar for Cohort 1 and Cohort 3, given that we did not have data on those debt levels; and (3) Cohort 1 might have used grants and/or business income to pay off pre-2021 debt before the 1-year survey recall period. As we note later, the revised definition of the consumption metric driven by these considerations empirically results in a higher estimated impact relative to the definition in the evaluation framework report.

Findings

Our findings are based on an analysis sample of 757 Cohort 1 participants and 890 matched Cohort 3 participants. Overall, 810 Cohort 1 participants and 1,056 Cohort 3 participants completed the household survey. After conducting matching to account for differences between the samples in terms of their demographic characteristics and baseline socio-economic characteristics, we were left with an analysis sample of 757 Cohort 1 and 890 Cohort 3 participants (See Appendix B for more information

on the sampling and matching approaches.)

The Cohort 1 analysis sample comprised 88 percent women, 35 percent refugees, and 24 percent youth; 8 percent had a disability (Table A1a). Cohort 1 respondents' mean age was 37 years at the time of selection for the program, about 30 percent were heads of their household, and their mean household size was 5 people. About two-thirds had received a grant from the program, averaging 566 JOD (\$798, or PPP \$2,106).¹ The Cohort 1 analysis sample was broadly similar in characteristics to the full group of Cohort 1 participants, although the former included slightly greater proportions of women, Jordanians, adults, and households in the lowest baseline asset quartile (Table A1a). These modest differences were driven by a combination of differential match rates (because Cohort 1 respondents with less common characteristics were less likely to find an exact match in Cohort 3), as well as small differences in survey response rates.

The matching approach successfully resulted in well-balanced Cohort 1 and Cohort 3 analysis samples with similar demographic characteristics and baseline socio-economic characteristics (Table A1b). The largest differences were that Cohort 3 participants were 5 percentage points less likely to have a disability (because disabilities were rarer in Cohort 3 than Cohort 1) and 6 percentage points more likely to have post-secondary education (the analysis described below controls for educational attainment).

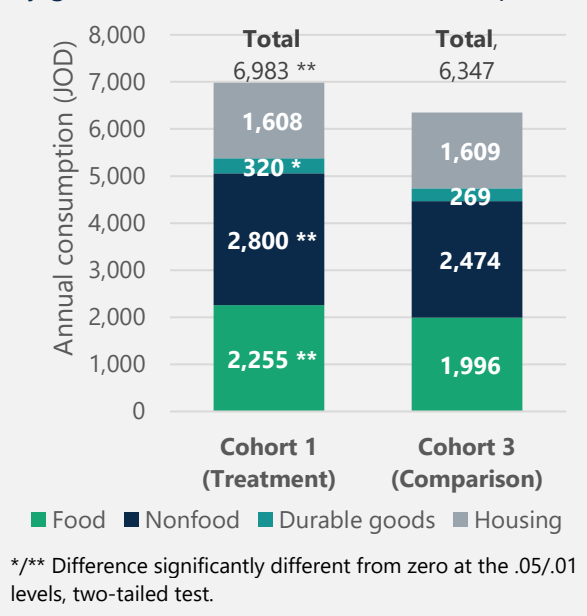
Impacts on the consumption metric

The program contributed to modest increases in total household consumption 23 months after grant disbursement, mainly driven by greater food and non-food consumption. The estimated average value of the revised consumption metric was 6,983 JOD per year for

¹ For clarity of presentation, we do not present conversions from JOD to dollars for the remainder of this brief. The nominal exchange rate is pegged, and JOD can be converted to dollars by multiplying by 1.41. One can also convert JOD to dollars using a purchasing

power parity (PPP)-adjusted exchange rate, which accounts for differences in the cost of living, by multiplying JOD by 3.72 (IMF 2024).

Figure 2. The positive impacts on annual household consumption were mainly driven by greater food and nonfood consumption



Cohort 1 households compared to 6,347 JOD for matched Cohort 3 households, a statistically significant, positive difference of 636 JOD, or 10 percent of the Cohort 3 average (Figure 2, Table A2a). In terms of the categories included in the consumption metric, Cohort 1 households had higher annual consumption of food (by 259 JOD), non-food goods and services (by 326 JOD), and durable goods (by 51 JOD) than Cohort 3 households, on average (Figure 2, Table A2a). The value of housing consumption was virtually identical across Cohort 1 and Cohort 3 households.

There were also positive, statistically significant impacts on the consumption metric for several subgroups, including female participants, adult and youth participants, Jordanians, and Irbid and Kufrsoum-based participants (Table A3). There

were also statistically significant impacts for three out of the four baseline asset quartiles, with the largest impacts occurring in the top two (wealthiest) quartiles. However, the differences in impacts for related subgroups were not statistically significant except for Jordanians versus refugees, providing little overall evidence of systematic differences in impacts by subgroup (results not shown).

We also examined how impacts on the consumption metric change based on the definition. The impact on total household consumption including debt repayments (which were included in the framework report definition) was lower (468 JOD, 0.12 standard deviations) due to lower reported repayments for pre-program debt by Cohort 1 relative to Cohort 3 (Table A2a). On a per-person basis (also included in the in the framework report definition), the impact was 116 JOD (0.13 SD) excluding debt repayments, and only 69 JOD (0.05 SD) and not statistically significant including debt repayments (Table A2b).² Overall, the revised definition of the consumption metric that is at the household level and excludes debt repayments—justified by the considerations described earlier—in practice results in a higher estimated impact relative to the definition in the framework report.

Impacts on household income, debt, and savings

Cohort 1 households reported higher average household incomes than Cohort 3, primarily driven by higher business incomes. Total self-reported annual household income for Cohort 1 households was 674 JOD (17 percent) higher than for Cohort 3, on average (Figure 3, Table A4).³

² We used the “adult equivalent” approach, which accounts for differences in consumption between adults and children and economies of scale when estimating per-person consumption.

