Geographic poverty targeting in social protection programs: Evidence from a nationwide policy experiment

Onur Altındağ, Stephen D. O'Connell, and Rim Achour

Abstract

How much should program impacts vary across places when cash transfers are implemented at national scale? We study this question using a nationwide experiment embedded in a humanitarian cash transfer program for Syrian refugees in Lebanon that randomizes households to alternative district budget allocation rules. While allocation rules shift resources toward demographically distinct households, average improvements in poverty outcomes are similar across households prioritized by these rules. In contrast, program impacts vary sharply across locations, with geographic heterogeneity dominating differences generated by targeting. Qualitative evidence links this variation to locally binding constraints, including housing obligations, accumulated debt, and institutional frictions.

Keywords: poverty targeting, poverty measurement, social protection, antipoverty programs, unconditional cash transfers, refugees, forced displacement, Lebanon.

JEL Classification: I38, I32, O12, D74

Contact Information Altındağ: Bentley University Department of Economics, IZA Institute of Labor Economics, and Economic Research Forum, oaltindag@bentley.edu. O'Connell: Emory University Department of Economics and IZA Institute of Labor Economics, soconnell@emory.edu. Achour: United Nations High Commissioner for Refugees, achour@unhcr.org.

Disclaimer: This study was approved by Emory University Institutional Review Board under protocols #3157 and #4634. A detailed pre-analysis plan was registered at https://www.socialscienceregistry.org/trials/9725.

Acknowledgement: We thank Zeynep Balcıoğlu and Aytuğ Sasmaz for their support with fieldwork and qualitative research. We also appreciate the input and guidance provided by many members of the UNHCR and WFP Lebanon staff throughout the planning, execution, and analysis phases, including Louise Abellard, Susanne Carl, Marco Principi, Simon Renk, and Milos Terzan. We are grateful for insightful comments and suggestions from Matthew Gudgeon, Olivier Sterck, Jade Siu, and conference and seminar participants at IPA, Tufts University, DevSouth, Northeastern University, the 2023 WFP Impact Evaluation Forum, The World Bank, and The Graduate Center, CUNY. The views and conclusions expressed in this study are those of the authors and do not necessarily reflect the positions or policies of any of the organizations involved.

1 Introduction

Unconditional cash transfers (UCTs) are among the most widely used social welfare programs worldwide, with well-documented impacts on household consumption, income, labor supply, school enrollment, food security, and psychological well-being (Crosta et al., 2024). In many developing countries, limited data on household income leads governments to rely on proxy means tests (PMTs) when determining eligibility. When programs are implemented nationally, however, policymakers face a more fundamental challenge that arises even before household-level targeting: how to allocate a program's budget across localities. This decision is typically guided by poverty maps, even though little is known *a priori* about how program impacts vary across markets. These knowledge gaps motivate the core question of this paper. How much heterogeneity should we expect in program effects across localities when a transfer program is implemented at scale, and why do these differences arise? Understanding how and why impacts vary across settings is central for designing scalable and effective antipoverty programs.

In this study, we present the results of a year-long, nationwide natural field experiment launched in mid-2021 that was designed to test for heterogeneity in program effects arising from alternative geographic allocation criteria in the program's design. The experiment was embedded within an ongoing, at-scale social assistance program implemented by humanitarian agencies in Lebanon during the 2021–2022 programming cycle. Prior to the experiment, transfers were allocated through a two-tiered system that determined eligibility and transfer amounts based on household vulnerability and budget constraints. First, the national budget was divided across the 26 districts in proportion to each district's share of the national poor, where poverty was defined as being below a threshold in monthly expenditure per capita. Second, within each district, a proxy means test identified the poorest households for eligibility.

During the study year, this system was modified to allow an experimental test of alternative ways of distributing the national program budget across districts. Throughout the paper, we refer to these alternatives as *geographic budget allocation rules*. Each rule specifies how the fixed national program budget is allocated across districts prior to household-level eligibility determination using the proxy means test.

For the study, implementing agencies divided the national budget equally across four experimental arms. Each arm distributed the same total dollar amount and supported the same number of households under a uniform per-household transfer value, but each relied on a different geographic budget allocation rule for distributing funds across districts. These rules prioritized distinct poverty metrics, described below, which generated meaningful differences in the total value of transfers districts would receive under alternative allocations due to the substantial variation in economic conditions and how poverty is experienced and expressed across areas. Under the experimental design, the four geographic budget allocation rules operated in parallel. *Households* within each district were then randomly assigned to one of the allocation rules, which determined the budget level available for their district and among the households assigned to that specific rule. This created exogenous variation in eligibility and transfer amounts for otherwise similar households living in the same district. From an aggregate perspective, our study design is an "A/B" test (or, given four treatment regimes, an "A/B/C/D" test) with no formal control group (Hanna and Karlan, 2017). As we describe below, however, we observe and exploit experimental variation in transfer eligibility among a subsample of households with marginal eligibility across rules in order to recover and analyze typical experimental estimands.

We designed the study to be able to simulate the application of each allocation rule to all households *ex ante* so that we could observe the transfer each household would receive under each of the four counterfactual budgets. We can then identify households whose transfer amount changes across rules and those for whom it does not. We refer to these groups as *marginal* and *inframarginal* beneficiaries, respectively. In our setting, the choice of poverty metric used for district-level budget allocation leaves transfer amounts unchanged for roughly 65% of households. The remaining 35% are marginal beneficiaries who receive different assistance amounts under at least one counterfactual allocation rule, and are therefore the households whose transfers are directly affected by changes induced by alternative geographic budget allocation rules. Consistent with our pre-registered analysis plan, marginal beneficiaries were then oversampled and surveyed to ensure adequate statistical power for estimating prespecified effect sizes.

The experimental design allows us to estimate program effects across multiple domains and to characterize treatment effect heterogeneity. We first estimate the average program effect for marginal beneficiaries at the national level. We find that the program improves all prespecified poverty indicators, with statistically significant effects ranging from 0.06 to 0.17 standard deviations. We also observe a decline in the share of children who are out of school of approximately 0.11 standard deviations. Taken together, these results indicate economically meaningful but quantitatively modest impacts.

Second, we compare program effects across the four groups of households marginal to each of the alternative budget allocation rules. The program reduces economic deprivation across all marginal beneficiary groups, with estimated effects ranging from a statistically insignificant 0.02 standard deviations to 0.21 standard deviations across the prespecified poverty outcomes. Despite this variation in point estimates across allocation rules, we do not detect economically meaningful or statistically significant differences in treatment effects across the marginal beneficiaries associated with any *particular* allocation rule. These results inform policymakers about the tradeoffs involved in selecting among alternative budget allocation rules, suggesting limited scope for improving aggregate outcomes through targeting-induced budget reallocation alone.

Third, the experimental design and large sample size allow us to estimate location-specific treatment effects across all 26 administrative districts. Program impacts vary substantially across localities, suggesting that local context is a first-order correlate of program effectiveness, even under uniform implementation and common design parameters. Positive and statistically significant effects are concentrated in a subset of districts, while estimates in many other districts are smaller and often statistically indistinguishable from zero. Importantly, districts that experience larger improvements in one outcome also tend to experience larger improvements across other poverty indicators, suggesting that local constraints operate jointly across dimensions of deprivation. Relative to the limited differences observed across alternative budget allocation rules, these findings point to geographic variation in heterogeneous treatment effects as a more consequential margin than targeting for understanding program outcomes.

We further examine treatment effect heterogeneity at the household level using Chernozhukov et al. (2025)'s flexible machine learning framework that allows treatment effects to vary with observed pre-intervention characteristics. This analysis confirms the strong geographic patterns documented in the district-level results, while also revealing systematic heterogeneity associated with household demographics, baseline vulnerability, and prior exposure to assistance. Program impacts are concentrated among more economically vulnerable households, including female-headed and single-parent households, those with lower baseline consumption, and households with weaker attachment to formal labor markets, whereas treatment effects are lower for households with higher baseline resources and more prime-age healthy men in the household.

We complement the quantitative analysis with a structured qualitative mapping exercise based on focus group discussions with 114 participants drawn from the experimental sample of marginal beneficiaries. Using a standardized coding framework applied to individual text segments from these discussions, we draw on recent methodological advances in open-text analysis (Ferrario and Stantcheva, 2022; Stantcheva, 2024) and combine large-language-model-assisted classification with systematic human verification. This approach allows us to identify binding constraints, coping strategies, and local shocks in a consistent manner.

The qualitative evidence sheds light on the mechanisms underlying the substantial heterogeneity documented in the experimental results. Across locations and household types, respondents describe economic hardship shaped by a sequence of severe and overlapping macroeconomic shocks during 2021–2022, including currency depreciation, dollarization of essential goods, fuel shortages, and disruptions to labor markets. These shocks shape household narratives, manifesting primarily through elevated housing costs, medical expenses, and food insecurity. Beneficiaries and non-beneficiaries report largely similar sets of constraints at comparable rates, suggesting that the economic crisis affects both groups similarly. Cash transfers are frequently described as providing partial buffering against extreme outcomes, such as delaying eviction, stabilizing food consumption, or reducing reliance on high-cost debt, rather than generating sustained improvements in overall living standards.

Importantly, the qualitative data reveal that the extent to which cash assistance can translate into observable welfare gains depends critically on local conditions. In settings characterized by informal housing arrangements and lower fixed costs, such as rural and informal settlements, respondents describe more direct pathways from transfers to food consumption and reductions in food-related coping behaviors. In contrast, in urban and peri-urban areas with dollarized rental markets, higher housing costs, and greater institutional frictions, transfers are more likely to be absorbed by rent obligations, medical expenses, or debt repayment, limiting their impact on measured consumption.

Taken together, the qualitative findings indicate that treatment effect heterogeneity reflects the interaction between a national social protection program and locally binding constraints in an environment of exceptional macroeconomic volatility. While the program delivers substantial resources relative to pre-transfer income, the scale and nature of contemporaneous shocks often limit households' ability to translate transfers into sustained welfare improvements, with outcomes depending critically on local market conditions and place of residence. These patterns underscore the central role of local context in shaping program outcomes when social protection programs are implemented at scale.

2 Literature and Contribution

A large literature studies the targeting of social assistance, examining both the determinants of targeting efficiency and the relationship between targeting choices and program effectiveness (Ravallion, 2009; Alatas et al., 2012, 2016; Stoeffler et al., 2016; Brown et al., 2018; Hanna and Olken, 2018; Karlan and Thuysbaert, 2019; Basurto et al., 2020; Premand and Schnitzer, 2020; Haushofer et al., 2022). This work typically compares alternative targeting approaches or evaluates the aggregate impacts of a given program design. Our study contributes to this literature by documenting and explaining treatment effect heterogeneity within a nationwide cash transfer program operating at scale in a real-world policy environment, and by highlighting the role of local context in mediating the program effectiveness.

In a setting of forcibly displaced households with limited access to formal insurance and labor markets, we document substantial variation in program benefits across localities. Drawing on extensive qualitative evidence, we show that this heterogeneity is closely linked to aggregate economic shocks and how these shocks interact with locally binding market and institutional constraints facing displaced households. These findings underscore the importance of local context for understanding the effectiveness of scalable social protection programs. More broadly, our study speaks to the program evaluation literature on site-specific heterogeneity (Allcott, 2015), context dependence (Pritchett and Sandefur, 2015), and the external validity of experimental estimates across settings and scales (Banerjee et al., 2017; List, 2022). We show that unobserved and location-specific factors generate pronounced and difficult-to-predict treatment effect heterogeneity, limiting the generalizability of estimates derived from a narrow set of program sites, even when implementation is uniform and contemporaneous at the national level.

We also contribute to the literature on geographic targeting, defined as the prioritization of program resources across localities.[1] In large-scale national transfer programs, initial budget allocations across localities are routinely determined using area-level poverty maps.[2] These geographic allocation rules shape both the socio-demographic composition of beneficiaries and the market environments in which transfers are ultimately used.

Despite extensive evidence on the average impacts of cash transfer programs, relatively little is known about how alternative geographic allocation rules affect treatment effect heterogeneity when programs are implemented at scale. Because localities differ in market structure, exposure to shocks, and capacity for risk sharing (Kinnan et al., 2020), the extent to which households can smooth consumption depends critically on local economic conditions (Hanna and Karlan, 2017). Our empirical results show that local market context plays a central role in mediating the effectiveness of cash transfers.

Our analysis is closely related to recent work by Garlick et al. (2025), who study treatment effect heterogeneity in large-scale poverty-alleviation interventions and find limited evidence of systematic heterogeneity with respect to observed household characteristics using flexible machine-learning methods. They conclude that cash transfers operate largely as a general tool, with relatively uniform effects across recipients along standard demographic and economic dimensions. In contrast to their work, we document substantial heterogeneity in program impacts across localities, even under uniform implementation and contemporaneous rollout at national scale. This divergence highlights the role of local market conditions, institutional frictions, and exposure to aggregate shocks as central drivers of treatment effect heterogeneity—dimensions that are largely orthogonal to the household-level covariates typically used for targeting.

Finally, our study relates to a growing program evaluation literature on the effectiveness of

[1] With the growing availability of spatial data and advances in predictive methods, geographic targeting has increasingly been used as a standalone allocation strategy in data-constrained settings (Elbers et al., 2007; Abelson et al., 2014; Aiken et al., 2022; Asher et al., 2021; Blumenstock et al., 2015; Smythe and Blumenstock, 2022).

