Distributional preference divergence in targeting foreign aid: Experimental evidence from aid workers, refugees, and a proxy means test in a humanitarian setting

Zeynep Balcioglu
Leiden University

Stephen D. O'Connell
Emory University

Aytug Sasmaz
Bryn Mawr College

July 2, 2026

Abstract

Aid institutions delivering social assistance balance donors' conceptions of appropriate program design with the need to maintain legitimacy among local beneficiaries. We present a conceptual framework for institutional targeting under dual accountability constraints and apply it to the case of cash transfer targeting among Syrian refugees in Lebanon. We recover distributional preferences for beneficiary selection in a discrete choice conjoint experiment given to humanitarian staff and refugees, and extract implicit preferences from the econometric proxy means test used in practice. Human respondents prioritize visible forms of vulnerability, while formal targeting models emphasize demographic predictors of expenditure. We find that donor compliance pressures outweigh legitimacy costs in shaping institutional behavior. The paper contributes to debates on governance, accountability, and institutional legitimacy in humanitarian and foreign aid settings.

Keywords: Cash Transfers; Targeting; Humanitarian Assistance; Refugees; Conjoint Analysis; Proxy Means Test; Institutional Legitimacy; Principal-Agent Problems

JEL codes: I38, O15, D63, D73, F22

Word count: 7961 (Monterey Language Services tool, excluding references and appendix)

Acknowledgments: We thank the staff of UNHCR Lebanon and WFP Lebanon for their collaboration and support during fieldwork. We are especially grateful to the numerous refugee outreach volunteers and humanitarian workers who participated in the study. This study was approved by the Emory University Institutional Review Board (Protocol 2026P000536). All views expressed are those of the authors and do not necessarily represent those of any affiliated organizations.

1 Introduction

How do international organizations balance competing accountability pressures when implementing policy? States delegate authority to IOs with the expectation of faithful execution, yet IOs simultaneously depend on perceived legitimacy among the populations they serve (Levi, 1997; Tallberg et al., 2018). The formal structure of delegation shapes the control mechanisms available to donors and the degree of autonomy afforded to implementing agencies (Hawkins et al., 2006), but international organizations are not passive transmission belts: they develop internal bureaucratic cultures and operational norms that mediate how donor mandates are interpreted and applied (Barnett and Finnemore, 2004). When upward accountability to donors and downward accountability to beneficiaries pull in different directions, institutional outcomes reveal which pressures dominate—and how organizational hierarchy filters the competing demands.

This paper develops an empirical strategy for recovering the relative strength of these pressures from observed policy outcomes. We study the allocation of humanitarian cash assistance to Syrian refugees in Lebanon, where a UN agency must choose which households to serve using a desk formula for eligibility ranking while maintaining legitimacy among displaced populations and navigating competing priorities within its own organizational hierarchy. The setting presents an instance of the common agency problem (Bernheim and Whinston, 1986): multiple principals—donor states, implementing staff at different organizational levels, and beneficiaries—simultaneously exert competing influence over the targeting agent. Donor states leverage financial support and monitoring mechanisms to influence implementers' behavior, often conditioning institutional autonomy on perceived performance or alignment with donor preferences (Brutger and Clark, 2023; Dietrich, 2016). Beneficiaries, concerned with fairness and lacking formal electoral channels for grievance, can impose costs on an implementer perceived as illegitimate—through non-cooperation, reputational damage, or appeals to alternative authorities (Easterly, 2006; Fox, 2007). Institutional behavior in these fragmented governance environments, where multiple international actors vie for influence and resources (Clark, 2021), is shaped by competing pressures that political dynamics within donor countries can amplify (Carnegie et al., 2024).

Targeted social assistance programs offer a particularly revealing window into these tensions. Cash transfers have become a cornerstone of social protection in contexts ranging from stable low-income countries to fragile and conflict-affected states (Gentilini, 2022; Harvey, 2005; Fiszbein and Schady, 2009). Yet targeting methods remain contentious. Technical approaches like proxy means tests offer standardized criteria but often fail to capture the full complexity of household vulnerability (Coady et al., 2004; Hanna and Olken, 2018). Participatory

approaches reflect local knowledge but are vulnerable to bias and elite capture (Platteau, 2004; Alatas et al., 2012). Even technically efficient models can produce distributions of aid that communities perceive as arbitrary or unfair (Alatas et al., 2012). In fragile contexts, where political accountability is weak, perceived fairness in resource distribution can substitute for formal legitimacy mechanisms (Pritchett et al., 2013; Mansuri and Rao, 2013; Olken, 2010), making the alignment—or misalignment—between institutional targeting rules and local conceptions of need a direct measure of how organizations navigate competing accountability demands.