³ Estimated total annual household consumption is substantially higher than estimated household income (for Cohort 3, 6,347 JOD versus 4,002 JOD). This is largely because the consumption metric includes non-expenditure items, specifically the estimated value of

consumption of durable goods as well as the estimated value of housing that is owned or used for free. Using a proxy measure for expenditure, which includes only the value of food consumption, non-food expenditures, and direct spending on housing, total household consumption for Cohort 3 decreases to 5,498 JOD—closer to total household income. Additionally, income is commonly underreported in low and middle-income

Most of this difference was driven by higher annual take-home incomes from household businesses, which were 498 JOD higher (more than double) in Cohort 1 than in Cohort 3, on average.⁴ This reflects that Cohort 1 participants were substantially more likely to report owning a business compared to Cohort 3 participants (63 versus 35 percent), and those with businesses had more businesses per household (1.2 versus 1.0). Cohort 1 participants also reported higher annual social assistance income (100 JOD) and wages (94 JOD) compared to Cohort 3 participants, although the latter difference was not statistically significant.⁵

About 23 months after grant disbursement, Cohort 1 households had high debt levels but few savings, on average. To further assess program households' financial status, we examined their reported debts and savings. About 85 percent of households in both Cohort 1 and Cohort 3 reporting having debts at the time of the survey, primarily from formal creditors, relatives, and friends (not shown). Average household debt was 3,447 for Cohort 1, which was 492 JOD (12 percent) lower than for Cohort 3, although this difference was not statistically significant because of the high variability in the debt measure (Figure 4, Table A4). Median debt levels were only about 850 JOD in Cohort 1 and 900 JOD in Cohort 3, which indicates that high average debt levels are driven by a relatively small fraction of households with heavy debt loads. Most of these heavily indebted households were Jordanian rather than

Figure 3. Positive impacts on annual household income were driven by take-home income from businesses

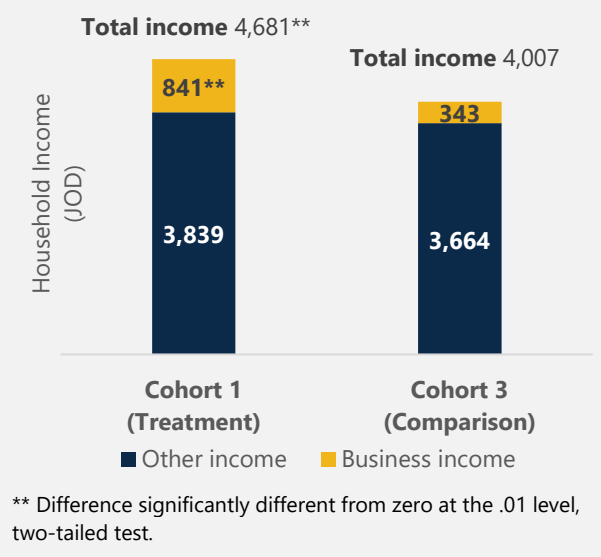
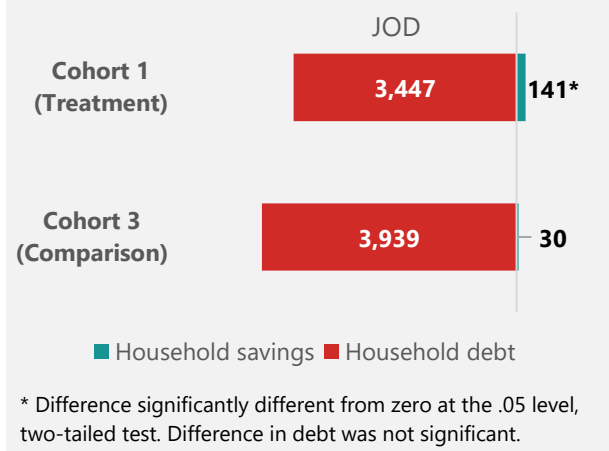


Figure 4. Program households had high levels of debt but few savings, on average



country contexts, especially among vulnerable populations who rely more on informal and seasonal work and may be reluctant to report income fully due to concerns about privacy, taxes, and eligibility for social protection programs (Deaton and Grosh 2000). This was a key reason for measuring consumption rather than income as a DIB metric.

Estimated total annual household income is similar to that reported in a recent nationally representative survey of Syrian refugee populations in Jordan (2,983 JOD for Cohort 3 refugees, versus 3,336 JOD in the national survey) (United Nations High Commission for Refugees 2023). Our proxy measure for expenditures is modestly higher than expenditures reported in that survey (4,925

for Cohort 3 refugees, versus 3,852 JOD in the national survey). We view these differences as plausible given differences in samples and measure definitions.

⁴ Take-home business income is defined as the amount taken out from household businesses to support family expenses, as opposed to business reinvestment.

⁵ Among Cohort 1 grantees, 85 percent reported having a business. The relatively high rates of business ownership by Cohort 3 although they had not yet received grants likely reflects that the survey captured businesses operated by other household members (about one in five Cohort 3 businesses) and that Cohort 3 participants might have started initial business activities in advance of grant receipt.

refugees and owed large amounts to formal creditors (not shown).

Fewer than 10 percent of households in Cohort 1 and Cohort 3 reported that they had any savings at the time of the survey. However, Cohort 1 households had modestly higher savings of 111 JOD compared to only 30 JOD among Cohort 3 households.

Discussion

The impacts on the consumption metric fall within the range of impacts that were used to set the thresholds for the DIB payments, which found impacts of between 0.07 and 0.38 standard deviations on expenditure and/or consumption, and impacts of between 0.12 and 0.30 standard deviations on income. Broadening our focus to include livelihoods-only or cash-only interventions from Jordan and Lebanon, we find that the impacts of the program were similar or slightly lower than other interventions, although these tended to focus on different populations or provide longer-term financial support to beneficiaries than NEF does.