[2] Mexico's PROGRESA is a leading example. The program first identified disadvantaged regions using aggregate poverty indicators, then selected communities with limited access to infrastructure and basic services, and finally applied a proxy means test to determine household-level eligibility.

social protection in humanitarian and displacement settings. Experimental evidence in these contexts has remained relatively limited until recently (Quattrochi et al., 2020), in part because the urgency and operational constraints of emergency response often preclude the design and implementation of randomized evaluations (Hanna and Karlan, 2017). Early studies examined the relative effectiveness of cash, in-kind, and voucher-based assistance among displaced populations (Hidrobo et al., 2014; Aker, 2017). More recent work has used experimental and quasi-experimental methods to study the economic and social impacts of humanitarian assistance programs for refugees (Schwab, 2019; Sterck and Delius, 2020; Sterck et al., 2020; Lehmann and Masterson, 2020; Masterson and Lehmann, 2020; MacPherson and Sterck, 2021; Aygün et al., 2021; Kurdi, 2021; Altındağ and O'Connell, 2022). We contribute to this literature by experimentally studying how household targeting and geographic allocation rules shape program effectiveness within an at-scale humanitarian cash transfer program. Our study provides evidence on treatment effect heterogeneity that is relevant for the design and evaluation of scalable social protection interventions in displacement settings.

3 Institutional Setting

As of 2022, more than 1.5 million forcibly displaced Syrians reside in Lebanon (Govt. of Lebanon & United Nations, 2023). Refugees live in non-camp settings and are spread throughout the country, with no statutory restrictions on mobility. The United Nations World Food Programme (WFP) and the United Nations High Commissioner for Refugees (UNHCR) support the refugee population in Lebanon through education, protection, shelter, and health care, among others. In collaboration with international and local NGOs, the UN agencies' primary form of assistance is through targeted monthly unconditional cash-based transfers (UCTs). These programs annually disburse over $250 million USD, reaching between 40% and 90% of the refugee population in recent years.

The assistance cycle operates on an annual basis, and beneficiary assignment uses a proxy-means test (PMT) targeting household expenditure per capita. Since 2016, the PMT has been based on an econometric model that uses survey and administrative data held by UNHCR. In 2021-22, the program benefit structure had three tiers. The poorest eligible households (roughly 40% of the population) received 800,000 Lebanese Pounds (LBP) per month (53 USD), plus 300,000 LBP (20 USD) for each of up to six family members. Depending on a set of programmatic background factors, the middle tier reaches approximately 45% of households and provides either 800,000 LBP in cash or 300,000 LBP for each of up to six family members in food voucher credit per month.[3] Those in the least-poor quintile receive no assistance. These transfer values are substantial:

[3]Neither potential beneficiaries nor implementing field staff were able to manipulate eligibility scores or randomization outcomes. Access to the data on scores and treatment arm assignment was highly restricted and, beyond ourselves,

a household of five eligible for the highest transfer value would receive approximately USD 153 per month. According to our survey data, the median monthly expenditure for a refugee household of five in June 2021 was 90 USD. Each of the treatment arms provides more than USD 65 million over the course of the study period, reaching more than 250,000 refugee families.[4]

A central issue in implementing a program of this size is the degree of heterogeneity in living conditions and economic constraints throughout the country. For example, households face different price levels and varied opportunities to access food, housing, and services depending on where they live. Our qualitative fieldwork and survey data both indicated that those living in informal settlements, which exist throughout the country, have stronger social support networks than those living in separate dwellings more prevalent in urban areas. Households near the Syrian border typically have access to markets and community networks in their native country, which reduce food and income insecurity. In remote areas, limited access to school and medical services primarily stems from high transportation costs. Conversely, in urban areas, this limited access is often a result of congestion and challenges resulting from higher capacity state institutions that relate to legal documentation requirements for enrollment. These varying factors highlight the complexities involved in allocating the national program budget across districts, as policymakers must account for the diverse economic conditions in the country.

During the 2021–2022 assistance cycle, humanitarian agencies considered four aggregate level poverty metrics to determine district-level budget distribution in their nationwide UCT program. The first one served as a benchmark and used traditional monetary poverty, as measured by expenditure per capita, with a poverty threshold set to an expenditure-based poverty line determined by a group comprised of experts from humanitarian agencies in Lebanon.[5] The second arm used food insecurity via the reduced Coping Strategies Index (rCSI), which measures the degree of food insecurity of a household via eight food coping strategies that the household engaged in during the week before the interview. The poverty threshold is a score of 18 or greater (out of 56), indicating high food insecurity. The third arm was based on the food consumption score (FCS), which is a proxy measure of a household's caloric intake based on the frequency of consumption across eight differentially weighted food categories over the previous week. A score of 42 or lower (out of 112) indicates

was available only to a small number of UN personnel tasked with program implementation. Altındağ and O'Connell (2022) confirm there is no evidence of manipulation in eligibility around score thresholds in the same setting in multiple prior annual cycles.

[4]All conversions in this paper use an exchange rate of LBP 15,000 per USD from June 2021. From 2021 onward, the Lebanese pound depreciated substantially leading to reductions in the real value of transfers. Nominal transfer values were adjusted by implementing organizations throughout 2022 to offset reductions in real values. The nominal values cited in the text refer specifically to transfers being made from September 2021 to March 2022, after which they were increased to offset currency depreciation.

[5]This poverty line reflects the consumption level required for a family of two adults and three children, one aged over five and other two aged under five, to satisfy basic needs such as food, shelter, heating, water, and clothing; see UNHCR (2023).

inadequate food consumption. The last arm targeted a multidimensional deprivation index (MDI) that aims to reflect multiple deprivations across basic needs of food, health, education, shelter, water supply, sanitation, hygiene (WASH), and safety. Binary deprivation indicators are aggregated across subcategories, resulting in an index that ranges from zero (not deprived) to one (deprived in all dimensions); a household with a score of .33 or greater is considered multidimensionally deprived.[6] These measures are frequently used by international organizations, governments, and humanitarian agencies to assess vulnerability and structure social assistance programs. In our setting, these outcomes are actively monitored and hold particular relevance for policy.

4 Data and Empirical Framework

4.1 Data

UNHCR database The foundation of the administration of the assistance program, as well as our analysis, is a database of all refugee households that have made themselves known to UNHCR in Lebanon.[7] This process provides refugees with official proof of identity, safeguarding them from forced return or detention while facilitating their enrollment in protection and assistance programs. Consequently, refugees have a strong incentive to be included in administrative records (Altındağ et al., 2021). The UNHCR database is regularly updated through mobile and in-person interactions with refugee families. These data include demographic details, past and current assistance records, and function similarly to a basic social register.

Vulnerability Assessment of Syrian Refugees in Lebanon Every year, UNHCR, WFP, and their partners administer a nationally representative vulnerability survey that collects data from a sample of households on an array of living conditions, protection concerns, employment, income, and other measures of well-being and deprivation. This survey, the Vulnerability Assessment of Syrian Refugees in Lebanon (VASyR), serves as the primary data source for the empirical analysis and can be linked to the UNHCR database using unique household IDs. The VASyR survey has been collecting comprehensive data on the well-being and expenditures of refugee families since 2016, and typically surveys 4,000 to 5,000 households per year across Lebanon. VASyR 2021 was conducted in May and June of 2021 and provides pre-intervention outcomes. The same data is also used for training the PMT model of household expenditure per capita as well as district-level poverty indicators used for budget allocation. We use VASyR 2022, collected in June and July of 2022 to measure post-intervention outcomes. Subsequent to our power calculations for this

[6] See World Food Programme (2008a), World Food Programme (2008b), and World Food Programme (2023) for official definitions and guidance on the construction of the rCSI, FCS, and MDI, respectively.

[7] Individuals and families interested in being enrolled with UNHCR are provided with appointments to collect biographical data and vulnerability information. Records of those found to no longer be in the country are inactivated in the database, and verification of refugees' whereabouts, family composition, and vulnerabilities takes place on an ongoing basis.

study, the implementing agencies surveyed 2,091 additional households during the same period, randomly drawn from marginal beneficiaries. Unlike a typical local RCT with household follow-up, our study relies on two nationally representative repeated cross-sections: VASyR 2021, used solely to construct the pre-intervention PMT model and budget allocations, and VASyR 2022, a post-intervention survey in which marginal beneficiaries were oversampled. Aside from this oversampling, VASyR 2022 was implemented following UNHCR's standard survey protocols by its in-house data collection teams and was not customized for the study.

Focus Group Discussions We conducted 12 focus group discussions with 114 participants between July 21 and July 29, 2022. Participants were randomly sampled from the experimental population of marginal beneficiary households for whom we identify local average treatment effects of the program. Within each randomized budget allocation arm, the focus groups included marginal households assigned to both the higher-transfer treatment arm and the lower-transfer control arm. Given practical constraints on conducting discussions in all locations, we selected four geographically and economically distinct districts Bekaa, Tripoli, Saida, and Beirut. These districts are spatially separated and differ in local market conditions, institutional environments, and exposure to contemporaneous shocks. The qualitative evidence therefore reflects the lived constraints of marginal households in settings that span the range of contexts in which the program operated, rather than a convenience or illustrative sample.

4.2 Study Design and Randomization

During the intervention, the program allocated funds to beneficiaries through a two-step process. First, the national program budget was divided into four equal parts. Each part was then distributed at the district level using a different allocation rule. These rules generally followed a simple principle: funds were allocated in proportion to each district's estimated share of the national poor. For example, if a district accounted for 20% of the national poor, it received 20% of the program's overall budget. The distinction between the four allocation rules were the poverty indicators used for distribution — monetary poverty, food insecurity, low food consumption, or multidimensional deprivation. These four budgets operated simultaneously within every district, effectively creating four parallel allocation regimes. Each district therefore received four different budgets, and *households* within a district were randomly assigned to one of them.

In the second step, each district applied a standard PMT separately for each budget, ranking households from the bottom up based on their predicted per capita expenditure.[8] Funds were then distributed progressively, with the poorest households receiving the highest value assistance package and the less poor receiving a lower tier of assistance, until the allocated district budget was

[8]See Altındağ et al. (2021) for further details for the PMT.

exhausted leaving the least poor out of the program.

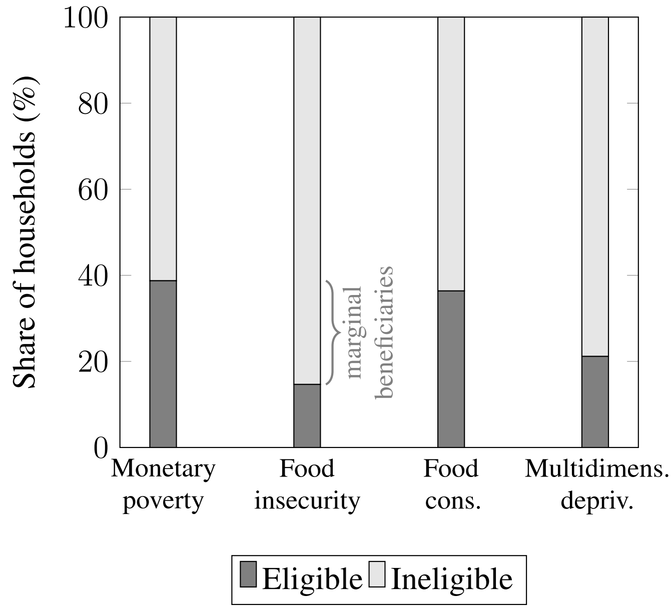
Each household had four potential assistance statuses that were observable to us *ex ante*, and their allocation to one arm therefore generated exogenous variation in assistance levels among otherwise identical households located in the same district. By examining the counterfactual assistance amounts for each household, we assess whether the assigned allocation rule changes the transfer value for each household across allocation rules. Households whose assistance amounts would have differed under a different allocation rule are the focus of the subsequent analysis, as they reveal the program trade-offs faced by policymakers. We refer to these households as *marginal* beneficiaries. In contrast, households for whom the district-level budget allocation rule had no impact on assistance levels are referred to as *inframarginal* beneficiaries. While eligibility simulations could be performed using administrative data alone, randomization was essential to causally identify differences in program outcomes across four budget arms.

Figure 1 provides a visual example of the randomized variation embedded in the design. Noting that a household's percentile ranking in the PMT remains stable across arms, the figure shows how different district allocation rules include more versus fewer households into the eligible set. Each bar corresponds to the program budget in the district in terms of the share of households that can be supported for a given aggregate allocation rule. Each bar is then divided into the share of households classified as eligible for assistance and those classified as ineligible. For the example used, the district of Zahle had high levels of monetary poverty and inadequate food consumption leading to those arms being more generous there compared to the food insecurity and multidimensional deprivation arms. Under an allocation rule based on monetary poverty, approximately 38 percent of households would qualify for full support. Under a rule prioritizing food insecurity, only 14 percent of households in Zahle would be supported. Households between the 14th and 38th percentiles are therefore marginal for this pair of allocation rules: their eligibility switches depending on whether the program uses monetary poverty or food insecurity as the primary criterion. The bracket in Figure 1 highlights this percentile range. Inframarginal households, by contrast, are those whose eligibility status does not change across the two rules.