Despite extensive literatures on delegation in IOs and on targeting in development economics, few studies directly measure whether the actors within an implementing organization—at different levels of the hierarchy—prioritize the same characteristics when allocating aid, or how their preferences compare to the formal rules the organization applies. This gap matters for the study of international organizations because it is precisely at the interface between organizational layers that competing accountability pressures are mediated. Whether central staff internalize donor priorities, field staff absorb local norms, or the implemented model reflects pressures not visible in any actor's stated preferences—these are empirical questions that speak to fundamental debates about agency autonomy, bureaucratic drift, and institutional legitimacy in global governance (Barnett and Finnemore, 1999; Nielson and Tierney, 2003).

This paper addresses these questions by developing a conceptual framework for institutional behavior under dual accountability pressures. Building on work examining delegation chains in foreign aid (Martens et al., 2002; Nielson and Tierney, 2003), we structure the problem as a common agency setting in which the implementing institution balances its own preferences, donor compliance incentives, and local legitimacy costs when choosing a targeting strategy. The framework yields testable expectations about which accountability pressures dominate institutional decision-making and provides a structure for empirically recovering the relative strength of these pressures from observed targeting preferences.

We make three contributions. First, we advance the study of IO accountability by recovering the *revealed* distributional preferences of actors at distinct levels of an implementing organization's hierarchy—central office staff, field implementers, and beneficiaries—and comparing them to the preferences embedded in the desk formula used to determine eligibility. This approach moves beyond survey-based measures of IO legitimacy or case-study accounts of organizational behavior to provide a quantitative decomposition of where, within an organization, competing accountability pressures bind. Second, we contribute to the delegation and common agency literatures by providing an empirical test that distinguishes among

competing models of institutional behavior—donor compliance, local legitimacy-seeking, internal compromise, and bureaucratic autonomy—using the implied targeting rankings of each actor group as the observable outcome. Third, we add to the targeting and social protection literatures by documenting the magnitude and character of divergence between econometric targeting and human judgments of vulnerability, showing that disagreement is not marginal but pervasive, driven by fundamentally different conceptions of what constitutes need.

Empirically, we recover distributional preferences from humanitarian staff, refugee outreach volunteers, and the desk formula used to allocate cash assistance to Syrian refugees in Lebanon. Using a discrete choice conjoint experiment, we compare these priorities and assess how divergences between human preferences and the implemented formula reveal underlying institutional constraints. We find that the desk formula lies closest to central office staff preferences across all divergence measures, consistent with donor-facing compliance as the dominant accountability pressure. Field staff and beneficiaries, whose preferences are relatively close to each other, diverge sharply from the desk formula—indicating that downward accountability exerts limited influence on actual targeting outcomes. When we translate these preference divergences into counterfactual targeting decisions, we find that when the budget allows the agency to serve only 25 percent of the population, more than 90 percent of the formula's selections would be contested by human actors.

2 Theoretical Foundations

This study builds on theories of principal-agent problems in aid delivery and institutional legitimacy, framing humanitarian cash transfer targeting as a problem of dual accountability: external donors monitor compliance with technical rules (Hawkins et al., 2006; Martinez-Bravo, 2017; Mansuri and Rao, 2013), while beneficiaries judge institutions through perceptions of fairness and procedural justice (Coady et al., 2004; Hanna and Olken, 2018). Proxy means tests offer standardized criteria for selecting beneficiaries but often miss dimensions of vulnerability salient to local populations, and agents closer to the ground may observe needs that desk formulas cannot capture. This extends the common agency framework (Bernheim and Whinston, 1986) to an applied political economy context in which competing principals shape aid implementation.

Theories of institutional legitimacy further highlight that institutions secure compliance not only through material incentives but also through perceived fairness (Levi, 1997; Bénabou and Tirole, 2006; North, 1990). International organizations develop internal bureaucratic cultures that shape how rules are interpreted, often in ways that diverge from both donor

intent and local expectations (Barnett and Finnemore, 2004), and their legitimacy depends on the perceived alignment of processes and outcomes with stakeholder values (Tallberg et al., 2018). In fragile contexts where formal political accountability is limited, perceptions of targeting fairness can substitute for traditional mechanisms of legitimacy (Devarajan et al., 2011; Pritchett et al., 2013); misalignments between technical criteria and local fairness norms can undermine trust, leading to dissatisfaction or disengagement among beneficiaries (Kosec and Mo, 2023; Stokes, 2005).

3 Conceptual Framework

Development organizations and social programs often operate under competing pressures: to deliver services in ways that are locally legitimate and inclusive, and to comply with rules and reporting standards imposed by external donors. To structure the empirical analysis, we present a conceptual framework in which an implementing institution balances its own preferences, donor compliance incentives, and local legitimacy costs when choosing a targeting strategy. The framework is intentionally parsimonious: it defines the key actors, their objectives, and the divergence measures that connect theory to data, without imposing equilibrium or deriving comparative statics. Its purpose is to organize the empirical expectations we test, not to characterize optimal institutional behavior.