Ultimately, there is limited rigorous evidence on the effects of livelihoods programming on income, consumption and/or expenditure in Lebanon, Jordan, or the broader Middle East and North Africa (MENA) region that enable direct comparisons of impacts or cost-effectiveness, especially for programs with an entrepreneurial rather than a jobs training focus. We hope that this study will help to contribute to filling an important evidence gap.

Next steps

In the final evaluation report, which we will submit a preliminary draft of by the end of 2024, we will expand on the findings in this brief by conducting additional analyses based on the survey data.

These include the following:

1. Impact estimates for additional outcomes related to social and economic wellbeing.
2. Descriptive statistics related to income generating activities, business financial metrics, and take-home business income for Cohort 1.
3. Impact estimates for matched Cohort 1 and Cohort 3 grantees.
4. An exploration of the potential implications of non-response on the impact estimates given the lower response rate for Cohort 1 relative to Cohort 3.⁶
5. Additional contextualization of our study findings in the available literature on livelihoods and/or cash transfers
6. Cost effectiveness estimates, pending the availability of cost data from NEF, to compare program costs to the estimated total value of the stream of household consumption impacts and calculate a return on investment; this would inform scale-up considerations and facilitate a comparison to other programs.

⁶ For example, we can use bounding approaches like Lee (2009), and compare outcomes for those who were

contacted easily for the survey to those who required additional follow-up.

Appendix A. Data tables

Table A1a. Representativeness of the Cohort 1 analysis sample

Characteristic	Cohort 1 population	Cohort 1 analysis sample	Difference
Demographic, household, and grant characteristics:			
Women (%)	82.4	88.4	-6.0
Youth (%)	30.5	23.5	7.0
Refugees (%)	30.4	35.0	-4.6
Mean age at the time of program intake (years)	35.1	36.8	-1.7
Head of household (%)	29.6	30.1	-0.5
Has a disability (%)	6.9	7.7	-0.8
Literate (%)	97.5	96.7	0.8
Mean household size at time of program intake	5.3	5.3	0.0
Education level (%)			
Less than secondary education	29.1	31.7	-2.6
Secondary school	41.6	42.9	-1.3
Post-secondary (technical or university)	29.2	25.4	3.8
Received a grant (%)	66.6	68.3	-1.7
Grant amount (for grantees only, JOD)	566	565	1
Baseline asset index			
Quartile 1 (%)	24.9	30.3	-5.4
Quartile 2 (%)	25.0	23.0	2.0
Quartile 3 (%)	25.1	23.9	1.2
Quartile 4 (%)	24.9	22.9	2.0
Baseline housing characteristics			
Owns home (%)	27.9	28.7	-0.8
Persons per room	1.9	1.9	0.0
Program site (%):			
Amman	26.2	25.0	1.2
Irbid	14.7	11.9	2.8
Kufrsoun	17.6	19.3	-1.7
Russeifa	15.0	15.9	-0.9
Zarqa	26.5	28	-1.5
Sample sizes	1,235	757	n.a.

Sources: Gender, refugee status, program site, disability status, and grant information are from NEF activity data. All other information is from NEF's vulnerability assessment.

Notes: n.a. = Not applicable. Youth are defined as being age 25 or under at the beginning of the program. The asset index is based on housing characteristics and durable goods ownership before Cohort 1 started the program, and is described in more detail in Appendix B. We do not test for statistical significance between the analysis sample and population because those groups are not mutually exclusive.

Table A1b. Baseline equivalence of the intervention and comparison samples after matching

Characteristic	Cohort 1 analysis sample	Cohort 3 analysis sample	Difference
Demographic and household characteristics			
Women (%)	88.4	88.4	0.0
Youth (%)	23.5	23.5	0.0
Refugees (%)	35.0	35.0	0.0
Mean age at the time of Cohort 1 program intake (years)	36.8	35.1	1.7**
Head of household (%)	30.1	31.5	-1.4
Has a disability (%)	7.7	2.9	4.8**
Literate (%)	96.7	98.1	-1.4
Mean household size at time of Cohort 1 program intake	5.3	5.3	0.0
Education level (%)			
Less than secondary education	31.7	30.2	1.5
Secondary school	42.9	38.8	4.1
Post-secondary (technical or university)	25.4	31.1	-5.7*
Baseline asset index			
Quartile 1 (%)	30.3	30.3	0.0
Quartile 2 (%)	23.0	23.0	0.0
Quartile 3 (%)	23.9	23.9	0.0
Quartile 4 (%)	22.9	22.9	0.0
Baseline housing characteristics			
Owns home (%)	28.7	27.6	1.1
Persons per room	1.9	1.7	0.2**
Location at time of Cohort 1 program intake (%)			
Amman	25.0	25.0	0.0
Irbid	11.9	11.9	0.0
Kufrsoum	19.3	19.3	0.0
Russeifa	15.9	15.9	0.0
Zarqa	28.0	28.0	0.0
Sample sizes	757	890	n.a.

*/** Difference significantly different from zero at the .05/.01 levels, two-tailed test using heteroskedasticity-robust standard errors.

Sources: Gender, age, refugee status, education level, and disability status are from NEF activity data. Other information was collected by NEF as part of the vulnerability assessment, with baseline asset ownership, housing information, household size, and location collected retrospectively from Cohort 3.

Notes: n.a. = Not applicable. Cohort 3 means and differences are estimated using coarsened exact matching weights. Youth are defined as being age 25 or under at the beginning of the program. The asset index is based on housing characteristics and durable goods ownership before Cohort 1 started the program, and is described in more detail in Appendix B. The asset index quartiles are based on a combination of housing information and assets at the time of cohort 1 beginning the program. The difference column may not exactly match the difference between group means due to rounding.