4.3 Counterfactual Framework and Estimation Strategy

Consider a household i with potential outcomes $Y_i(Z)$ and a transfer amount $T_i(Z)$, representing the amount household i would receive if assigned to budget allocation rule Z. Here, Z takes values from the set $Z \in \{1, 2, 3, 4\}$, corresponding to the four budget allocation rules. Household index i ranges over the set $\{1, 2, 3, ..., N\}$, and transfer amount T_i can vary between zero and a positive amount. Since household i is randomly assigned to Z, the exogeneity assumption holds, meaning that potential outcomes and transfer amounts are independent of the assigned district budget:

Figure 1: Marginal and inframarginal beneficiaries, Zahle District



$$\forall j \in \{1,2,3,4\}, \quad Y_i(j), T_i(j) \perp\!\!\!\perp Z$$

For household i assigned to allocation rule $Z \in \{j,k\}$ where $j \neq k$, we can observe the transfer amounts the family would receive under each assignment. Based on this, we can categorize the household into one of two distinct scenarios:

$$T_i = \begin{cases} \text{Inframarginal beneficiary} & \text{when } T_i(j) = T_i(k) \text{ (equivalently, } \Delta T_{j,k} = 0). \\ \text{Marginal beneficiary} & \text{when } T_i(j) \neq T_i(k) \text{ (equivalently, } \Delta T_{j,k} \neq 0), \end{cases} \tag{1}$$

The difference between the means of the two outcome distributions provides the difference in average economic well-being between households due to the assigned budget allocation rule:

$$\tau_{jk} = E[Y|Z=j] - E[Y|Z=k] \tag{2}$$

which can be further decomposed by the beneficiary types described in equation 1:

$$\tau_{jk} = \begin{cases} \tau_{jk|\Delta T_{j,k}=0} = E[Y|Z=j, \Delta T_{j,k}=0] - E[Y|Z=k, \Delta T_{j,k}=0] \\ \tau_{jk|\Delta T_{j,k}\neq 0} = E[Y|Z=j, \Delta T_{j,k}\neq 0] - E[Y|Z=k, \Delta T_{j,k}\neq 0] \end{cases} \tag{3}$$

where the aggregate difference is a weighted average of the two:

$$\tau_{jk} = \tau_{jk|\Delta T_{j,k}=0} \times Pr(\Delta T_{j,k} = 0) + \tau_{jk|\Delta T_{j,k}\neq 0} \times Pr(\Delta T_{j,k} \neq 0) \tag{4}$$

If the assignment is exogenous and the exclusion restriction holds, $\tau_{jk|\Delta T_{j,k}=0}$ should be zero in expectation for all outcomes. For these inframarginal households, assignment to allocation rule j versus k does not change the amount of assistance received, and therefore cannot affect outcomes.

The term $\tau_{jk|\Delta T_{j,k}\neq 0}$ captures outcome differences among households whose transfer amount changes under the two allocation rules. This group can be further partitioned based on the direction of the transfer change:

$$\Delta T_{j,k} > 0 \quad \text{and} \quad \Delta T_{j,k} < 0.$$

Households with $\Delta T_{j,k} > 0$ receive a larger transfer under allocation rule j than under k, while households with $\Delta T_{j,k} < 0$ receive a smaller transfer. For households with $\Delta T_{j,k} > 0$, assignment to j induces an increase in treatment intensity, allowing us to interpret differences in outcomes as causal effects of additional cash assistance.

For the subset of households with $\Delta T_{j,k} > 0$, the research design identifies a local average treatment effect corresponding to the causal impact of an increase in cash assistance. This estimand is given by

$$\tau_{jk,LATE} = \frac{E[Y|Z=j, \Delta T_{j,k} > 0] - E[Y|Z=k, \Delta T_{j,k} > 0]}{E[T|Z=j, \Delta T_{j,k} > 0] - E[T|Z=k, \Delta T_{j,k} > 0]} \tag{5}$$

To estimate $\tau_{jk,LATE}$ using Two-Stage Least Squares (2SLS), we proceed with the following two-stage estimation among the sample of households for whom $\Delta T_{j,k} > 0$:

$$T_i = \pi_0 + \pi_1 \mathbf{1}(Z_i = j) + \eta_i, \tag{6}$$

where T_i is the transfer amount, $\mathbf{1}(Z_i = j)$ is an indicator for being assigned to allocation rule j, and η_i is an error term. The coefficient π_1 captures the first-stage relationship between allocation rule and the amount of transfer that a household receives.

The second-stage specification follows as:

$$Y_i = \alpha + \tau_{jk,LATE}\widehat{T_i} + \varepsilon_i, \tag{7}$$

where Y_i is the outcome of interest, and $\widehat{T_i}$ is the predicted transfer amount from the first stage. The coefficient $\tau_{jk,LATE}$ captures the causal effect of cash transfers on the outcome among marginal beneficiaries (households for which $\Delta T_{j,k} > 0$).

In this setting, $\tau_{jk,\text{LATE}}$ provides an unbiased estimate of the local average treatment effect of an additional transfer amount under three key assumptions: (i) Z is randomly assigned; (ii) PMT rankings and transfer amounts cannot be manipulated by potential beneficiaries; and (iii) Z affects outcomes only through its impact on cash assistance. We empirically test the validity of the random assignment in the next section. Because PMT scores and the eligibility assessment methodology are known only to a small number of central staff and researchers, manipulation of scores is highly unlikely and formal tests of this provide supporting empirical evidence.[9] The overall size of the cash assistance program is less than one percent of Lebanon's GDP, supporting the claim that our estimates of direct effects are not biased by changes in price levels or other aggregate market-level features.[10] Our research design, however, does not separately identify indirect effects of the program, which should be used when characterizing total program effects.

The 2SLS estimates in equation 5 represent the average program effect for marginal beneficiaries prioritized for a larger assistance amount under budget arm j versus k. We pool the data across all marginal beneficiaries from each pairwise counterfactual with budget arm j, which recovers the weighted average across the pairwise samples with the weights corresponding to the relative sizes of the marginal beneficiary groups for each comparison.[11] This allows us to estimate Equation 5 using (i) the pooled sample of marginal beneficiaries across all randomized budget allocation arms, (ii) separately within each randomized budget allocation arm, and (iii) separately by district.

4.4 Pre-registration and Multiple Hypothesis Testing

We pre-registered an analysis plan specifying hypotheses, validity checks, outcome definitions, and regression specifications. The outcomes and specifications analyzed in this paper correspond exactly to the primary outcomes specified in the preregistration and are implemented as planned; for the purposes of this paper, we present a pooled-sample version of our main effects in Table 3 and apply multiple hypothesis test corrections by family as specified in the preregistration using Anderson (2008). The treatment effect heterogeneity presented in Sections 5.2.2 and 5.2.3 were not pre-registered, are explicitly exploratory, and are included to focus on understanding heterogeneity and policy design at scale. The preregistration and the results of all prespecified analyses are publicly available on the Social Science Registry at https://www.socialscienceregistry.org/trials/9725.

[9] See Altındağ and O'Connell (2022) for details.

[10] For context, the cash transfer program studied in Egger et al. (2022) amounted to over 15 percent of local GDP. The study found large impacts on consumption and assets for recipients, significant positive spillovers to non-recipient households and firms, minimal price inflation, and estimated a local transfer multiplier of 2.4.

[11] In this specification we cluster standard errors by household to account for households that are marginal beneficiaries across more than one counterfactual comparison.

4.5 Balance tests and validation of intervention and empirical design

Appendix Table 1 shows that households in the post-intervention survey were balanced across treatment arms in pre-intervention characteristics. $F-$statistics and corresponding p-values from a joint hypothesis test based on a specification that regresses each demographic variable of interest on indicators for three of four treatment arms fail to reject equality across means in all cases.

Next we test whether the allocation rules include in the program populations that are more likely to experience the specific type of deprivation that it intends to target. Empirical evidence supporting this in our setting is provided in Table 1. The first column presents four baseline poverty rates for populations in each allocation rule, derived from a simple Ordinary Least Squares (OLS) model. Using the baseline survey data, we regress pre-intervention poverty indicators on the vector of indicators for assigned treatment arm with the monetary poverty arm as the reference group. The specification is the sample analog to Equation 2 allowing for multiple treatment groups:

$$\begin{aligned} Y_i = \beta_0 + \beta_1 * 1[Z_i = \{Food\ insecurity\ targeting\}] \\ + \beta_2 * 1[Z_i = \{Food\ consumption\ targeting\}] \\ + \beta_3 * 1[Z_i = \{Multidimensional\ deprivation\ targeting\}] + u_i \end{aligned} \tag{8}$$

The table reports $\beta_0, \beta_1, \beta_2$ and β_3 across rows, with the indicator for each type of poverty (Y_i) specified in panels. The outcome and coefficients are scaled to be in percentage points out of 100. As expected, the first column shows the randomization balance in initial poverty rates across experimental budget allocation rules. Column 3 presents the same regression results limited to households eligible for the most generous assistance package. This confirms that each experimental arm shifted the beneficiary profile toward the type of poverty targeted by that arm, though also incidentally towards others. For instance, 46.3 percent of beneficiaries receiving the highest value assistance package under the monetary poverty allocation rule are food insecure–a share that would be 13.2 percentage points higher if district-level budgets had been allocated based on food insecurity rates instead. Taken together, these results show that the intervention operated as intended. First, the randomization produced comparable groups, and this remains true in the endline data using pre-intervention characteristics. Second, the beneficiary profiles in each arm indicate that targeting based on area-level poverty indicators successfully directs assistance toward the intended poor population. Finally, because the intervention was randomized at the population level and the endline survey was independently re-sampled from that same population, rather than following up the original baseline households, the analysis is not subject to differential attrition or concerns about follow-up rates across treatment arms.

Table 1: Tests of balance and targeting effect on beneficiary profiles

	Full sample	std. err.	Beneficiaries	std. err.
Outcome: % expenditure poor				
Monetary poverty arm mean	85.28		95.1	
I[Z=Food insecurity arm]	-0.76	(1.44)	-6.16***	(1.52)
I[Z=Food consumption arm]	-1.55	(1.46)	-0.88	(1.34)
I[Z=Multidimensional deprivation arm]	0.1	(1.41)	-5.72***	(1.47)
N	4953		2477	
Outcome: % food insecure				
Monetary poverty arm mean	46.42		46.27	
I[Z=Food insecurity arm]	0.76	(2.00)	13.18***	(2.79)
I[Z=Food consumption arm]	0.58	(2.00)	4.29	(2.95)
I[Z=Multidimensional deprivation arm]	2.07	(1.98)	8.49***	(2.76)
N	5017		2494	
Outcome: % with inadequate food consumption				
Monetary poverty arm mean	41.88		39.45	
I[Z=Food insecurity arm]	3.13	(1.98)	5.6**	(2.78)
I[Z=Food consumption arm]	-0.56	(1.98)	6.28**	(2.92)
I[Z=Multidimensional deprivation arm]	-1.01	(1.95)	0.61	(2.71)
N	5017		2494	
Outcome: % multidimensionally deprived				
Monetary poverty arm mean	11.23		11.36	
I[Z=Food insecurity arm]	-1.4	(1.23)	0.25	(1.8)
I[Z=Food consumption arm]	-0.92	(1.24)	-0.03	(1.88)
I[Z=Multidimensional deprivation arm]	-0.89	(1.23)	3.91**	(1.87)
N	5017		2494	

Note: Table contains means of pre-intervention poverty rates and differences relative to households in the monetary poverty targeting arm among all sampled households (Columns 1 and 2) and post-intervention beneficiaries (Columns 3 and 4). Data are from the the pre-intervention survey; sample sizes vary due to outcome missingness. Results of the t-test of mean differences between beneficiaries and non-beneficiaries are indicated by $^{*}p < .10$; $^{**}p < .05$; $^{***}p < .01$. **Reading:** The Full Sample column tests for balance in the indicated outcome across treatment assignment. 85.28% of households assigned to the monetary poverty targeting arm were expenditure poor, and this rate is statistically indistinguishable across all targeting arms. The Beneficiaries column tests for the effect on targeting: 95.1% of beneficiaries under the monetary poverty targeting arm were expenditure poor, which is statistically significantly higher than beneficiary households subject to food insecurity or multidimensional deprivation targeting by 6.16 and 5.72 percentage points, respectively. Targeting poor food consumption results in beneficiaries who are no less likely to be expenditure poor, however.

Characterizing marginal beneficiaries

Similar to the right-hand side of Table 1, we show the characteristics of marginal beneficiaries prioritized by each budget allocation arm. Table 2 presents descriptive statistics on demographics, poverty levels, and well-being measures for marginal beneficiaries in the "control" condition across budget allocation rules, with each column testing mean differences relative to those in the monetary poverty column.[12] Two insights emerge from this analysis. First, households targeted by consumption-based indicators (expenditure poverty and food consumption poverty) are relatively similar to each other, with few small differences from marginal beneficiaries in the monetary poverty allocation arm — and when differences are statistically significant, they are often of a small economic magnitude. Second, consumption-based geographic allocation rules prioritize starkly different demographic groups compared to the other vulnerability-based allocation strategies. Marginal beneficiaries in consumption-based allocation arms have higher ability to borrow and have higher baseline assets. Despite higher nominal consumption, beneficiaries in the vulnerability-based arms (food insecurity and multidimensional deprivation) have lower baseline ability to smooth consumption due to smaller household sizes, lower assets, greater financial market exclusion, and limited social support. They are less likely to have close friends, feel reliant on social connections for credit, or perceive their community as supportive and cohesive compared to the beneficiaries in consumption-based targeting groups.

5 Results

5.1 Program effects

First, we estimate the reduced form and local average treatment effects of the program among the pooled marginal beneficiary groups. The estimation proceeds in three steps. First, we identify all beneficiaries assigned to treatment arms j or k, where arm j offers a larger transfer amount than arm k (that is, $T(j) > T(k)$), representing the population prioritized under allocation rule j. Second, because a household may face up to three counterfactual transfer amounts, we stack the data for each (j, k) pair with $T(j) > T(k)$. Third, we pool all $j = 1, 2, 3, 4$ to estimate the overall program effect across marginal beneficiary types. We estimate the first-stage and local average treatment effects using Equations 6 and 7. Standard errors are clustered at the household level to account for within-household correlation.