Players and Objective

The baseline setup consists of three agents:

1) The **institution** (I) chooses a targeting strategy T_t and is seeking to sustain funding and local legitimacy while maximizing over its own preferences.
2) The **local beneficiaries** (B) form perceptions of legitimacy based on the institution's alignment with local expectations.
3) The **donor** (D) monitors adherence to its own preferred targeting strategy and allocates funding based on observed technical compliance.

The institution's objective function is:

$$\max_{T_t} W = u(-d_I) + R(-d_D) - C(-d_B)$$

where d_I is the divergence between the ideal targeting strategy of I (T_I) and the implemented one (T_t), d_D is the divergence between the ideal targeting strategy of D (T_D) and the implemented one (T_t), and d_B is the divergence between the ideal targeting strategy of B (T_B) and the implemented one (T_t).

The functions $u(\cdot)$, $R(\cdot)$, and $C(\cdot)$ have the natural monotonicity properties: institutional payoff falls in own divergence, donor revenue falls in donor divergence, and legitimacy cost rises in beneficiary divergence. Appendix A states the formal assumptions.

Divergence Measures

What is observable to actors in practice is not a vector of preference weights but the resulting *ranking* of households for assistance. Rank-based divergence measures are therefore the natural empirical counterpart to the distances d_I, d_D, and d_B in the framework: they capture disagreement over who should be prioritized and in what order, rather than over abstract coefficient magnitudes. We require measures that are symmetric, bounded, and sensitive to the ordering of elements. The specific metrics used—including the Kendall tau distance, Spearman rho distance, and top-half overlap measures—are defined in the Analysis section below.

Expectations

The framework describes how institutions balance upward and downward accountability constraints when choosing targeting strategies. By examining the relative divergence of the implemented strategy from the ideal strategies of donors, the institution itself, and beneficiaries, we can infer which accountability pressures dominate institutional decision-making.

In our empirical setting, we do not directly observe donor preferences. Instead, we observe:

- T_t: The actual targeting strategy implemented by the institution (the desk formula),
- T_C: Preferences of country or central-office staff (a proxy for donor-facing internal constraints),
- T_F: Preferences of field staff (a proxy for operational and contextual judgment),
- T_B: Preferences of beneficiaries or the local population (a proxy for the legitimacy anchor).

This gives us four positions in the targeting space. The donor's targeting preference T_D remains latent.[1] By analyzing the relative divergences among these four observed positions,

[1]We treat central-office preferences as the closest observable proxy for donor-facing accountability pressures, since country-office staff are most directly responsible for reporting to donors and ensuring compliance with program rules. An alternative interpretation is that central staff preferences align with the desk formula not because of donor compliance pressure but because of greater familiarity with the PMT's construction and logic. We cannot fully distinguish between these channels with the present design. However, the conjoint instrument presented unlabeled household profiles and did not reference the targeting formula, reducing the scope for mechanical anchoring on known PMT criteria. We return to this issue when interpreting the results.

we can infer which accountability pressures—upward towards donors or downward towards beneficiaries—dominate institutional decision-making and how internal mediation processes reflect those pressures.

The framework yields six expectations, each mapping a distinct accountability mechanism to a pattern of relative proximity among T_t, T_C, T_F, and T_B. Table 1 summarizes them; the empirical analysis evaluates each in turn.

Table 1: Expectations on Patterns of Relative Preference Divergence

Expectation	**Pattern of Relative Preference Divergence**
1. Donor-facing compliance	$\lvert T_t - T_C\rvert \ll \lvert T_t - T_F\rvert$ and $\lvert T_t - T_C\rvert \ll \lvert T_t - T_B\rvert$
2. Local legitimacy	$\lvert T_t - T_B\rvert \ll \lvert T_t - T_C\rvert$ and $\lvert T_t - T_B\rvert \ll \lvert T_t - T_F\rvert$
3. Field-level pragmatism	$\lvert T_t - T_F\rvert \ll \lvert T_t - T_C\rvert$ and $\lvert T_t - T_F\rvert \ll \lvert T_t - T_B\rvert$
4. Internal bureaucratic compromise	$\lvert T_C - T_F\rvert \gg 0$ and $T_t \in [T_C, T_F]$
5. Latent donor targeting	None of the above, and $\lvert T_t - T_C\rvert \gg 0$ and $\lvert T_t - T_F\rvert \gg 0$ and $\lvert T_t - T_B\rvert \gg 0$
6. No meaningful accountability trade-off	$T_t \approx T_C \approx T_F \approx T_B$

Notes: T_t, T_C, T_F, and T_B denote the targeting preferences of the desk formula, central office staff, field staff, and beneficiaries, respectively. $\lvert T_a - T_b\rvert \ll \lvert T_a - T_c\rvert$ indicates that a's preferences are much closer to b's than to c's. Each row predicts a pattern of relative proximity among these four positions consistent with a distinct accountability mechanism dominating institutional targeting.