Table A2a. Impacts on total annual consumption for the full sample, overall and by consumption category

Sample	Cohort 1 treatment mean, μ_T (JOD)	Cohort 3 comparison mean, μ_C (JOD)	Difference, $\mu_T - \mu_C$ (JOD)	Standard deviation in comparison, σ_C (JOD)	Standardized difference, $(\mu_T - \mu_C)/\sigma_C$ (standard deviations)
Total household consumption					
Revised consumption metric:	6,983	6,347	636**	2,833	0.22**
Total household consumption, excluding debt repayments					
Total household consumption, including debt repayments	7,806	7,299	508**	3,738	0.14**
Direct consumption (proxy for expenditure) ^a	6,053	5,498	555**	2,387	0.23**
Total household consumption by category					
Food items	2,255	1,996	259**	965	0.27**
Non-food items	2,800	2,474	326**	1,776	0.18**
Durables	320	269	51*	451	0.11*
Housing	1,608	1,609	-1	614	0
Debt repayments	823	952	-129	1,823	-0.07
Sample sizes	757	890	n.a.	n.a.	n.a.

*/** Difference significantly different from zero at the .05/.01 levels, two-tailed test using heteroskedasticity-robust standard errors.

Source: Household survey.

Notes: n.a. = Not applicable. Comparison means and treatment-comparison differences are estimated using coarsened exact matching weights and regression adjusted with controls for household size and its square, age at the time of program entry for Cohort 1 and its square, the continuous asset index, and education level (less than secondary, secondary, or more than secondary). We conducted top- and bottom-coding separately by cohort for each consumption category to avoid outliers unduly influencing the findings. Food, nonfood, and housing were top and bottom coded to the 99th and 1st percentiles, respectively, by cohort. Durables goods were top coded to the 99th percentile by cohort. Debt repayments were top coded to the 95th percentile by cohort due to more extreme outliers. The difference column may not exactly match the difference between group means due to rounding.

^aDirect consumption includes the value of food consumption, expenditure on non-food goods and services, and rent payments; it excludes durable goods and estimated rent for owner-occupied housing, which are included in the primary consumption measure.

Table A2b. Impacts on adult equivalent consumption per-capita for the full sample, overall and by consumption category

Sample	Cohort 1 treatment mean, μ_T (JOD)	Cohort 3 comparison mean, μ_C (JOD)	Difference, $\mu_T - \mu_C$ (JOD)	Standard deviation in comparison, σ_C (JOD)	Standardized difference, $(\mu_T - \mu_C)/\sigma_C$ (standard deviations)
Per-capita consumption					
Per-capita consumption, excluding debt repayments	1,936	1,819	116**	871	0.13**
Per-capita consumption including debt repayments	2,154	2,085	69	1,109	0.06
Per-capita consumption by category					
Food items	621	568	53**	281	0.19**
Non-food items	763	698	66*	509	0.13*
Durables	87	75	11	121	0.09
Housing	465	478	-14	250	-0.05
Debt repayments	218	266	-48*	511	-0.09*
Sample sizes	757	890	n.a.	n.a.	n.a.

*/** Difference significantly different from zero at the .05/.01 levels, two-tailed test using heteroskedasticity-robust standard errors.

Source: Household survey.

Notes: n.a. = Not applicable. Adult equivalent calculations use the OECD equivalence scale which assigns a value of 1 to the first household member, 0.7 for each additional adult aged 14 or older, and 0.5 to each child. Comparison means and treatment-comparison differences are estimated using coarsened exact matching weights and regression adjusted with controls for household size and its square, age at the time of program entry for Cohort 1 and its square, the continuous asset index, and education level (less than secondary, secondary, or more than secondary). We conducted top- and bottom-coding separately by cohort for each per-capita consumption category to avoid outliers unduly influencing the findings. Food, nonfood, and housing were top and bottom coded to the 99th and 1st percentiles, respectively, by cohort. Durables goods were top coded to the 99th percentile by cohort. Debt repayments were top coded to the 95th percentile by cohort due to more extreme outliers. The difference column may not exactly match the difference between group means due to rounding.

Table A3. Impacts on total annual consumption excluding debt repayments (the revised consumption metric), by subgroup

Sample	Cohort 1 treatment sample size	Cohort 3 comparison sample size	Cohort 1 treatment mean, μ_T (JOD)	Cohort 3 comparison mean, μ_C (JOD)	Difference, $\mu_T - \mu_C$ (JOD)	Standard deviation in comparison, σ_C (JOD)	Standardized difference, $(\mu_T - \mu_C)/\sigma_C$ (standard deviations)
By gender							
Women	669	765	6,908	6,268	640**	2,760	0.23**
Men	88	125	7,555	6,976	578	3,275	0.18
By age							
Youth	178	281	7,410	6,433	976**	2,965	0.33**
Adult	579	609	6,852	6,310	542**	2,789	0.19**
By refugee status							
Refugees	265	288	5,173	4,959	214	1,923	0.11
Jordanians	492	602	7,958	7,100	858**	2,971	0.29**
By location							
Amman	189	203	7,286	6,936	350	3,108	0.11
Irbid	90	155	6,822	5,582	1240**	2,505	0.49**
Kufrsoun	146	191	7,212	6,364	848**	2,730	0.31**
Russeifa	120	146	6,883	6,259	624	2,439	0.26
Zarqa	212	195	6,681	6,189	492	2,889	0.17
By baseline asset quartile							
Quartile 1	229	301	5,434	5,001	433*	2,080	0.21*
Quartile 2	174	229	6,061	5,914	147	2,326	0.06
Quartile 3	181	179	7,493	6,650	844**	2,681	0.31**
Quartile 4	173	181	9,427	8,252	1174**	3,191	0.37**

*/** Difference significantly different from zero at the .05/.01 levels, two-tailed test using heteroskedasticity-robust standard errors.