Table 3 reports the first-stage, ITT, and 2SLS estimates, where the ITT coefficient corresponds to the numerator of Equation 5. Outcomes in the ITT and 2SLS specifications are z-standardized to facilitate comparisons across indicators, while the first-stage outcome is measured as transfer amounts in millions of Lebanese pounds. We report household-level poverty outcomes in the first

[12] A marginal household in the control arm is defined as a household that would be eligible to receive a larger transfer amount in / allocation rule j than in at least one other allocation rule k, but is assigned to k.

Table 2: Control means for marginal beneficiaries by geographic allocation arm

Measure	Monetary poverty	Food insecurity	Food consumption	Multidimensional deprivation
Demographics				
Household size	4.897	3.707***	4.797	3.645***
Share HH age 0-5	0.212	0.155***	0.192*	0.160***
Share HH age 50+	0.084	0.091	0.085	0.095
Share of nondisabled working-age males	0.127	0.224***	0.141**	0.246***
Female-headed household	0.284	0.158***	0.266	0.150***
Disability in household	0.159	0.154	0.156	0.109***
Share with no education	0.145	0.074***	0.125	0.081***
Share with high school education or above	0.304	0.327	0.317	0.331
Targeting score (predicted exp. per cap.)	2.500	3.128***	2.582***	3.233***
Well-being measures				
Livelihood coping strategies index (z-score)	5.338	5.216	5.353	5.369
WASH index (z-score)	-0.083	0.025***	-0.017***	0.006***
Shelter condition index (z-score)	0.010	-0.042*	0.004	0.038
Rental debt (MM LBP)	1.081	1.238*	1.118	1.186
Durable goods index	0.014	-0.027***	0.016	-0.008
Productive assets index	0.075	-0.016***	0.045	-0.016***
Social cohesion				
Has close friends	0.857	0.782***	0.843	0.828
Neighbors could care for children	0.628	0.645	0.601	0.604
Could borrow from social circle	0.804	0.751***	0.800	0.795
Willing to assist others	0.102	0.093	0.086	0.103
Community is supportive	0.602	0.533***	0.552**	0.486***
Community helps in emergency	0.685	0.597***	0.616***	0.540***
N	3338			

Note: Table contains means of control group marginal beneficiaries under each targeting strategy. Data are from the post-intervention survey. Sample size is N = 3,338 across all tests, made consistent across tests to account for missing outcome records. Each household appears in the sample only once. The results of t-tests of mean differences between the sample indicated in the column relative to the beneficiaries marginal to the Monetary Poverty arm are indicated by *p $<$ .01; **p $<$ .05; ***p $<$ .01 **Reading:** Households marginally prioritized by targeting monetary poverty have 4.89 people, on average. Households marginally prioritized by food insecurity targeting have 3.7 members on average, and the difference between this mean and that in the monetary poverty column is statistically significant at the 1% level.

four columns and pre-specified outcomes specific to children in the latter four columns. The first stage shows that assignment to a more generous district budget increases transfers by about LBP 0.92 million on average. The 2SLS estimates therefore capture the effect of an additional LBP 1 million (approximately USD 66 monthly during the study period) for marginal beneficiaries whose transfers increased due to random assignment. As expected, the reduced-form and 2SLS estimates are very similar, reflecting the fact that the first-stage increase is close to one million LBP, the denominator of the Wald estimator in Equation 5.

Table 3: Program effects across marginal beneficiaries, primary poverty outcomes, pooled sample

Outcome:	Expenditure per capita	Reduced Coping Strategies	Food Consumption	Multidimensional Deprivation	Child labor	Child out of school	Child sick	Female child married
Specification: ITT								
coef.	0.15***	0.08***	0.11***	0.06***	0.02	-0.11***	0.03	0.05
(se)	(0.02)	(0.02)	(0.03)	(0.02)	(0.04)	(0.03)	(0.03)	(0.06)
N	6444	6464	6464	6246	3398	3398	3746	1227
Specification: 2SLS (LATE)								
coef.	0.17***	0.09***	0.12***	0.06***	0.02	-0.11***	0.03	0.05
(se)	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.06)
N	6444	6464	6464	6246	3398	3398	3746	1227
Specification: First stage (outcome: transfer value in MM LBP)								
coef.	0.92***	0.92***	0.92***	0.92***	0.94***	0.94***	0.9***	0.92***
(se)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
N	6444	6464	6464	6246	3398	3398	3746	1227

Note: Table contains estimates of estimands described in Section 4.3 of the text. p-values are adjusted for multiple hypothesis testing within primary and secondary outcome families based on Anderson (2008). Statistical significance is indicated by *q < .10; **q < .05; ***q < .01. **Reading:** Top panel: marginal households have .15 standard deviation higher ln(expenditure per capita) when they receive a higher transfer due to being assigned to the arm in which they would receive a higher benefit. Bottom panel: marginal households have a .17 standard deviation higher ln(expenditure per capita) when they receive 1MM LBP more in assistance due to being assigned to the arm in which they would receive a higher benefit.

The program improves all household poverty indicators. Estimated effects range from 0.06 standard deviations for multidimensional poverty to 0.17 standard deviations for log expenditure per capita, with intermediate effects of 0.09 standard deviations for reduced food coping strategies and 0.12 standard deviations for food consumption. All estimates are statistically significant and economically meaningful, though they do not imply large reductions in poverty. Overall, the program's impact on economic well-being is positive but modest. In contrast, we find little evidence of improvements in child well-being outcomes. Effects on child labor, child sickness, and early marriage are small and statistically indistinguishable from zero. The only exception is school attendance: the share of children who are out of school declines by 0.11 standard deviations. Because all child outcomes are coded such that higher values indicate worse outcomes, this negative estimate reflects an improvement and is quantitatively comparable in magnitude to the effects observed on improvements in food consumption and coping strategy outcomes.

5.2 Program Effect Heterogeneity

We next examine how treatment effects on household-level outcomes vary across alternative allocation rules, locations, and household characteristics. We begin by comparing impacts across the four budget allocation rules, asking whether different geographic allocation strategies generate meaningful differences in program outcomes. Finding little evidence of economically or statistically meaningful differences in treatment effects across the distinct groups of marginal beneficiaries induced by these rules, we then turn to geographic heterogeneity. We show that program impacts vary substantially across locations, indicating that local market conditions and institutional environments play a dominant role in shaping effectiveness. Finally, we estimate household-level heterogeneity using flexible machine learning methods to characterize how treatment effects vary with observable baseline characteristics and to identify the correlates of high- and low-effect households. The last section integrates qualitative evidence collected after the intervention to interpret the mechanisms underlying the observed heterogeneity, providing a contextual explanation for why program impacts differ so substantially across places.

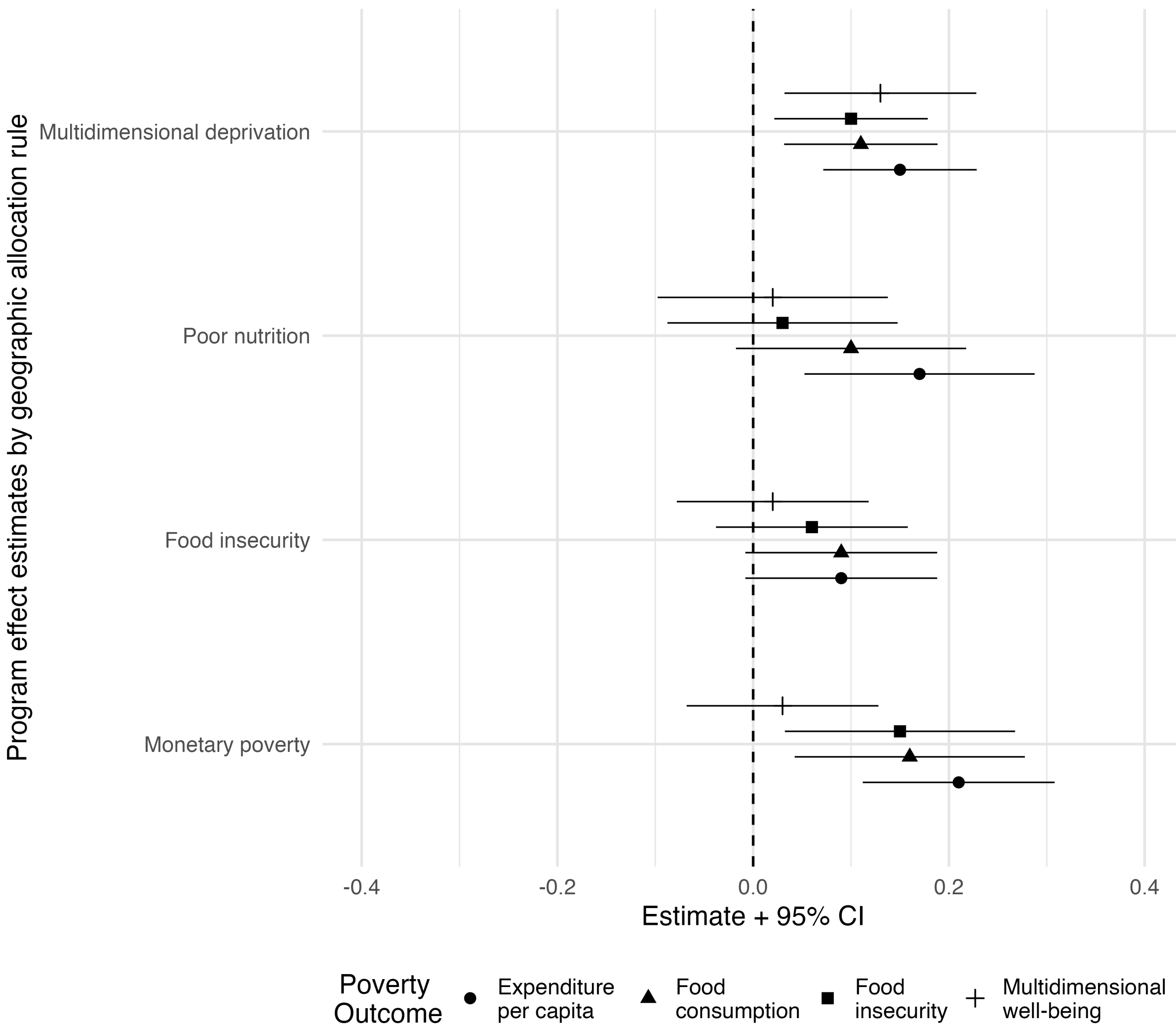
5.2.1 Heterogeneity by randomized allocation rules

This subsection examines whether treatment effect heterogeneity is associated with the poverty metric used to allocate resources across districts. We compare program impacts across groups of marginal beneficiaries induced by alternative budget allocation rules to assess whether these rules are linked to systematic differences in outcomes. Specifically, we test whether households whose transfer amounts change under different allocation rules experience differential treatment effects, holding program design, location, and implementation constant.

Figure 2 presents 2SLS estimates for marginal beneficiaries separately by budget allocation rule. The program effects are positive across all primary poverty indicators and all randomized budget allocation arms, although the point estimates vary. The geographic budget allocation rule based on monetary poverty yields sizable improvements in per capita expenditure (0.21 SD), coping strategies (0.15 SD), and food consumption scores (0.16 SD). The budget allocation based on multidimensional deprivation mitigates all deprivation measures consistently, with effects between 0.11 SD and 0.15 SD. In contrast, allocation rules based on food insecurity or inadequate nutrition produce limited and often statistically insignificant gains, with effects between 0.02 SD and 0.10 SD except for per capita expenditure.

Overall, no single allocation rule consistently generates the largest improvements across outcomes. In only a subset of cases do the largest effects align with the poverty indicator used for allocation, most notably for monetary poverty and multidimensional deprivation. By contrast, allocation rules based on food insecurity or inadequate food consumption prioritize households with worse baseline outcomes along those dimensions but do not produce correspondingly larger treatment effects. Although estimated treatment effects among marginal beneficiaries vary across allocation methods, we cannot statistically reject equality of program effects across rules. This suggests that no single budget allocation rule can be expected to differentially improve aggregate outcomes along any specific dimension of well-being.

Figure 2: Local average treatment effect estimates by geographic allocation rule



Note: Figure depicts LATE effects for each sample and across four primary outcomes. The groups on the vertical axis indicate the allocation rule to which the sample households are marginal beneficiaries. Coefficients are in units of the outcome's standard deviation. **Reading:** Households marginal to the multidimensional deprivation arm are positively impacted in all four outcome measures when receiving a higher transfer as a result of being assigned to that arm. Households marginal to food consumption targeting increase expenditures, with effects on other outcomes remaining statistically indistinguishable from zero.

It is clear, however, that marginal beneficiaries in the budget allocation arm that uses multidimensional deprivation are the only group that exhibits treatment effects that are statistically different from zero in all four primary poverty outcomes. This ends up being the case because it is the only budget arm that yields treatment effects statistically distinguishable from zero on multidimensional deprivation outcomes.[13] Households prioritized under the monetary poverty based allocation experience statistically significant improvements in three outcomes, expenditure poverty, food security, and food consumption, but not in multidimensional deprivation. Allocation based on poor food consumption improves expenditure-based poverty and food consumption, although the latter is only marginally significant. Finally, allocation based on food insecurity prioritizes households for whom none of the estimated treatment effects are statistically distinguishable from zero.

Overall, the program improves the primary poverty outcomes it targets, but we find little evidence of meaningful treatment effect heterogeneity across households that are marginal to different budget allocation rules. This limited variation across alternative allocation rules contrasts sharply with the substantial heterogeneity in program impacts observed across geographic locations that we document below. Taken together, this pattern suggests that once a cash transfer program is implemented at scale, local market conditions shared by beneficiaries within a given area play a more important role in shaping program outcomes than differences in the beneficiary set induced by alternative budget allocation rules.