Using the Framework

The framework enables us to infer the relative bindingness of legitimacy versus compliance constraints using only observed targeting strategies and stakeholder preferences. By locating the actual targeting decision T_t in the space defined by T_B, T_D, and T_I, and measuring the divergence of T_t from each ideal strategy, we recover institutional trade-offs and their likely accountability anchor through revealed-preference reasoning. Because donor preferences are not directly observable, the framework triangulates across implemented allocations and stakeholder preferences using the divergence measures outlined above.

4 Context and Actors

4.1 Syrian Refugees in Lebanon

Since 2011, Lebanon has hosted approximately 1.5 million Syrian refugees, the highest per capita refugee population in the world. Lebanon enforces a no-camp policy, resulting in refugees being widely dispersed across urban, rural, and informal settlements alongside host

communities. Socioeconomic conditions facing Syrian refugees are dire: most have limited access to legal residency, are largely restricted to the informal labor market, and experience exploitation, poverty, and precarious living conditions.

The situation worsened after Lebanon's economic collapse in 2019, which strained livelihoods and increased the demand for social assistance. According to the 2022 Vulnerability Assessment of Syrian Refugees in Lebanon (UNHCR, 2022), 90 percent of Syrian refugees require assistance to meet basic needs, and nearly 89 percent report resorting to harmful coping strategies such as reducing expenditures on food, healthcare, and education.

4.2 The Cash Transfer Program

The humanitarian response to Syrian displacement in Lebanon is heavily cash-based, with a joint UN cash transfer program serving as the largest initiative. The program provides monthly unconditional cash transfers to refugee households deemed socioeconomically vulnerable.

Eligibility for the cash transfer is determined by a desk formula — an annually recalibrated proxy means test that predicts household expenditure from demographic and socioeconomic variables collected in administrative records. The formula scores and ranks households centrally based on weights determined by an econometric prediction model, with those falling below a budget-determined threshold deemed eligible for assistance (Altındağ et al., 2021). This results in an arbitrary cut-off that leaves many households excluded from the program that are similarly vulnerable to those eligible for assistance. In our field consultations with staff and refugees (described in Section 5), respondents consistently reported confusion and resentment over this process. Because economic conditions among Syrian refugees in Lebanon are relatively homogeneous, small predicted differences in expenditure—invisible to community members—can determine eligibility. Respondents described the selection process as arbitrary or akin to a lottery, despite some general awareness of the desk formula. Staff and refugees also identified specific vulnerable groups—such as small households, elderly individuals, or those living in remote areas—that the formula's reliance on per-capita expenditure tends to exclude. The implementing agency has attempted to address these gaps through quotas and supplemental programs, but grievances persist. Complicating matters further, the agency occupies a quasi-state role for Syrian refugees in Lebanon, providing protection, health services, shelter assistance, and livelihoods support (Kagan, 2011; Deardorff Miller, 2017). Dissatisfaction with targeting practices thus has implications for the perceived legitimacy of humanitarian governance.

4.3 Actors and Units of Analysis

Our analysis focuses on four sets of actors:

- **Central Office Staff:** Individuals involved in the design and oversight of the cash transfer targeting system.
- **Field-Level Implementers:** Staff working directly with refugee communities, mediating between institutional policies and local realities.
- **Beneficiaries:** Syrian refugees eligible for or affected by the cash transfer program, whose perspectives reflect lived experiences of vulnerability and perceptions of fairness.
- **The desk formula:** The proxy means test used to determine program eligibility, whose simulated conjoint responses serve as a benchmark. The construction of this benchmark is described in Section 5.2.

5 Study Design and Data Collection

To ensure that our conjoint experiment reflected the actual considerations shaping vulnerability assessments in practice, we developed its attributes through a structured, two-stage process. This approach allowed us to ground the design in real-world institutional experience and to reflect the divergence between formal model-based targeting and human judgments.

Stage One: Focus Groups and Structured Fieldwork. We conducted focus group discussions with field staff and refugee outreach volunteers across multiple regions of Lebanon, alongside a structured qualitative exercise in which participants were presented with anonymized real-world household profiles and asked to (a) determine whether each household should receive assistance, and (b) make pairwise choices between households under a hypothetical resource constraint. These tasks approximated the decisions frontline actors make in practice, while revealing the attributes that drove prioritization.

Both activities revealed consistent tensions between the desk formula and how staff and community members judged vulnerability on the ground. Participants placed high priority on households with elderly or disabled heads, serious medical conditions, single female heads of household, and those facing high dependency burdens or recent displacement. Particular concern was expressed for households with children exposed to protection risks or those relying on harmful coping mechanisms such as begging or sending children to work. In contrast, households with a working-age male, shared housing arrangements, or longer duration in Lebanon were often viewed as relatively less vulnerable, even when their predicted expenditure was low. These patterns underscored the disconnect between formal targeting formulas and local judgments, and highlighted vulnerability dimensions that were both socially salient and

underrecognized in the existing targeting system.