Source: Household survey.

Notes: n.a. = Not applicable. Comparison means and treatment-comparison differences are estimated using coarsened exact matching weights and regression adjusted with controls for household size and its square, age at the time of program entry for Cohort 1 and its square, the continuous asset index, and education level (less than secondary, secondary, or more than secondary). We conducted top- and bottom-coding separately by cohort for each consumption category to avoid outliers unduly influencing the findings. Food, nonfood, and housing were top and bottom coded to the 99th and 1st percentiles, respectively, by cohort. Durable goods were top coded to the 99th percentile by cohort. Differences in impacts across related subgroups were not statistically significant except for the difference between Jordanians and refugees, which was statistically significant at the 0.05 level. The difference column may not exactly match the difference between group means due to rounding.

Table A4. Impacts on total annual income, and household savings and debt

Sample	Cohort 1 treatment mean, μ_T	Cohort 3 comparison mean, μ_C	Difference $\mu_T - \mu_C$	Standard deviation in comparison, σ_C	Standardized difference, $(\mu_T - \mu_C)/\sigma_C$ (standard deviations)
Total annual income					
Total annual household income (JOD)	4,681	4,007	674**	2,841	0.24**
Business ownership					
Owns a business (%)	63.4	35.2	28.2**	47.6	0.59**
Average number of businesses owned, among business-owning households	1.2	1.0	0.2**	0.14	1.46**
Annual income, by source					
Business income (JOD)	841	343	498**	856	0.58**
Income from wages (JOD)	2,583	2,489	94	2,644	0.04
Income from pensions	608	562	46	1,433	0.03
Social assistance income (JOD)	495	395	100*	725	0.14*
Remittances, family support, and income from assets (JOD)	154	219	-65	881	-0.07
Savings/Debt					
Household savings (JOD)	141	30	111*	356	0.31*
Household debt (JOD)	3,447	3,939	-492	7,938	-0.06
Sample sizes	757	890	n.a.	n.a.	n.a.

*/** Difference significantly different from zero at the .05/.01 levels, two-tailed test using heteroskedasticity-robust standard errors.

Source: Household survey.

Notes: n.a. = Not applicable. Comparison means and treatment-comparison differences are estimated using coarsened exact matching weights and regression adjusted with controls for household size and its square, age at the time of program entry for Cohort 1 and its square, the continuous asset index, and education level (less than secondary, secondary, or more than secondary). We conducted top- and bottom-coding to avoid outliers unduly influencing the findings. Household savings and income variables were top coded to the 99th percentile by cohort. Household debt was top coded to the 95th percentile by cohort due to more extreme outliers. Take-home business income is defined as the amount taken out from household businesses to support family expenses, as opposed to business reinvestment. The difference column may not exactly match the difference between group means due to rounding.

Appendix B. Technical details for the impact evaluation

Impact evaluation design

The design of the evaluation leverages the staggered rollout of the intervention across cohorts to compare the outcomes of participants in Cohort 1 (including grantees and non-grantees) about 23 months after grant disbursement with the situation of a matched sample of those who recently entered Cohort 3. This matched comparison design aims to limit the possibility that differences in outcomes between the two cohorts are driven by differences in participant characteristics, making it more plausible to attribute these differences to the impact of the program. Specifically, because Cohorts 1 and 3 were selected for the program using a similar approach, we expect them to be broadly similar in terms of vulnerability and unobserved characteristics related to program participation and outcomes (for example, motivation and entrepreneurial spirit).

To further limit the possibility that differences in outcomes between the two cohorts are driven by differences in observed demographic and economic participant characteristics, we implemented coarsened exact matching (CEM) (Iacus et al. 2012). This method divides members of the two cohorts into mutually exclusive strata, or groups, defined by a combination of participant demographic characteristics and self-reported household assets and characteristics in late 2021 when Cohort 1 was selected.⁷ For example, one possible stratum is female Jordanians who were under 25, received the program at the Irbid *Siraj* center, were in the lowest quartile of the sample in terms of household assets in late 2021, and had a household size of 3 or fewer members at the time. We then reweight the comparison group (Cohort 3) so that its distribution across strata is identical

to that of the treatment group (Cohort 1), making it as similar as possible to the treatment group in terms of the matching characteristics. Thus, all comparison group observations in the same stratum receive the same weight. In this way, the comparison group is adjusted so that it is as similar as possible to the treatment group in terms of the matching characteristics, enabling us to produce credible impact estimates.

Sampling approach

The CEM approach is applied to a sample of survey respondents with data on the relevant characteristics and outcomes. It typically results in some sample loss because respondents in CEM strata that have only treatment or control observations (that is, unmatched individuals) are dropped. In this study, we planned to survey a sample of participants rather than all participants given our sample size targets and wanted to minimize subsequent sample loss due to unmatched respondents at the analysis stage. Therefore, we conducted an initial stage of matching before sampling that was intended to focus the sample on individuals who were more likely to be matched, while recognizing that the final matching could only be conducted once we knew who in the sample responded to the survey.