5.2.2 Heterogeneity by location

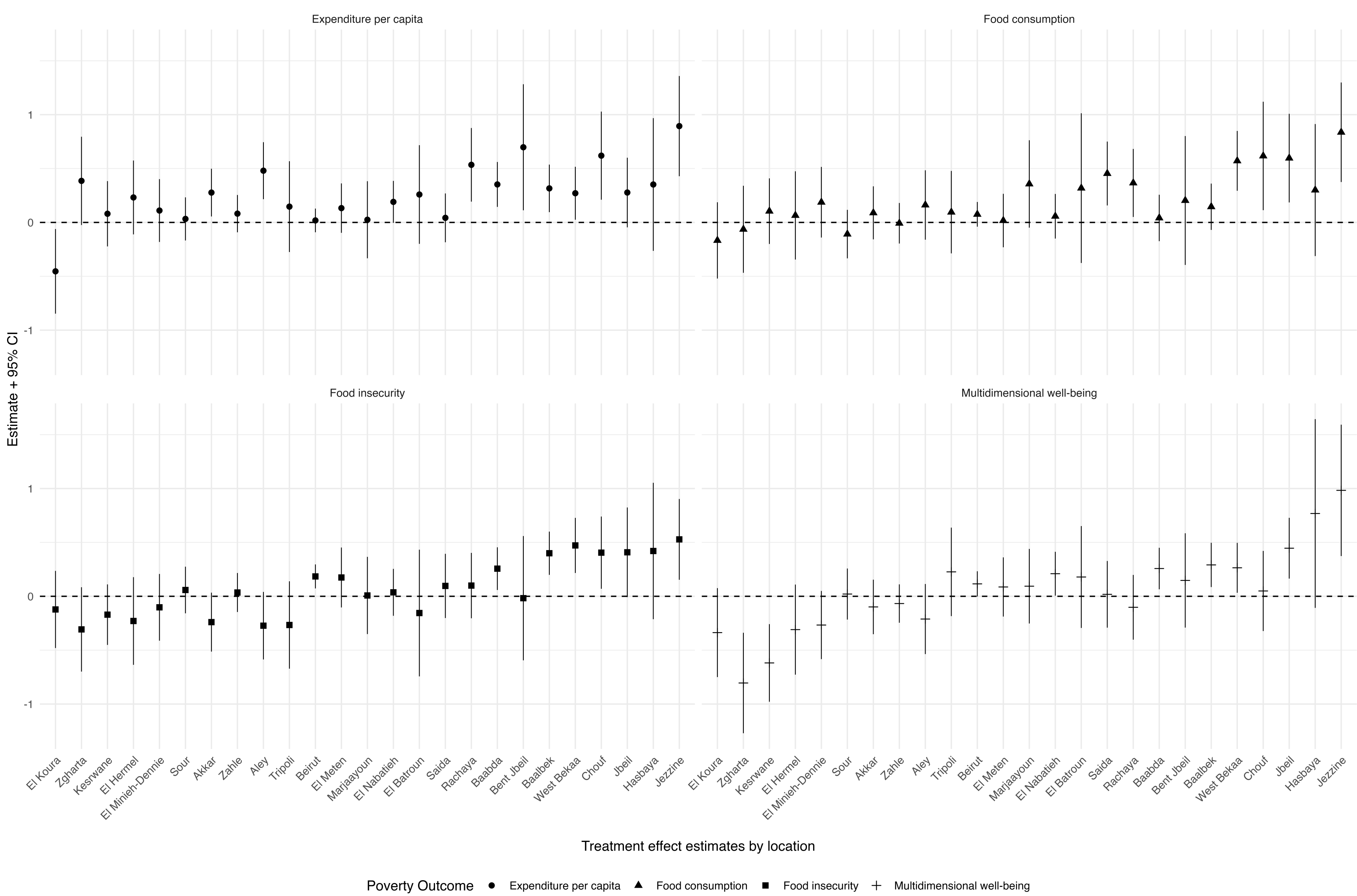
Having found little evidence that alternative budget allocation rules generate meaningful differences in program impacts, we next explore geographic heterogeneity in treatment effects. This analysis assesses whether program impacts vary systematically across locations that differ in local market conditions.

Figure 3 shows the estimated program effects across each district and by each of the main poverty outcomes. Effects are plotted based on the rank of the average effect magnitude in each outcome panel, and shapes indicate the outcome of interest. Variation in location-specific heterogeneity is substantial, ranging from a few negative point estimates to 0.5 SD treatment effects or larger; F-tests confirm that this heterogeneity is unlikely a result of sampling variation. Districts are ordered similarly across panels, where it becomes apparent that the same subset of districts tend to drive the overall improvements in well-being across all outcomes. Effects in the upper tail are large enough to statistically reject zero on their own, and there is a substantial positive correlation in effect sizes across outcomes. That is, the districts with the largest effect sizes tend to overlap

[13] This result is attributable to improvements to living quarters via lower physical crowding and reductions in physical insecurity.

across all outcomes that we investigate; the median Pearson correlation coefficient of effect sizes by district across outcomes is .58, with a range from .33 to .75. These results suggest location-specific factors are important determinants of program effects, and joint hypothesis tests reject a null of true differences across districts being zero with $p < 0.01$ for all four outcomes.

Figure 3: Local average treatment effect estimates by district



Note: Figure depicts local average treatment effects for each district sample and across four primary outcomes. Districts are ordered according to average coefficient size. Joint hypothesis tests reject a null of true differences across districts being zero with $p < 0.01$ for all four outcomes.

5.2.3 Heterogeneity by households

In exploring the dimensions of heterogeneity, we now shift our focus from geographic and allocation-based differences to variation at the household level. This allows us to understand how individual household characteristics interact with the program's design, offering a more granular view of which types of households benefit most and why. By doing so, we connect the broader patterns of heterogeneity back to the central theme of how context and beneficiary characteristics together shape the program's effectiveness at scale.

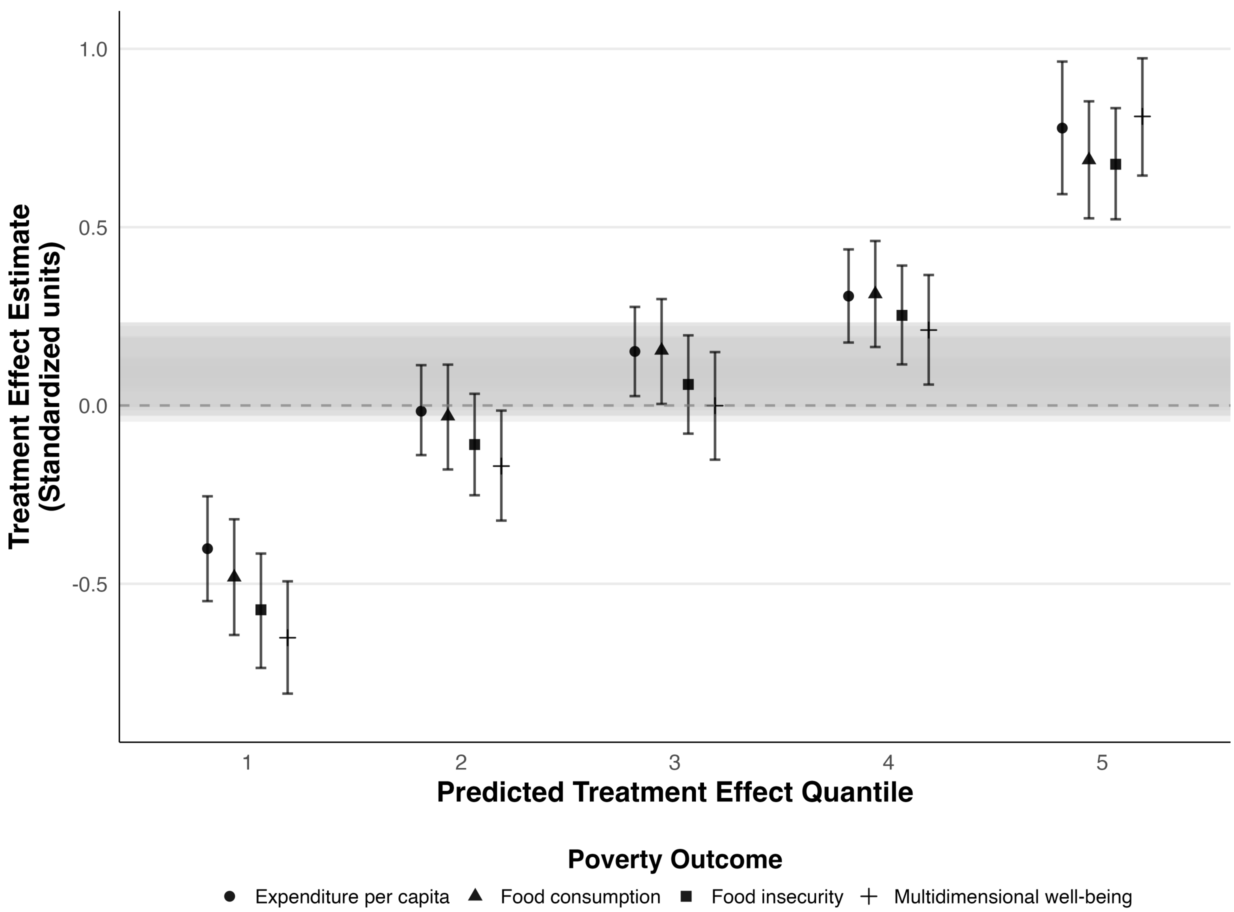
To do this, we apply the Generic Machine Learning (GenericML) framework of Chernozhukov et al. (2025), which allows for agnostic inference on treatment effect heterogeneity without imposing parametric assumptions.[14] The procedure follows in three steps. First, we use random forests to generate out-of-sample predictions of household-level reduced-form treatment effects based on the full set of baseline covariates, including household demographics, composition, district of residence, randomized allocation rule, and baseline poverty measures. Second, we sort households into five quintiles according to their predicted treatment effects and estimate Group Average Treatment Effects (GATES) for each quintile. If treatment effects are heterogeneous and the predictions capture this heterogeneity, GATES estimates should increase monotonically from the lowest to the highest quintile. Third, we conduct a Classification Analysis (CLAN) to characterize which baseline covariates systematically differ across quintiles, thereby identifying the observable factors most strongly associated with treatment effect heterogeneity.

Figure 4 and Table 4 jointly summarize treatment effect heterogeneity for the primary poverty outcomes using the GATES and CLAN frameworks. For each outcome, households are partitioned into five equally sized groups based on their predicted treatment effects, which we refer to as quintiles Δ_1 through Δ_5, with Δ_1 denoting the lowest predicted-effect group and Δ_5 the highest. Figure 4 reports the average treatment effect within each predicted-effect quintile, while Table 4 reports mean baseline covariate values by quintile for the log consumption per capita outcome. Corresponding CLAN results for baseline demographics and household characteristics for the remaining primary outcomes are qualitatively similar and reported in Appendix Tables 2–4.

The results reveal statistically and economically meaningful heterogeneity in program impacts driven by observable household characteristics. Households in the lowest predicted-effect groups (Δ_1 and Δ_2) experience zero or negative program effects, while households in Δ_3 through Δ_5 exhibit sizable improvements. For all primary outcomes, the Δ_5 effects exceed 0.5 standard deviations and

[14] We implement GenericML using the R package of Welz et al. (2022). Conditional average treatment effects are estimated using random forests with repeated sample splitting (100 splits), and standard errors are clustered at the household level. All outcomes are standardized to have mean zero and unit variance, and BLP and GATES specifications control for baseline predicted log per capita expenditure. Given the experimental design with known treatment probabilities, the propensity score is set to a constant.

Figure 4: Treatment Effects by Predicted Quantile



Note: Figure depicts heterogeneous treatment effects across predicted treatment effect quantiles. Grey bands show 95% confidence interval for the BLP β_1 coefficient. Error bars show 95% confidence intervals for quantile-specific effects. Points are spaced horizontally for clarity. Quantile 1 contains households with the lowest predicted treatment effects; Quantile 5 contains households with the highest predicted treatment effects.

lie well above the overall mean effect, whereas Δ_1 effects are negative and well below it.

Household demographics and baseline vulnerability exhibit systematic and largely monotonic relationships with treatment effect size. Female-headed and single-parent households are overrepresented in the highest predicted-effect quintile, while households with a greater share of non-disabled working-age men benefit less from the program. Lower baseline predicted per capita expenditure is also strongly associated with larger gains. Prior exposure to social protection dampens marginal effects, with previous cash beneficiaries disproportionately represented in the lowest predicted-effect quintile. These patterns indicate that observable vulnerability is an important correlate of program impact.

At the same time, location emerges as one of the strongest predictors of treatment effects across all outcomes. The geographic composition of predicted-effect quintiles broadly aligns with the district-level treatment effects documented in Figure 3. Households residing in poorer, more rural, and more informal districts such as Akkar, West Bekaa, and Baalbek are heavily overrepresented in the highest predicted-effect groups, while households in urban and service-oriented districts such as Beirut and Tripoli are disproportionately concentrated in the lowest predicted-effect groups. For example, 12.5 percent of households in the highest predicted-effect quintile reside in Akkar, compared to only 3.3 percent in the lowest quintile, whereas the opposite pattern holds for Beirut. These results indicate that household-level vulnerability is systematically related to treatment effects, while geographic location remains a strong and independent correlate of heterogeneity.