Stage Two: Translation to Experimental Design. The most frequently cited indicators of vulnerability in staff and refugee judgments became the attributes used in the conjoint experiment. These were selected to span economic, demographic, health, and social protection factors, with particular emphasis on dimensions that drove disagreement between staff and the desk formula. The resulting design allows us to evaluate how these attributes are prioritized by different actors and to compare human preferences with the desk formula used to allocate aid.

5.1 Conjoint Experiment Design

Building on the findings of the focus groups, we designed a discrete choice conjoint experiment to systematically elicit distributional preferences from humanitarian staff and refugee outreach volunteers. The experiment was intended to mimic real-world decision-making under resource constraints, while allowing us to isolate the marginal effect of different household characteristics on prioritization judgments.

In each task, respondents were shown a pair of hypothetical refugee household profiles and asked to select which household should be prioritized for cash assistance. Profiles were described using attributes identified through the qualitative stages, with levels reflecting the range of variation observed in the field. Attributes captured key dimensions of vulnerability—including disability status, serious medical conditions, household composition, asset adequacy, and child protection risks—that had been repeatedly cited by field actors as important but underemphasized in the desk formula.

Each attribute was randomly assigned one level per profile, and the resulting profiles were independently randomized within each choice set. Respondents completed four paired choice tasks, yielding eight profile evaluations per respondent. The experiment was administered online via Qualtrics in March 2021, with instruments available in both English and Arabic. Respondents completed the survey remotely on personal devices at a time of their choosing. The conjoint experiment was designed and fielded by the implementing agency as an internal program improvement exercise within its cash transfer operation in Lebanon. The goal was to elicit staff and volunteer priorities for targeting criteria as an input to operational decision-making, and participation took place in the ordinary course of agency program activities.

In total, 35 central office staff, 155 field staff, and 380 refugee outreach volunteers completed

the experiment, producing 1,520 staff observations and 3,040 refugee observations.[2] This structure allows us to estimate average marginal component effects (AMCEs) for each attribute level on the likelihood of selection, separately for each respondent group.

Table 2 summarizes the ten attributes and their associated levels used in the experiment. The design aimed to balance realism and parsimony: levels were crafted to reflect actual vulnerability conditions observed in the field, while maintaining cognitive tractability for respondents. Household descriptions were anonymized and neutrally worded to avoid priming or social desirability effects.

[2]The central office staff sample is small in absolute terms but represents a near-census of the relevant population within the country operation. Methodological guidance on discrete choice experiments indicates that as few as 20 respondents per version can yield reliable aggregate preference estimates when each respondent completes multiple choice tasks (Lancsar and Louviere, 2008). A systematic review of 69 healthcare DCEs found that 32 percent had total sample sizes below 100, with small expert and specialist samples treated as standard practice (de Bekker-Grob et al., 2015). Our 35 central office respondents, each completing eight profile evaluations, produce 280 choice observations—well above the minimum thresholds suggested by these benchmarks. We further note that our core findings are robust across all four rank-based divergence measures, reducing concern that the central staff results are driven by sampling noise.

Table 2: Attributes and Levels Used in Conjoint Experiment

Attribute	Levels
Arrival Year	2013; 2016; 2020
City	Beirut; Tripoli; Zahle; Nabatieh
Adult Education	None; Primary; Secondary
Household Composition	Large family with two elderly members and five children; Large family with five children; Small family with elderly member; Small family with single female head of household; Very large family with eleven children
Disability Status	No disability; Head of household with disability; Non–head of household with disability; Two household members with disabilities
Medical Condition	No medical condition; Head of household with medical condition; Non–head of household with medical condition
Household Assets	Adequate; Mildly inadequate; Highly inadequate
Debt Situation	No debt; Small and flexible debt; Large and flexible debt; Large and urgent debt
Coping Strategy	No coping strategy; Begging; High-risk jobs; Selling non-productive goods; Selling productive goods
Assistance History	Receiving assistance; Not receiving assistance

Note: The table lists all attributes and levels included in the conjoint experiment used to elicit preferences for household targeting. Each respondent was shown a series of paired household profiles randomly drawn from these attributes and levels, and asked to indicate which household should be prioritized for assistance. Attributes capture key vulnerability dimensions commonly used in humanitarian targeting frameworks, including health, dependency, economic status, and prior support. All levels were pretested to ensure contextual relevance and clarity for respondents. *Reading:* Each row represents an attribute whose possible values (levels) were randomly varied across household profiles. For example, the "Medical Condition" attribute alternated between "No medical condition," "Head of household with medical condition," and "Non–head of household with medical condition."