Specifically, we applied the following steps to select the survey sample:

1. We conducted initial CEM using the following matching characteristics: gender, refugee status, youth status (25 years or younger, or older than 25), program site, household size (3 or less, 5–8 or 9 or more) education level (less than secondary, secondary, or more than secondary), and asset index quartile. We

⁷ We obtained most of the information about matching characteristics from NEF's vulnerability assessment, which was used to screen participants shortly before they were selected (between February and April 2022 for Cohort 1 and between January and February 2024

for Cohort 3). For characteristics that were not time invariant, such as household assets, NEF asked these questions retrospectively for Cohort 3 about their situation in late 2021.

selected these matching variables to optimize baseline balance while retaining as much of the sample as possible. To estimate the asset index, we conducted a principal components analysis based on pre-program housing characteristics and durable goods owned at the time when Cohort 1 was starting the program.⁸ Out of all program participants in the two cohorts, we matched 1,189 of the 1,235 participants in Cohort 1 (96 percent) and 2,231 of the 2,472 participants in Cohort 3 (90 percent) using this approach.⁹

2. We included all 1,189 matched Cohort 1 participants (1,189) in our survey sample because the cohort was smaller than Cohort 3 and because we anticipated higher levels of nonresponse given that these participants had left the program almost two years prior to the survey.
3. We then randomly selected a sample of 929 out of the 2,231 matched Cohort 3 participants as our primary sample, allocating this sample across strata defined by mutually exclusive combinations of gender, refugee status, youth status, and program site, in proportion to strata's share of the population of participants.¹⁰ The rest of the matched Cohort 3 participants (1,302) served as replacements in the case of nonresponse. Specifically, when participants from the primary sample were unreachable, unwilling to participate, or unavailable, we selected replacements from a randomly sorted list of potential

replacements from the same stratum. If replacements from that stratum were exhausted, we drew replacements from the most similar stratum with replacements remaining.

Data collection approach

Survey development. Our survey development approach sought to ensure that the impact survey captured as much of respondents' household consumption as possible without become overly burdensome. As a starting point, we used data tables from the Jordan Household Expenditure and Income Survey 2017–18 to identify an initial list of consumption items that were likely to comprise the largest share of consumption. Specifically, we included in our initial list the smallest number of items that collectively contributed 90 percent of total consumption. However, recognizing that program participants might differ from the typical Jordanian household, we adjusted this initial list by: (1) cross-referencing the consumption survey conducted with refugees in Jordan in 2021 for UNHCR's vulnerability assessment framework; and (2) using what we learned during survey pretesting and piloting. These changes to the list included adding, removing, combining, or disaggregating items to make the final list more relevant to the consumption of program participants and minimize respondent burden while maintaining accuracy.

Training and implementation. The data collection team participated in a 3-day in-person training on the survey. The training included training on protection of human subjects and

⁸ We used the initial list of potential asset index variables identified in the evaluation framework report as a starting point, and excluded durable goods that were extremely common or extremely rare, as well as housing characteristics that showed a weak relationship with self-reported household expenses or had limited variation. We applied the principal components analysis to Cohort 1 and used it to predict the index for Cohort 3. We then divided all participants into asset index quartiles based on the Cohort 1 distribution.

⁹ These numbers reflect the sample after we corrected for duplicates across Cohorts 1 and 3 and dropouts from Cohort 3.

¹⁰ These are different from the CEM strata and were intended to broadly ensure the representativeness of the sample and provide Mindset with a primary and replacement sample using an approach they were familiar with from the income-generating activity surveys conducted as part of the DIB.

vulnerable populations, a detailed review of the survey questions, and practice conducting interviews using a series of respondent scenarios. Following training, we then conducted a day of piloting with both Cohort 1 and Cohort 3 participants.

Data collection occurred between late May and late July 2024, a mean of 22.6 months following grant distribution for Cohort 1, and between 1 and 3 months following the beginning of training—and prior to grant disbursement—for Cohort 3. Twenty-three different enumerators conducted the survey in teams of two or three, each with the support of a logistical coordinator who coordinated with potential respondents while in the field. The study used a verbal informed consent statement and procedure approved by the King Hussein Cancer Center Institutional Review Board (IRB) in Jordan and Health Media Lab IRB in the United States. Surveys lasted a mean of almost 50 minutes.

Sample sizes and response rates. Overall, 810 Cohort 1 respondents and 1,056 Cohort 3 respondents completed the survey. We had initially planned to have the same number of respondents (about 930) in both cohorts but adjusted to include more in Cohort 3 when it became clear that we would not reach this target in Cohort 1 given challenges with the response rate. The final response rate was 68 percent for Cohort 1 and 88 percent for Cohort 3. Achieving a higher response rate for Cohort 1 almost two years after the end of the program using the available contact information proved challenging, despite several measures we took to try to increase it. First, we implemented a systematic tracking effort to reach and update contact information via SMS and phone for Cohort 1 participants at two time points between the 10-month income-generating activity survey and the 23-month impact survey. During the survey, we also worked closely with NEF and *Siraj* centers to reach out to grantees, encourage participation in

the survey, and obtain updated information about their availability and willingness to participate. Finally, we conducted a small number of surveys by phone (about 2 percent of the total for Cohort 1) for respondents who had moved to other communities within Jordan or who were unable or unwilling to participate in person.

Ultimately, 379 sampled program participants in Cohort 1 were unreachable, or unable or unwilling to participate (Table B1). Of these, 217 were unreachable by phone, in large part due to disconnected phone numbers. 143 participants refused to participate, and an additional 24 were unreachable or incomplete in the field. 33 participants had migrated to other countries, and an additional 4 were incarcerated or deceased. 117 Cohort 3 participants who were sampled initially or added from the replacement list also did not respond to the survey. The reasons were mostly similar to Cohort 1, except that they had lower rates of out-migration.





Table B1. Final statuses of survey non-respondents

Status	Cohort 1	Cohort 3
Unreachable by phone	176	64
Refused to participate	143	40
Unavailable due to migration, incarceration, or death	36	2
Unreachable or incomplete in the field	24	11
Total incomplete	379	117

Additional details on the calculation of the consumption metric and other outcomes

As mentioned earlier, the consumption metric is defined as the average monetary value of annual household consumption. Table B2 summarizes the data collected and the estimation approach for each category of consumption.