Table 4: CLAN estimates by covariate and GATES group: Log Cash Expenditure per Capita

Variable	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	$\Delta_5 - \Delta_1$
Household composition and age structure						
Household size	4.026	4.469	4.420	4.219	3.663	-0.385***
Female-headed household	0.152	0.147	0.181	0.219	0.299	0.15***
Share of household members aged 0–5	0.146	0.183	0.188	0.190	0.189	0.043***
Share of household members aged 6–10	0.101	0.119	0.129	0.131	0.106	0.005
Share of household members aged 11–17	0.115	0.132	0.131	0.124	0.105	-0.007
Share of male members aged 18–50	0.221	0.199	0.184	0.168	0.155	-0.065***
Share of household members aged 50+	0.110	0.071	0.073	0.079	0.101	-0.01
Labor and education						
Share of adults working as laborers	0.275	0.261	0.247	0.239	0.232	-0.047***
Share of adults working in services	0.087	0.087	0.096	0.107	0.112	0.022*
Share of adults with medium education (levels 3–4)	0.449	0.475	0.481	0.488	0.491	0.041*
Social protection						
Single parent	0.038	0.054	0.072	0.086	0.093	0.055***
Share of household members with medical condition	0.171	0.130	0.123	0.125	0.136	-0.034**
Share of household members with a disability	0.012	0.006	0.007	0.007	0.008	-0.003
Previous cycle cash transfer beneficiary	0.432	0.485	0.452	0.404	0.337	-0.089***
Previous cycle food assistance beneficiary	0.234	0.260	0.262	0.255	0.218	-0.017
Log predicted per capita expenditure (baseline)	12.611	12.543	12.532	12.538	12.564	-0.044***
District of residence						
Beirut	0.177	0.130	0.113	0.102	0.096	-0.082***
Zahle	0.106	0.140	0.132	0.101	0.047	-0.057***
Minieh-Dennieh	0.086	0.073	0.057	0.045	0.034	-0.054***
Koura	0.055	0.028	0.019	0.010	0.006	-0.047***
Saida	0.045	0.041	0.026	0.018	0.013	-0.032***
Marjaayoun	0.044	0.051	0.038	0.025	0.015	-0.028***
Sour (Tyre)	0.053	0.039	0.037	0.035	0.031	-0.019*
Tripoli	0.032	0.029	0.022	0.016	0.018	-0.016*
Hasbaya	0.012	0.012	0.010	0.006	0.003	-0.009*
Metn	0.038	0.039	0.044	0.041	0.031	-0.007
Batroun	0.023	0.012	0.013	0.013	0.018	-0.007
Keserwan	0.023	0.023	0.023	0.025	0.025	0.003
Zgharta	0.039	0.028	0.034	0.038	0.041	0.004
Jbeil	0.010	0.012	0.015	0.018	0.015	0.003
Bint Jbeil	0.001	0.006	0.009	0.012	0.012	0.009**
West Bekaa	0.047	0.064	0.074	0.070	0.058	0.01
Hermel	0.006	0.022	0.026	0.025	0.022	0.015***
Jezzine	0.004	0.006	0.009	0.013	0.020	0.018***
Chouf	0.001	0.004	0.006	0.013	0.025	0.022***
Rachaya	0.015	0.034	0.042	0.042	0.039	0.023***
Baabda	0.031	0.036	0.045	0.058	0.060	0.026**
Nabatieh	0.044	0.023	0.029	0.036	0.077	0.032***
Aley	0.010	0.015	0.026	0.045	0.050	0.039***
Baalbek	0.037	0.066	0.077	0.076	0.104	0.069***
Akkar	0.033	0.045	0.066	0.098	0.125	0.088***
N						

Note: Table reports CLAN estimates of covariate means for effects on cash expenditure per capita. Columns Δ_1 through Δ_5 report mean values of each covariate within quintiles of the estimated conditional average treatment effect, where Δ_1 corresponds to households with the lowest predicted gains from treatment and Δ_5 to those with the highest predicted gains. The final column reports the difference in means between the highest- and lowest-gain groups ($\Delta_5 - \Delta_1$). Sample based on marginal beneficiaries in post-intervention survey. Sample size is 6,847. Statistical significance of this difference is indicated by $^{*}p < .10$; $^{**}p < .05$; $^{***}p < .01$. **Reading:** Households predicted to benefit more from the program (higher GATES groups) are more likely to be female-headed or single-parent households, and have lower baseline predicted per capita expenditure.

6 Qualitative analysis: Explaining Treatment Effect Heterogeneity

Despite uniform program implementation over the same period and within the same country, results from multiple analytical approaches reveal meaningful variation in program effects across local, economic, and demographic dimensions. In this section, we use qualitative evidence to contextualize these patterns by examining the environments in which beneficiaries use cash transfers and to identify plausible sources of the observed treatment effect heterogeneity. As described in Section 4.1, we conducted 12 focus group discussions with 114 participants drawn from the experimental sample of marginal beneficiaries. Importantly, participants were sampled from both beneficiary and non-beneficiary households for whom local average treatment effects are identified. Discussions were held in four geographically and economically diverse districts: Beirut, Tripoli, West Bekaa, and Saida.

We use these discussions to interpret potential sources of treatment effect heterogeneity. Specifically, we adopt a structured constraint-mapping approach that leverages qualitative evidence to identify binding constraints, coping strategies, and local conditions that shape how households are able, or unable, to translate cash transfers into welfare gains. This analysis is not intended to estimate program impacts or establish causal mechanisms, but to document common constraints and narratives across settings rather than to measure their relative prevalence or causal importance.

6.1 Qualitative coding approach

Our qualitative analysis proceeds in three steps. First, we segment the focus group transcripts into discrete text units corresponding to individual participant contributions in response to moderator questions. Each segment typically consists of one or more sentences spoken by a single respondent and may express a constraint, coping behavior, or lived experience related to prevailing economic conditions. These segments constitute the unit of analysis and may reflect multiple constraints simultaneously. Second, we apply a deductive coding framework developed in consultation with the humanitarian organization's field offices, which interact with the refugee population on a daily basis. The framework comprises 30 prespecified constraint categories spanning monetary shocks (such as food prices, rental costs, medical expenditures, debt obligations, school fees, and transportation costs), non-monetary barriers (such as housing market constraints, documentation problems, institutional exclusion, and lack of community support), and macro-level shocks (such as currency devaluation, the bread and fuel crises, the COVID-19 pandemic, and the Beirut port explosion). We apply a conservative tagging protocol: a tag is assigned only when direct textual evidence clearly matches the code definition and ambiguous cases or those needing interpretation or inference are left untagged by default. Each assigned code is accompanied by a verbatim justification drawn directly from the participant's statement. Third, we implement a hybrid human–machine coding procedure. A large language model applies the thematic classifications to each transcript segment and generates initial code assignments with short textual justifications. Two research assistants, blinded to the composition of participants in each focus group discussion, then independently reviewed each coded transcript, validating assignments against the original text and the codebook definitions. Any discrepancies are resolved through discussion, and final codes are assigned by consensus. This approach combines the scalability of automated text classification with human judgment in ambiguous cases.

Finally, we synthesize the coded evidence to characterize the local environments in which the program operates. We aggregate constraint patterns at the district level and compare narratives across beneficiary and non-beneficiary respondents. This organization allows us to identify commonalities and differences in economic conditions, local market frictions, social relations between refugees and host communities, and access to services. The resulting qualitative portrait helps contextualize the quantitative heterogeneity results and assess which dimensions of treatment effect variation are consistent with the constraints and coping mechanisms described by participants themselves.

6.2 Overall narrative from focus group discussions

Overall, the focus group discussions reveal the unprecedented economic and institutional collapse that Lebanon experienced during 2021–2022. Roughly one-third of coded segments reference crisis-related constraints. These include housing and medical costs rendered unaffordable by dollarization and hyperinflation, acute shortages of cooking gas and bread, and the erosion of purchasing power as the Lebanese pound lost more than ninety percent of its value. Debt, medical costs, housing instability, and social tensions are mentioned at nearly identical rates across both beneficiaries and non-beneficiaries. When beneficiaries describe the effects of transfers, they emphasize partial buffering against extreme outcomes, such as delaying eviction, stabilizing food consumption, or reducing reliance on high-cost debt, rather than sustained improvements in living standards. These patterns suggest that cash assistance provides limited relief from a severe and ongoing macroeconomic shock rather than sustained improvements in poverty.

6.3 Geographic heterogeneity

Table 5 reports district-specific treatment effect estimates from the intervention for the primary program outcomes in districts where focus group discussion participant households lived. Treatment effects are largest and most consistently positive in West Bekaa. In this district, the program increases log cash expenditure per capita by 0.27 standard deviations and generates statistically significant improvements in reduced coping strategies (rCSI), food consumption (FCS), and multidimensional deprivation. Estimated effects in Beirut are smaller and mostly statistically indistinguishable from zero, with the exception of a reduction in coping strategies and a modest improvement in multidimensional deprivation. Estimates for Saida and Tripoli are more mixed. In Saida, the program is associated with a large and statistically significant improvement in food consumption, but effects on other outcomes are imprecisely estimated. In Tripoli, point estimates vary in sign and magnitude across outcomes, and none of the primary poverty or coping outcomes is estimated with precision. Sample sizes in Saida and Tripoli are substantially smaller than in Beirut and West Bekaa, leading to wider confidence intervals and greater uncertainty around the estimated effects. Accordingly, we emphasize the contrast between Beirut and West Bekaa, where estimates are both more precisely measured and more informative about systematic geographic heterogeneity.

Focusing on broad patterns in the discussion topics, the qualitative data yield two insights: housing costs, medical expenses, and food insecurity are salient across all districts, but evidence of the salient binding constraints varies markedly by location. West Bekaa is characterized by a rural informal settlement context where housing payments, though frequently mentioned, are less likely to accumulate as formal debt obligations. Respondents in this district frequently describe direct pathways from cash transfers to food consumption, including purchasing food, meeting children's needs, and reducing meal skipping. Bekaa also has the highest rate of mentions of employment unavailability (6.9% vs. 2.8% in Beirut). These features appear to allow transfers to translate more directly into improvements in consumption and food security. Relative to West Bekaa, Beirut respondents more often report that transfers are absorbed by housing-related debt (6.7% vs. 2.5% incidence, respectively) and institutional frictions such as documentation (5.1% vs. 2.5%) and school access (6.3% vs. 2.5%) barriers. These constraints limit the extent to which additional cash translates into measurable gains in consumption or broader well-being outcomes. Another striking difference between these two locales is the degree to which macro-level shocks were referenced: in Bekaa, the crises related to basic needs – the fuel/gas subsidy removal and bread shortages – were referenced far more than in Beirut (4.4% and 5.4% vs 1.6% and 0.8%, respectively) while Beirut respondents indicated greater difficulties arising from the diffuse impacts spanning consumption, services, and household functioning associated with currency devaluation and COVID-related restrictions (4.0% and 4.7% vs. 1.0% and 0.0%, respectively).

Taken together, the quantitative estimates and qualitative evidence indicate that identical transfer values interact with distinct, locally binding constraints across districts. As a result, program impacts vary systematically across locations, with larger effects emerging in settings where cash assistance can be more readily converted into consumption and food security improvements. These patterns underscore that limits to scalability arise not only from program design but also from persistent differences in local economic and institutional environments.

7 Discussion

This study presents experimental evidence of substantial heterogeneity in program effects from an at-scale social protection program implemented in Lebanon. The program targeted forcibly

Table 5: Local average treatment effect estimates by district and outcome

	Expenditure per capita	rCSI	FCS	MDDI
Beirut				
Estimate	0.019	0.184***	0.075	0.115*
Std. Error	(0.056)	(0.057)	(0.058)	(0.060)
N	933	935	935	799
Saida				
Estimate	0.042	0.096	0.454***	0.018
Std. Error	(0.116)	(0.152)	(0.151)	(0.158)
N	200	203	203	202
Tripoli				
Estimate	0.147	-0.266	0.095	0.227
Std. Error	(0.215)	(0.207)	(0.195)	(0.209)
N	142	142	142	137
West Bekaa				
Estimate	0.271**	0.472***	0.571***	0.264**
Std. Error	(0.125)	(0.130)	(0.142)	(0.118)
N	452	458	458	456

Note: Table reports district-specific treatment effect estimates for alternative welfare and coping outcomes. Each cell presents the estimated coefficient from a district-by-outcome regression specification, with heteroskedasticity-robust standard errors in parentheses. Poverty outcomes are unit standardized and expenditure per capita is measured in logarithms. Statistical significance is indicated by $^{*}p < .10$; $^{**}p < .05$; $^{***}p < .01$. **Reading:** Each district block should be read independently. For example, in West Bekaa, treatment increases log cash expenditure per capita by 0.27 SD and is associated with statistically significant improvements in rCSI, FCS, and dietary diversity, while having no detectable effect on school enrollment. In contrast, estimates for Beirut are smaller and generally statistically indistinguishable from zero, except for a reduction in coping strategies.

displaced Syrian refugee households in a fragile institutional setting characterized by political instability, social tension between host and refugee communities, and limited access to formal insurance and labor markets. These features make the setting both policy relevant and analytically challenging, as program implementation coincided with a sequence of severe and overlapping macroeconomic shocks.

During the 2021–2022 study period, Lebanon experienced hyperinflation, sharp currency depreciation, the breakdown of public services, and cascading disruptions to food, fuel, housing, and labor markets. These shocks sharply increased poverty and vulnerability for both refugee and host populations, while simultaneously undermining the functioning of local markets. In such an environment, predicting *ex ante* how cash transfers would translate into welfare improvements is inherently difficult even under uniform program design and implementation.

Our study shows that alternative geographic budget allocation rules can effectively prioritize populations with different demographic and socioeconomic characteristics. While allocation rules shift which households receive larger transfers, they do not materially change the local market environment in which transfers are spent. Identical transfer amounts interact with local constraints and informal institutions that differ markedly across districts, generating pronounced heterogeneity in treatment effects.