The randomized attribute structure allows us to estimate the marginal effect of each characteristic on selection probability—the AMCE coefficients that serve as implicit weights when translated into a scoring function—enabling direct comparison across institutional layers and between human preferences and the desk formula.

5.2 Benchmarking Against the Desk Formula

To compare human judgments directly against formal program rules, we constructed a benchmark version of the conjoint experiment using the desk formula. We generated a synthetic dataset of 2,000 hypothetical household profiles with the same attributes and levels as the conjoint design, applied the official desk formula coefficients and scoring procedures

from the 2020–2021 targeting algorithm to calculate predicted household expenditures, and randomly paired profiles to identify the lower-expenditure profile in each pair. This translates the formula's continuous expenditure predictions into the same binary choice format faced by human respondents, enabling direct comparison of AMCEs across all four samples—staff responses, refugee responses, and formula-based simulations.

5.3 Prediction Sample

To translate estimated preferences into counterfactual targeting rankings, we require a common set of real household profiles that can be scored under each group's AMCE coefficients. We construct this prediction sample by drawing on two administrative data sources from the same program context: the 2022 Vulnerability Assessment of Syrian Refugees in Lebanon (VASyR), which provides household-level measures of asset adequacy, debt burden, and coping behavior; and the agency's 2022 program administration records, which provide demographic and socioeconomic characteristics including household size and composition, disability status, medical conditions, geographic location, arrival year, adult education, and assistance history. We merge these sources at the household level using masked case identifiers.

Because the conjoint experiment uses categorical attribute levels rather than continuous variables, we map each household's observed characteristics into the closest conjoint level. Arrival year is binned to the nearest design value (2013, 2016, or 2020); geographic location is aggregated from district to the four regional categories used in the experiment (Beirut, Tripoli, Zahle, Nabatieh); household composition is classified into the five conjoint categories based on household size, presence of elderly members, and head-of-household gender; and disability, medical condition, and education variables are binned using threshold rules applied to household-level shares. Economic variables—asset adequacy, debt situation, and coping strategy—enter directly from VASyR using the same categorical scales as the conjoint design.

The resulting prediction sample contains 5,070 registered refugee households. Each household is characterized by the same ten attributes used in the conjoint experiment, enabling us to apply each group's estimated AMCE vector as a linear scoring function and produce a complete preference-based ranking of the sample for each actor group.

6 Analysis

6.1 Empirical Tests Using Conjoint Data

1. Recovering Preference Vectors

Separate AMCE models are estimated for each group using a specification of the form:

$$Y_{ij} = \alpha_j + \mathbf{X}'_{ij}\boldsymbol{\beta}_g + \varepsilon_{ij}$$

where Y_{ij} is the choice or ranking outcome, $\mathbf{X}_{ij}$ is the vector of profile attributes, and $\boldsymbol{\beta}_g$ is the AMCE vector for group $g \in \{B, F, C, t\}$. These estimates constitute the empirical analogs to the theoretical T_B, T_F, T_C, and T_t.

2. Testing Differences in Preferences Across Groups

To assess whether different stakeholder groups express statistically distinct targeting preferences, we estimate pairwise pooled models comparing two groups at a time. For each pairwise comparison, we pool the corresponding data and include interaction terms between attribute levels and a binary group indicator. The model is specified as:

$$Y_{ij} = \alpha_j + \sum_k X_{ijk} \cdot \beta_k + \sum_k (X_{ijk} \cdot G_g) \cdot \delta_k + \varepsilon_{ij}$$

where X_{ijk} indicates attribute level k in profile i for respondent j, G_g is a group indicator, β_k captures AMCEs for the reference group, and δ_k captures differential effects for group g.

3. Quantifying Divergence in Targeting Rankings

Pairwise pooled models identify whether any attribute-level effects differ across groups, but they do not provide a joint metric of how far apart the targeting decisions implied by each group's preferences actually are. Because what is observable to actors in the real world is not a vector of regression coefficients but the resulting *ranking* of households for assistance, we focus on rank-based divergence measures that directly capture disagreement over who should be prioritized.[3]

[3]As a complementary approach, we also report in Appendix A tests based on Mahalanobis distances between AMCE vectors with associated Wald tests for pairwise divergence. While these provide a useful multivariate summary of coefficient-level differences, they operate in the space of estimated attribute weights rather than in the space of implied targeting decisions, making them less directly interpretable in terms of the framework's accountability structure.

To construct empirical analogs of the targeting strategies T_B, T_F, T_C, and T_t, we extend each group's estimated AMCE model into a preference-based scoring function. Using the prediction sample described in Section 5.3, we apply each group's estimated AMCE coefficients to produce a predicted priority score $\hat{s}_{ig} = \mathbf{X}_i'\hat{\boldsymbol{\beta}}_g$ for every household, inducing a complete ranking for each group. We then compare rankings pairwise using the Kendall tau distance, Spearman rho distance, Spearman footrule distance, and top-half mismatch rate motivated in Section 3.