Table B2. Details on calculation of the consumption metric

Category	Data collected in the household survey	Calculation approach
 Food	<ul style="list-style-type: none"> Quantity of food consumed by household members in the previous seven days The quantities and total costs of food purchased in the prior 30 days 	<ol style="list-style-type: none"> Calculate typical unit prices for each type of food, survey location and cohort, based on survey data Apply the unit price to the quantity consumed to estimate the total value of weekly consumption per food item Sum across all goods and multiply weekly consumption by the number of weeks per year to produce a total annual estimate
 Non-food	<ul style="list-style-type: none"> Expenditure on transportation, personal care products, utilities, tobacco products, clothing, maintenance of home and vehicles, health care, education, and other miscellaneous items Reporting periods varied between 1, 3, and 12 months depending on the item and typical consumption patterns in Jordan 	<ol style="list-style-type: none"> Calculate annualized values by category Sum across annualized estimates to produce a total annual estimate
 Durable goods	<ul style="list-style-type: none"> Purchase cost of each durable good owned Year of purchase Respondents' estimate of the current market value of the good 	<ol style="list-style-type: none"> Convert purchase costs to 2024 values using the Jordanian Consumer Price Index Use the difference between the present value of the purchase cost and the current value to estimate the annual depreciation rate for each type of good Estimated the value the household derives from the good (intuitively, the cost at which they might be willing to rent it out) as the current value plus interest minus annual depreciation.
 Housing	<ul style="list-style-type: none"> Rent payments (annual or monthly) Respondent-estimated cost to rent current home, if owned or used for free 	Convert actual or estimated rent to annual values to produce a total annual housing cost

We calculated additional outcomes related to households' financial situations as follows:

/ **Income:** We collected self-reported data on monthly income from household businesses, employment, assets and pensions, social assistance, and remittances and family support, and aggregated across sources to estimate total annual household income.

/ **Debt:** We collected self-reported data on households' current debt levels from various formal and informal sources such as banks, microcredit institutions, and family members, and aggregated to produce an estimate of total household debt load.

/ **Savings:** We collected self-reported data on households' current savings held in various modes such as banks, cash, jewelry, and savings groups, and aggregated across sources to estimate total household savings.

Analysis approach

We rematched the respondents who completed the survey using CEM based on the characteristics shown in Table B3. Like in the initial matching before sampling, we selected these characteristics to optimize baseline balance using CEM while

retaining as much of the sample as possible.¹¹ Out of all survey respondents, 757 out of the 810 respondents in Cohort 1 and 890 out of the 1,056 respondents in Cohort 3 were matched; these comprised the final analysis sample.

We then used regression analysis to estimate impacts by comparing outcomes in the matched treatment and comparison groups by estimating the following regression equation using ordinary least squares:

$$Y_i = \alpha + \beta_1 T_i + \beta_2 X_i + \varepsilon_i$$

where Y_i is the outcome for participant i ; T_i is an indicator for treatment, equal to 1 for the treatment group (Cohort 1) and 0 for the comparison group (Cohort 3); and ε_i is a random error term. Although our matching approach minimized pre-existing differences between treatment and comparison groups, we included a set of control variables, X_i , in the model to account

for any remaining imbalance between the two groups (Ho et al. 2007; Stuart 2010). X_i consists of indicators for categorical matching variables (respondent gender, refugee status and location at baseline), continuous versions of the remaining matching variables (age, age squared, household size, household size squared, and the continuous asset index), as well as education level (less than secondary, secondary, or more than secondary), which we did not include in the final matching because it led to additional sample size loss without improving overall baseline balance.¹² The coefficient, β_1 , provides the regression-adjusted estimate of the impact of the program, which is the difference in outcome means between the treatment and comparison groups after controlling for potential confounders that may contribute to the difference.

Table B3. Matching characteristics for CEM

Demographic characteristics	Asset index (four quartiles)	
	Housing characteristics	Durable goods owned
<ul style="list-style-type: none"> • Gender • Refugee status (refugee versus Jordanian) • Youth (25 years or younger versus older than 25 at the time that Cohort 1 was selected) • Program site (five sites) • Household size (3 or less; 4-7; 8 or more) 	<ul style="list-style-type: none"> • Owns home • Number of persons per room 	<ul style="list-style-type: none"> • Car • Computer • Air conditioner • Microwave • Gas/electric heater • Water heater/cooler • Electric fan • Vacuum cleaner • Clothes iron • Television

Note: Information on matching characteristics was obtained from NEF’s vulnerability assessment, which participants complete before entering the program. All matching characteristics were measured as of late 2021, when Cohort 1 was entering the program; NEF obtained some of this information retrospectively for Cohort 3 when this cohort completed the vulnerability assessment in early 2024.

¹¹ The estimates were not sensitive to including disability status and postsecondary education—the two characteristics where there was a modest baseline imbalance—in the matching process. This would have

ensured balance along these characteristics but led to sample size loss.

¹² The estimates were not sensitive to including these control variables.

Appendix C. Comparison to findings from the literature

There is limited rigorous evidence on the effects of livelihoods programming on income, consumption and/or expenditure in Lebanon, Jordan, or the broader Middle East and North Africa (MENA) region that enable direct comparisons, especially for programs with an entrepreneurial rather than a jobs training focus.

The impacts on the consumption metric (excluding debt repayments) are within the range of impacts found in the reference studies that were used to set the thresholds for the DIB payments (Table C.1). These studies from Sub-Saharan Africa and South Asia, which were used for benchmarking because they had both a cash grant and training component, had impacts on consumption or expenditure between 0.07 and 0.38 standard deviations (compared to our estimated impact of 0.22 standard deviations on consumption) and impacts of between 0.12 and 0.30 standard deviations on income (compared to our estimated impact of 0.24 standard deviations).