These findings align with a broader body of work emphasizing the importance of site-specific heterogeneity (Allcott, 2015), context dependence (Pritchett and Sandefur, 2015), and the limits of external validity in experimental evaluations conducted across settings and scales (Banerjee et al., 2017; List, 2022). In our setting, even though the program was implemented uniformly and simultaneously at the national level, treatment effects vary markedly across districts, suggesting that unobserved and place-specific factors can generate sizable and difficult-to-predict differences in treatment effects.

Qualitative evidence helps clarify some of the mechanisms underlying this heterogeneity. Focus group discussions reveal that households often face multiple competing needs, including debt repayment, eviction risk driven by accumulated rent obligations, rising medical expenses, and constrained labor market opportunities. In locations where accumulated housing debt and rent obligations dominate household budgets, cash transfers are frequently absorbed by debt repayment or rent, limiting their impact on consumption and other welfare measures. In contrast, in settings with lower fixed costs and more informal housing arrangements, transfers more readily translate into improved food security and reductions in adverse coping strategies. These mechanisms highlight how program impacts depend not only on household vulnerability, but also on local market structure and institutional context in ways that are difficult to anticipate at the design stage.

While the magnitude of estimated effects is shaped by Lebanon's exceptional macroeconomic collapse, the underlying mechanism we document is not unique to this setting.[15] The interaction between cash transfers and locally binding market frictions is likely to arise in many contexts in which social protection programs are scaled under post-conflict conditions, weak institutional capacity, or heightened macroeconomic volatility. Our results therefore suggest that geographic heterogeneity driven by local constraints is likely to be a general feature of large-scale social protection programs, even if the specific magnitudes and outcome profiles remain context dependent. As a result, understanding how and why program impacts differ requires close attention to local economic conditions and social dynamics.

Our study highlights a central challenge for the science of scaling social protection programs. Even when implemented uniformly at the national level, program impacts vary sharply across locations due to persistent differences in locally binding economic and institutional constraints. Scaled experiments therefore provide critical benchmarks for average effects and for detecting heterogeneity, but they do not, on their own, identify where uniform program designs will be most

[15] The World Bank characterizes Lebanon's economic and financial crisis as "among the top three most severe crises globally since the mid-nineteenth century" (World Bank, 2021).

effective. Our findings point to the importance of treating local context as a core object of analysis in scaling decisions, rather than as residual noise. For policymakers, this implies that effective scaling may require adaptive program components, context-specific complements, or deliberate variation across markets. For researchers, it underscores the value of multi-site designs and context-aware evaluations that explicitly anticipate how local constraints mediate impacts at scale.

References

ABELSON, B., K. R. VARSHNEY, AND J. SUN (2014): "Targeting direct cash transfers to the extremely poor," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1563–1572.

AIKEN, E., S. BELLUE, D. KARLAN, C. UDRY, AND J. E. BLUMENSTOCK (2022): "Machine learning and phone data can improve targeting of humanitarian aid," *Nature*, 603, 864–870.

AKER, J. C. (2017): "Comparing Cash and Voucher Transfers in a Humanitarian Context: Evidence from the Democratic Republic of Congo," *The World Bank Economic Review*, 31, 44–70.

ALATAS, V., A. BANERJEE, R. HANNA, B. A. OLKEN, R. PURNAMASARI, AND M. WAI-POI (2016): "Self-Targeting: Evidence from a Field Experiment in Indonesia," *Journal of Political Economy*, 124, 371–427.

ALATAS, V., A. BANERJEE, R. HANNA, B. A. OLKEN, AND J. TOBIAS (2012): "Targeting the Poor: Evidence from a Field Experiment in Indonesia," *American Economic Review*, 102, 1206–40.

ALLCOTT, H. (2015): " Site Selection Bias in Program Evaluation *," *The Quarterly Journal of Economics*, 130, 1117–1165.

ALTINDAĞ, O. AND S. O'CONNELL (2022): "The short-lived effects of unconditional cash transfers to refugees," .

ALTINDAĞ, O., S. O'CONNELL, A. ŞAŞMAZ, Z. BALCIOĞLU, P. CADONI, M. JERNECK, AND A. KUNZE FOONG (2021): "Targeting humanitarian aid using administrative data: model design and validation," *Journal of Development Economics*, 148.

ANDERSON, M. L. (2008): "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103, 1481–1495.

ASHER, S., T. LUNT, R. MATSUURA, AND P. NOVOSAD (2021): "Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in India using the shrug open data platform," *The World Bank Economic Review*, 35, 845–871.

AYGÜN, A., M. G. KIRDAR, M. KOYUNCU, AND Q. STOEFFLER (2021): "Keeping Refugee Children in School and Out of Work: Evidence from the World's Largest Humanitarian Cash Transfer Program," *IZA Discussion Paper*.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukerji, M. Shotland, and M. Walton (2017): "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application," *Journal of Economic Perspectives*, 31, 73–102.

Basurto, M. P., P. Dupas, and J. Robinson (2020): "Decentralization and efficiency of subsidy targeting: Evidence from chiefs in rural Malawi," *Journal of Public Economics*, 185, 104047.

Blumenstock, J., G. Cadamuro, and R. On (2015): "Predicting poverty and wealth from mobile phone metadata," *Science*, 350, 1073–1076.

Brown, C., M. Ravallion, and D. van de Walle (2018): "A poor means test? Econometric targeting in Africa," *Journal of Development Economics*, 134, 109–124.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernández-Val (2025): "Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments," *Econometrica*, 93, 1165–1170.

Crosta, T., D. Karlan, F. Ong, J. Rüschenpöhler, and C. R. Udry (2024): "Unconditional Cash Transfers: A Bayesian Meta-Analysis of Randomized Evaluations in Low and Middle Income Countries," Working Paper 32779, National Bureau of Economic Research.

Egger, D., J. Haushofer, E. Miguel, P. Niehaus, and M. Walker (2022): "General Equilibrium Effects of Unconditional Cash Transfers: Experimental Evidence from Kenya," *Econometrica*, 90, 2603–2643.

Elbers, C., T. Fujii, P. Lanjouw, B. Özler, and W. Yin (2007): "Poverty alleviation through geographic targeting: How much does disaggregation help?" *Journal of Development Economics*, 83, 198–213.

Ferrario, A. and S. Stantcheva (2022): "A Guide to Using Open-Ended Questions in Surveys," *American Economic Association: Papers and Proceedings*, 112, 1–20.

Garlick, R., J. Walker, and K. Orkin (2025): "Heterogeneity in Effects of Poverty-Alleviating Interventions," *AEA Papers and Proceedings*, 115, 335–339.

Govt. of Lebanon & United Nations (2023): "Lebanon Crises Response Plan 2023," Tech. rep., Government of Lebanon and United Nations.

Hanna, R. and D. Karlan (2017): "Chapter 7 - Designing Social Protection Programs: Using Theory and Experimentation to Understand How to Help Combat Poverty," in *Handbook of*

Economic Field Experiments, ed. by A. V. Banerjee and E. Duflo, North-Holland, vol. 2 of *Handbook of Economic Field Experiments*, 515–553.

HANNA, R. AND B. A. OLKEN (2018): "Universal basic incomes versus targeted transfers: Anti-poverty programs in developing countries," *Journal of Economic Perspectives*, 32, 201–226.

HAUSHOFER, J., P. NIEHAUS, C. PARAMO, E. MIGUEL, AND M. W. WALKER (2022): "Targeting Impact versus Deprivation," Working Paper 30138, National Bureau of Economic Research.

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.

KARLAN, D. AND B. THUYSBAERT (2019): "Targeting Ultra-Poor Households in Honduras and Peru," *The World Bank Economic Review*, 33, 63–94.

KINNAN, C., K. SAMPHANTHARAK, R. TOWNSEND, AND D. A. V. COSSIO (2020): "Propagation and insurance in village networks," Tech. rep., National Bureau of Economic Research.

KURDI, S. (2021): "The nutritional benefits of cash transfers in humanitarian crises: evidence from Yemen," *World Development*, 148, 105664.

LEHMANN, M. C. AND D. T. R. MASTERSON (2020): "Does Aid Reduce Anti-refugee Violence? Evidence from Syrian Refugees in Lebanon," *American Political Science Review*, 114, 1335–1342.

LIST, J. A. (2022): *The Voltage Effect*, Penguin Books Limited.

MACPHERSON, C. AND O. STERCK (2021): "Empowering refugees through cash and agriculture: A regression discontinuity design," *Journal of Development Economics*, 149, 102614.

MASTERSON, D. AND M. C. LEHMANN (2020): "Refugees, Mobilization, and Humanitarian Aid: Evidence from the Syrian Refugee Crisis in Lebanon," *Journal of Conflict Resolution*, 64, 817–843.

PREMAND, P. AND P. SCHNITZER (2020): "Efficiency, Legitimacy, and Impacts of Targeting Methods: Evidence from an Experiment in Niger," *The World Bank Economic Review*, 35, 892–920.

PRITCHETT, L. AND J. SANDEFUR (2015): "Learning from Experiments When Context Matters," *American Economic Review*, 105, 471–75.

QUATTROCHI, J., J. C. AKER, P. VAN DER WINDT, AND M. VOORS (2020): “Contributions of experimental approaches to development and poverty alleviation: Field experiments and humanitarian assistance,” *World Development*, 127, 104830.

RAVALLION, M. (2009): “How Relevant Is Targeting to the Success of an Antipoverty Program?” *The World Bank Research Observer*, 24, 205–231.

SCHWAB, B. (2019): “Comparing the Productive Effects of Cash and Food Transfers in a Crisis Setting: Evidence from a Randomised Experiment in Yemen,” *The Journal of Development Studies*, 55, 29 – 54.

SMYTHE, I. S. AND J. E. BLUMENSTOCK (2022): “Geographic microtargeting of social assistance with high-resolution poverty maps,” *Proceedings of the National Academy of Sciences*, 119, e2120025119.

STANTCHEVA, S. (2024): “Understanding People’s Beliefs, Misperceptions, and Policy Preferences,” NBER Working Paper 32421, National Bureau of Economic Research, Cambridge, MA.

STERCK, O. AND A. DELIUS (2020): “Cash Transfers and Micro-Enterprise Performance: Theory and Quasi-Experimental Evidence from Kenya,” Tech. rep.

STERCK, O. C., C. RODGERS, J. SIU, M. F. STIERNA, AND A. BETTS (2020): “Cash transfer models and debt in the Kalobeyei settlement,” .

STOEFFLER, Q., B. MILLS, AND C. DEL NINNO (2016): “Reaching the Poor: Cash Transfer Program Targeting in Cameroon,” *World Development*, 83, 244–263.

UNHCR (2023): “2022 Update of the Non-Food Expenditure Basket: Basic Assistance Sector - September 2022,” Tech. rep., United Nations High Commissioner for Refugees.

WELZ, M., A. ALFONS, M. DEMIRER, AND V. CHERNOZHUKOV (2022): *GenericML: Generic Machine Learning Inference*, r package version 0.2.3.

WORLD BANK (2021): “Lebanon Economic Monitor: Lebanon Sinking (to the Top 3),” Tech. rep., World Bank, Washington, DC.

WORLD FOOD PROGRAMME (2008a): “The Coping Strategies Index Field Methods Manual,” Tech. rep.

——— (2008b): “Food consumption analysis: Calculation and use of the food consumption score in food security analysis,” Tech. rep.

——— (2023): “Essential Needs Assessment Guidance Note,” Tech. rep.

Appendix Figures and Tables

Appendix Table 1: Means and tests of assignment balance in endline sample

Measure	Means by budget allocation arm: Monetary poverty	Food insecurity	Food consumption	Multidim. deprivation	Tests: F	p-value	N
Panel A: Full endline sample							
Household size	4.43	4.50	4.43	4.50	0.65	0.58	6914
% aged 0-5	17.42	17.92	18.52	18.07	0.87	0.46	6914
% aged 50 +	8.26	8.82	8.75	7.79	0.88	0.45	6914
% male aged 18-50	19.12	19.07	18.48	19.37	0.57	0.64	6914
% female headed	21.34	20.14	20.31	21.65	0.59	0.62	6914
% has disabled member	15.93	15.85	15.07	15.30	0.23	0.87	6914
% no education	11.93	12.59	12.60	11.57	0.62	0.60	6914
% secondary education	31.44	30.87	30.66	30.97	0.15	0.93	6914
pred. exp. p.c. (000 LBP)	285.18	283.65	283.41	284.01	0.20	0.90	6914
Panel B: Marginal beneficiaries in endline sample							
Household size	4.22	4.24	4.16	4.23	0.32	0.81	3978
% aged 0-5	17.55	18.57	18.96	18.70	0.89	0.45	3978
% aged 50 +	8.37	9.20	9.65	8.05	1.13	0.33	3978
% male aged 18-50	18.79	18.55	18.84	18.77	0.05	0.99	3978
% female headed	20.91	20.41	18.98	20.82	0.49	0.69	3978
% has disabled member	14.10	14.82	14.18	14.51	0.09	0.97	3978
% no education	10.52	11.20	11.92	10.86	0.55	0.65	3978
% secondary education	32.44	31.85	30.42	31.81	0.56	0.64	3978
pred. exp. p.c. (000 LBP)	287.61	286.11	289.49	286.46	0.58	0.63	3978

Note: Table presents means and tests of covariate balance among marginal beneficiaries. Data come from administrative records prior to treatment assignment. Panels indicate the sample used, and the means of each variable are presented across columns for each treatment arm sample. Panel A contains the full post-intervention sample, and Panel B contains these tests for the marginal beneficiary population. The F statistic and its corresponding p value come from the joint hypothesis tests that mean differences across all subgroups relative to the monetary poverty arm are zero. **Reading:** Among marginal beneficiaries assigned to the monetary poverty targeting arm, the average household size was 4.22 (see Panel B). Average household size among marginal beneficiaries assigned to other targeting arms were 4.24, 4.16, and 4.23. An F-test fails to reject the joint null hypothesis that the latter three means are equal to 4.22. Overall, randomized assignment to targeting arms achieved balance in baseline covariates among the full endline sample and among marginal beneficiaries in the endline sample. Baseline tests in the full sample are available in Table 3 of the pre-analysis plan, and similarly showed strong balance across targeting arms.