6.2 Distributional Preferences Across Stakeholders

Figure 1 reports the average marginal component effects (AMCEs) estimated separately for the four groups: central office staff, field staff, refugee respondents, and the desk formula. AMCE coefficients and standard errors are available in Appendix Table 1, and marginal means are reported in Appendix Figure 1. Across groups, the sign and magnitude of coefficients reveal distinct patterns in how each actor prioritizes household characteristics, vulnerability indicators, and demographic compositions when evaluating hypothetical cases.

All human respondent groups—central staff, field staff, and refugees—display positive valuations for households with high debt burdens and inadequate assets. Households with "large and urgent" or "large and flexible" debt and with "highly inadequate" assets receive the highest positive coefficients, indicating a shared humanitarian logic emphasizing acute material deprivation. Arrival year exerts weak and inconsistent effects: earlier arrivals (2016) are slightly penalized, especially by central staff, consistent with a mild preference for recently displaced households.

Refugees prioritized many of the same markers as humanitarian staff but exhibited distinct sensitivities. They placed stronger weight on households with members who were disabled or chronically ill, with positive coefficients for both "head with medical condition" and "multiple disabilities." They also penalized very large households with many working-age members more heavily than staff, and—consistent with qualitative focus-group findings—strongly favored households resorting to begging as a coping strategy, interpreting it as an indicator of child labor or extreme economic distress. Refugees further emphasized supporting families who had not previously received assistance, a dimension indirectly reflected in the high valuation of acute coping behaviors.

Geographic location generates pronounced divergence across groups. Central staff assign moderate positive weights to all regions—particularly Nabatieh, Tripoli, and Zahle—indicating perceived disadvantage in peripheral areas. Field staff's preferences are closer to zero, reflecting local familiarity and smaller perceived regional gaps. By contrast, the desk formula assigns

very large positive weights, especially for Zahle (≈ 0.39) and Tripoli (≈ 0.33), implying that algorithmic prioritization amplifies geographic differentiation far beyond what human respondents imply.

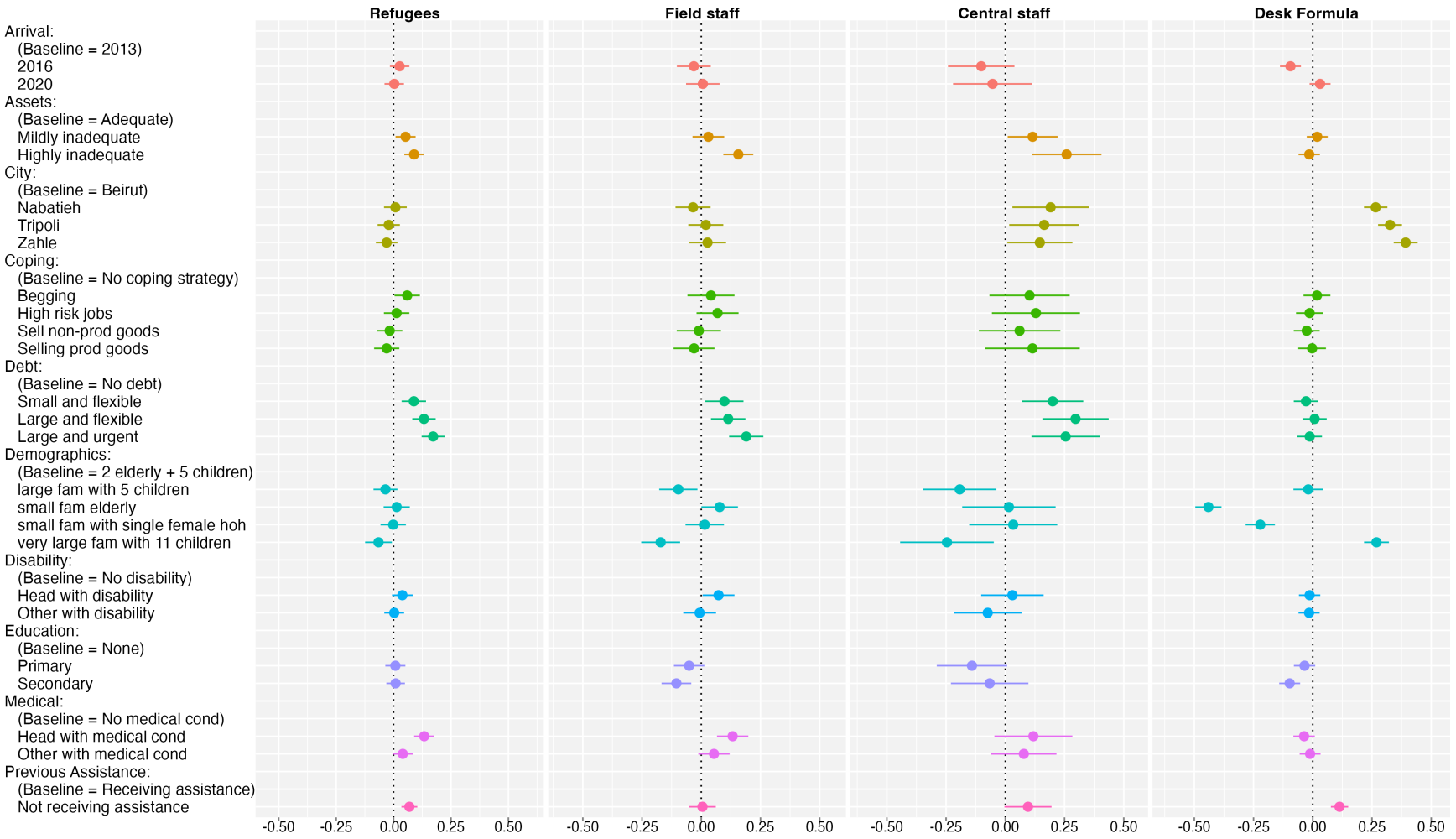
Indicators of economic distress such as begging, high-risk jobs, and selling productive assets receive positive weights from all human groups, but central and refugee respondents emphasize them more than the field group. The desk formula, in contrast, down-weights or neutralizes these behavioral signals, likely because such variables are highly correlated with asset or debt indicators already embedded in its expenditure-based prediction algorithm.