Broadening our focus to include livelihoods-only or cash-only interventions from Jordan and Lebanon, we find the following relevant comparisons to our estimated impacts, which in percentage terms were 10 percent for consumption and 16 percent for income:

- / A cash-for-work program implemented in a Jordanian refugee camp increased income by 23 percent, but only among semi-skilled workers (Lombardini and Mager 2019).
- / A long-term (16-22 months) World Food Program cash transfer program targeting Syrian refugees in Lebanon increased total household expenditures by 20 percent, including significant increases in food expenditures (Chaaban et. al. 2020).
- / A randomized wage voucher program in Jordan targeting recent female community college graduates (who are likely different from our

study population in many ways) showed initial positive effects on employment and income (17 percent among employed respondents) that faded once the voucher program ended (Groh et. al. 2012). Additional study arms that provided training or both training and vouchers did not show any significant effects.

In general, we consider our findings to be within the range of estimated impacts on consumption or expenditure, and income, from the limited existing literature for similar interventions outside MENA and for cash and livelihoods interventions in Jordan and Lebanon.

Table C.1. Summary of reference studies used to benchmark the consumption metric

Country	Program name	Program description	Impacts on enterprise profits (SDs)	Impacts on income (SDs)	Impacts on household consumption (SDs)
Uganda	Youth Opportunities Program (YOP)	Group training and business start-up support, whereby groups of youth in conflict-affected areas applied for grants averaging \$382 per person to start nonagricultural businesses.	0.16 (2 years) 0.17 (4 years)	<i>Not available</i>	0.18 (4 years)
Liberia	Action on Armed Violence (AoAV)	Residential coursework for male former fighters in agriculture and animal husbandry, along with counseling, life skills classes, and a package of tools and supplies to start business, valued at \$125.	No effect after 14 months	0.12 (14 months)	<i>Not available</i>
Sri Lanka	Start and Improve Your Business (SIYB)	A widely used entrepreneurship training course from the International Labour Organization that helps participants select and operationalize feasible business ideas. Training was for urban women and included cash grants of \$129.	0.16 (7-8 months) 0.18 (15-16 months) No effect after 2 years No effect for potential new business owners at any time point	No effect after 7, 15, and 24 months	<i>Not available</i>
Uganda	Start and Improve Your Business (SIYB)	Similar to the program in Sri Lanka (above), with cash grants of \$200.	No effect after 9 months	<i>Not available</i>	No effect after 9 months
Uganda	AVSI Women’s Income Generating Support (WINGS)	Business skills training for rural women, cash grants of \$150 to support businesses, and ongoing support through supervision and self-help group formation.	<i>Not available</i>	0.30 (16 months)	0.38 (16 months)
Bangladesh	BRAC’s Targeting the Ultra Poor (TUP) Program	Intensive skills training for rural women related to livestock businesses. Participants chose among six livestock packages, to the value of \$140.	<i>Not available</i>	0.15 (2 years) 0.17 (4 years)	0.07 (2 years) 0.10 (4 years)

Source: Keaveney et al., 2018

References

- Borkum, Evan, Paolo Abarcar, Laura Meyer, and Matt Spitzer. "Jordan Refugee Livelihoods Development Impact Bond Evaluation Framework." Washington, DC: Mathematica, 2022.
- Chaaban, Jad, Nisreen Salti, Hala Ghattas, Wael Moussa, Alexandra Irani, Zeina Jamaluddine, and Rima Al-Mokdad. "Multi-purpose cash assistance in Lebanon: Impact evaluation on the well-being of Syrian refugees." Cairo, Egypt: Economic Research Forum, 2020. Available at <https://erf.org.eg/publications/multi-purpose-cash-assistance-in-lebanon-impact-evaluation-on-the-well-being-of-syrian-refugees/>.
- Deaton, Angus and Grosh, M. "Consumption", in "Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study, M. Grosh and P. Glewwe, Editors. 2000. Washington, DC: The World Bank, 2000.
- Groh, Matthew, Nandini Krishnan, David J. McKenzie, and Tara Vishwanath. "Soft skills or hard cash? The impact of training and wage subsidy programs on female youth employment in Jordan." World Bank Policy Research Working Paper 6141 (2012). Washington, DC: The World Bank. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2116139.
- Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis*, vol. 15, no. 3, 2007, pp. 199-236.
- Iacus, Stefano M., Gary King, and Giuseppe Porro. "Causal inference without balance checking: Coarsened exact matching." *Political Analysis*, vol. 20, no. 1, 2012, pp. 1-24.
- International Monetary Fund "Implied PPP Conversion Rate, 2024." Available at <https://www.imf.org/external/datamapper/PPPEX@WEO/OEMDC/IRQ>.
- Keaveney, Erika, Mike Duthie, and Andrea Hur. "Development Impact Bond (DIB) for Improving Livelihood for Syrian Refugees and Vulnerable Populations in Jordan and Lebanon: Evaluation Design Report." Arlington, VA: Social Impact, 2018.
- Lee, David S. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies*, vol 76, no. 3, 2009, pp.1071-1102.
- Lombardini, Simone, and Franziska Mager. "Livelihoods in the Za'atari Camp: Impact evaluation of Oxfam's Cash for Work activities in the Za'atari camp (Jordan)." Oxford, UK: , Oxfarm, 2019. Available at <https://oxfamilibrary.openrepository.com/bitstream/handle/10546/620883/er-livelihoods-za%27atari-camp-cash-for-work-221019-en.pdf>.
- Stuart, Elizabeth A. "Matching methods for causal inference: A review and a look forward." *Statistical Science: a Review Journal of the Institute of Mathematical Statistics*, vol. 25, no. 1, 2010, pp. 1-21.
- United Nations High Commissioner for Refugees (UNHCR). "Socio-economic situation of refugees in Jordan Q2 2023". Amman, Jordan: UNHCR, 2023. Available at <https://data.unhcr.org/en/documents/details/103118>.