Appendix Table 2: CLAN estimates by covariate and GATES group: Reduced Coping Strategies Index (rCSI)

Variable	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	$\Delta_5 - \Delta_1$
Household composition and age structure						
Household size	4.070	4.291	4.341	4.270	3.901	-0.187*
Female-headed household	0.178	0.199	0.216	0.223	0.185	0.007
Share of household members aged 0–5	0.189	0.182	0.181	0.181	0.163	-0.026**
Share of household members aged 6–10	0.111	0.119	0.121	0.125	0.113	0.005
Share of household members aged 11–17	0.106	0.122	0.131	0.133	0.114	0.007
Share of male members aged 18–50	0.189	0.180	0.174	0.175	0.204	0.014
Share of household members aged 50+	0.057	0.086	0.096	0.095	0.104	0.044***
Labor and education						
Share of adults working as laborers	0.263	0.248	0.235	0.243	0.263	0
Share of adults working in services	0.082	0.083	0.084	0.094	0.143	0.059***
Share of adults with medium education (levels 3–4)	0.514	0.462	0.455	0.470	0.484	-0.034
Social protection						
Single parent	0.051	0.070	0.080	0.085	0.058	0.009
Share of household members with medical condition	0.154	0.146	0.139	0.125	0.119	-0.033***
Share of household members with a disability	0.004	0.007	0.011	0.010	0.007	0.003
Previous cycle cash transfer beneficiary	0.336	0.420	0.474	0.499	0.389	0.042*
Previous cycle food assistance beneficiary	0.267	0.298	0.276	0.229	0.165	-0.105***
Log predicted per capita expenditure (baseline)	12.565	12.535	12.530	12.540	12.614	0.047***
District of residence						
Akkar	0.114	0.098	0.069	0.057	0.034	-0.077***
Minieh-Dennieh	0.093	0.067	0.050	0.044	0.045	-0.047***
Zgharta	0.066	0.038	0.028	0.023	0.022	-0.044***
Koura	0.051	0.028	0.019	0.013	0.009	-0.042***
Aley	0.051	0.032	0.025	0.020	0.012	-0.041***
Zahle	0.092	0.142	0.131	0.101	0.054	-0.042***
Tripoli	0.048	0.028	0.018	0.013	0.012	-0.038***
Batroun	0.029	0.019	0.015	0.009	0.006	-0.022***
Hermel	0.026	0.028	0.025	0.016	0.004	-0.02***
Sour (Tyre)	0.045	0.047	0.038	0.034	0.028	-0.019*
Keserwan	0.034	0.026	0.023	0.020	0.020	-0.012
Marjaayoun	0.036	0.037	0.038	0.035	0.029	-0.007
Nabatieh	0.058	0.036	0.029	0.031	0.053	-0.007
Hasbaya	0.006	0.012	0.010	0.009	0.004	-0.001
Bint Jbeil	0.007	0.009	0.009	0.009	0.006	-0.001
Rachaya	0.010	0.054	0.061	0.039	0.010	0
Jbeil	0.010	0.010	0.014	0.015	0.022	0.012*
Jezzine	0.004	0.006	0.009	0.013	0.019	0.015***
Chouf	0.004	0.007	0.009	0.010	0.020	0.015**
Saida	0.026	0.022	0.019	0.023	0.048	0.023**
Metn	0.020	0.032	0.038	0.045	0.050	0.026***
West Bekaa	0.006	0.044	0.093	0.111	0.058	0.05***
Baabda	0.026	0.034	0.036	0.044	0.086	0.058***
Baalbek	0.012	0.032	0.080	0.126	0.102	0.091***
Beirut	0.085	0.091	0.097	0.130	0.218	0.137***

Note: Table reports CLAN estimates of covariate means for effects on reduced coping strategies. Columns Δ_1 through Δ_5 report mean values of each covariate within quintiles of the estimated conditional average treatment effect, where Δ_1 corresponds to households with the lowest predicted gains from treatment and Δ_5 to those with the highest predicted gains. The final column reports the difference in means between the highest- and lowest-gain groups ($\Delta_5 - \Delta_1$). Sample based on marginal beneficiaries in post-intervention survey. Sample size is 6,847. Statistical significance of this difference is indicated by *p $<$ .10; **p $<$ .05; ***p $<$ .01. **Reading:** Households predicted to benefit more from the program (higher GATES groups) are more likely to be female-headed or single-parent households, and have lower baseline predicted per capita expenditure.

Appendix Table 3: CLAN estimates by covariate and GATES group: Food Consumption Score (FCS)

Variable	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	$\Delta_5 - \Delta_1$
Household composition and age structure						
Household size	3.866	4.190	4.259	4.272	4.236	0.385***
Female-headed household	0.219	0.210	0.203	0.199	0.169	-0.05**
Share of household members aged 0–5	0.168	0.185	0.181	0.180	0.180	0.012
Share of household members aged 6–10	0.090	0.107	0.120	0.131	0.140	0.05***
Share of household members aged 11–17	0.112	0.126	0.130	0.127	0.114	0.001
Share of male members aged 18–50	0.184	0.182	0.181	0.182	0.196	0.01
Share of household members aged 50+	0.142	0.094	0.079	0.067	0.051	-0.09***
Labor and education						
Share of adults working as laborers	0.215	0.239	0.253	0.265	0.274	0.06***
Share of adults working in services	0.091	0.084	0.090	0.097	0.126	0.037***
Share of adults with medium education (levels 3–4)	0.422	0.442	0.470	0.504	0.554	0.13***
Social protection						
Single parent	0.061	0.079	0.080	0.073	0.051	-0.01
Share of household members with medical condition	0.177	0.147	0.133	0.119	0.110	-0.071***
Share of household members with a disability	0.013	0.009	0.007	0.006	0.004	-0.009***
Previous cycle cash transfer beneficiary	0.460	0.467	0.435	0.409	0.354	-0.107***
Previous cycle food assistance beneficiary	0.222	0.269	0.277	0.264	0.196	-0.031
Log predicted per capita expenditure (baseline)	12.586	12.545	12.541	12.544	12.569	-0.018
District of residence						
Zahle	0.121	0.146	0.117	0.082	0.045	-0.076***
Zgharta	0.058	0.044	0.034	0.025	0.019	-0.041***
Sour (Tyre)	0.066	0.039	0.030	0.029	0.028	-0.04***
Minieh-Dennieh	0.063	0.070	0.072	0.063	0.035	-0.028**
Baabda	0.060	0.047	0.047	0.042	0.034	-0.026**
Koura	0.032	0.032	0.025	0.020	0.010	-0.019***
Tripoli	0.031	0.029	0.025	0.022	0.012	-0.019**
Beirut	0.147	0.118	0.111	0.109	0.130	-0.019
Akkar	0.062	0.082	0.088	0.088	0.055	-0.006
Batroun	0.016	0.016	0.018	0.016	0.013	-0.003
Hermel	0.015	0.023	0.026	0.020	0.015	0
Keserwan	0.026	0.020	0.020	0.025	0.028	0
Hasbaya	0.003	0.009	0.012	0.010	0.007	0.003
Aley	0.025	0.029	0.028	0.026	0.032	0.006
Nabatieh	0.041	0.035	0.037	0.042	0.050	0.012
Rachaya	0.020	0.028	0.042	0.047	0.034	0.015
Baalbek	0.073	0.063	0.066	0.070	0.088	0.016
Jezzine	0.004	0.006	0.009	0.013	0.022	0.019***
Marjaayoun	0.026	0.028	0.032	0.038	0.047	0.02**
Jbeil	0.006	0.010	0.010	0.016	0.028	0.019***
Bint Jbeil	0.003	0.001	0.003	0.006	0.026	0.023***
Saida	0.010	0.022	0.032	0.041	0.035	0.025***
Metn	0.028	0.028	0.034	0.041	0.058	0.034***
Chouf	0.001	0.003	0.006	0.007	0.034	0.031***
West Bekaa	0.026	0.047	0.066	0.086	0.086	0.06***

Note: Table reports CLAN estimates of covariate means for effects on food consumption score. Columns Δ_1 through Δ_5 report mean values of each covariate within quintiles of the estimated conditional average treatment effect, where Δ_1 corresponds to households with the lowest predicted gains from treatment and Δ_5 to those with the highest predicted gains. The final column reports the difference in means between the highest- and lowest-gain groups ($\Delta_5 - \Delta_1$). Sample based on marginal beneficiaries in post-intervention survey. Sample size is 6,847. Statistical significance of this difference is indicated by *p $<$ .10; **p $<$ .05; ***p $<$ .01. **Reading:** Households predicted to benefit more from the program (higher GATES groups) are more likely to be female-headed or single-parent households, and have lower baseline predicted per capita expenditure.

Appendix Table 4: CLAN estimates by covariate and GATES group: Multidimensional Deprivation Index (MDDI)

Variable	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	$\Delta_5 - \Delta_1$
Household composition and age structure						
Household size	4.244	4.280	4.280	4.206	4.197	-0.045
Female-headed household	0.193	0.207	0.202	0.208	0.217	0.018
Share of household members aged 0–5	0.169	0.184	0.191	0.195	0.176	0.008
Share of household members aged 6–10	0.123	0.123	0.118	0.114	0.123	0.001
Share of household members aged 11–17	0.130	0.122	0.121	0.120	0.126	-0.006
Share of male members aged 18–50	0.186	0.178	0.174	0.170	0.165	-0.021**
Share of household members aged 50+	0.076	0.080	0.084	0.083	0.092	0.016
Labor and education						
Share of adults working as laborers	0.269	0.243	0.240	0.238	0.241	-0.028*
Share of adults working in services	0.070	0.088	0.096	0.106	0.120	0.049***
Share of adults with medium education (levels 3–4)	0.448	0.462	0.473	0.489	0.490	0.043**
Social protection						
Single parent	0.069	0.081	0.072	0.068	0.063	-0.006
Share of household members with medical condition	0.136	0.137	0.131	0.134	0.134	-0.005
Share of household members with a disability	0.011	0.008	0.007	0.006	0.005	-0.005*
Previous cycle cash transfer beneficiary	0.432	0.445	0.431	0.424	0.426	-0.003
Previous cycle food assistance beneficiary	0.288	0.282	0.253	0.227	0.214	-0.07***
Log predicted per capita expenditure (baseline)	12.534	12.531	12.538	12.551	12.565	0.031***
District of residence						
Minieh-Dennieh	0.105	0.071	0.054	0.036	0.029	-0.072***
Akkar	0.108	0.097	0.077	0.060	0.038	-0.069***
Zgharta	0.077	0.045	0.027	0.018	0.012	-0.065***
Keserwan	0.068	0.031	0.017	0.009	0.003	-0.066***
Zahle	0.117	0.131	0.113	0.095	0.072	-0.051***
Koura	0.050	0.027	0.018	0.014	0.009	-0.041***
Rachaya	0.047	0.044	0.038	0.029	0.021	-0.024**
Hermel	0.029	0.030	0.024	0.015	0.006	-0.021***
Aley	0.038	0.031	0.029	0.026	0.021	-0.02**
Sour (Tyre)	0.041	0.041	0.044	0.042	0.032	-0.009
Tripoli	0.026	0.024	0.024	0.023	0.021	-0.005
Batroun	0.017	0.017	0.018	0.018	0.014	-0.003
Saida	0.021	0.033	0.039	0.032	0.020	-0.002
Bint Jbeil	0.005	0.008	0.011	0.009	0.005	0
Chouf	0.009	0.008	0.009	0.012	0.012	0.003
Marjaayoun	0.023	0.035	0.045	0.045	0.029	0.005
Metn	0.026	0.036	0.045	0.045	0.039	0.011
Jezzine	0.002	0.006	0.009	0.011	0.020	0.017***
Hasbaya	0.002	0.003	0.006	0.011	0.021	0.02***
West Bekaa	0.038	0.063	0.074	0.074	0.074	0.036***
Jbeil	0.002	0.006	0.009	0.017	0.039	0.036***
Nabatieh	0.019	0.035	0.045	0.050	0.060	0.039***
Baabda	0.026	0.029	0.033	0.045	0.092	0.063***
Baalbek	0.047	0.053	0.066	0.089	0.113	0.063***
Beirut	0.032	0.077	0.113	0.160	0.166	0.13***

Note: Table reports CLAN estimates of covariate means for effects on multidimensional deprivation. Columns Δ_1 through Δ_5 report mean values of each covariate within quintiles of the estimated conditional average treatment effect, where Δ_1 corresponds to households with the lowest predicted gains from treatment and Δ_5 to those with the highest predicted gains. The final column reports the difference in means between the highest- and lowest-gain groups ($\Delta_5 - \Delta_1$). Sample based on marginal beneficiaries in post-intervention survey. Sample size is 6,847. Statistical significance of this difference is indicated by $^{*}p < .10$; $^{**}p < .05$; $^{***}p < .01$. **Reading:** Households predicted to benefit more from the program (higher GATES groups) are more likely to be female-headed or single-parent households, and have lower baseline predicted per capita expenditure.