The starkest divergences arise for family composition. Central and field staff both penalize very large families with many dependents and are broadly neutral toward small families headed by women or with elderly members. The desk formula reverses these patterns, favoring larger households (AMCE = 0.27 for eleven children) and sharply penalizing elderly and female-headed households (-0.44 and -0.22, respectively). These results suggest that the proxy means test's demographic structure favors household size as a predictor of consumption potential, not vulnerability per se, introducing a systematic bias against nontraditional or low-labor-capacity households.

Human respondents consistently favor households with health or disability-related challenges, although effect sizes are modest (AMCEs around 0.1–0.13). The desk formula again diverges, assigning near-zero or negative coefficients to the same attributes, indicating limited sensitivity to direct welfare deficits not captured by expenditure regressors.

Taken together, the AMCEs reveal three systematic features. First, human respondents across all groups privilege markers of acute financial stress and health vulnerability, forming a coherent humanitarian logic of need. Second, the desk formula—optimized for expenditure prediction—places greater emphasis on demographic and regional correlates of consumption than on direct vulnerability cues such as disability, chronic illness, or coping behavior. Third, the largest divergences occur for demographic and regional characteristics, where the formula's prioritization diverges most sharply from both staff and refugee preferences. This divergence highlights the trade-off between technical accuracy and perceived fairness, motivating the next section in which we examine these differences more formally by comparing the implied targeting rankings of each group using the rank-based divergence measures defined above.

Figure 1: AMCEs across profile characteristics — central staff, field staff, refugees, and the desk formula



Notes: Average marginal component effects (AMCEs) from a conjoint experiment in which respondents selected which of two refugee household profiles should receive cash assistance. Points are estimates; horizontal lines are 95% confidence intervals. An AMCE of 0.10 indicates a 10 percentage point increase in selection probability relative to the baseline level (listed in gray). Human-group estimates are based on 35 central staff, 155 field staff, and 380 refugee outreach volunteers (4,560 choice observations, March 2021). Desk formula coefficients derive from the proxy means test applied to 5,070 households in the prediction sample.

6.3 Testing framework expectations using rank-based divergence measures

Table 3 reports pairwise divergences of the implied targeting rankings across all six pairs among {Central, Field, Beneficiary, Desk Formula}, using the Kendall tau distance, Spearman rho distance, Spearman footrule distance, and top-half mismatch rate defined in the preceding subsection. Bootstrap standard errors (10,000 replications) are reported in parentheses.

Table 3: Pairwise rank divergence measures (bootstrap SE)

Group pair	Kendall τ	Spearman ρ	Spearman footrule	Top-half mismatch
Beneficiary vs Field	0.274	0.178	0.393	0.234
	(0.058)	(0.079)	(0.077)	(0.073)
Beneficiary vs Central	0.315	0.235	0.434	0.270
	(0.062)	(0.087)	(0.080)	(0.073)
Beneficiary vs Desk Formula	0.699	0.786	0.873	0.700
	(0.056)	(0.077)	(0.056)	(0.067)
Field vs Central	0.261	0.165	0.366	0.248
	(0.058)	(0.076)	(0.078)	(0.064)
Field vs Desk Formula	0.649	0.740	0.868	0.680
	(0.045)	(0.064)	(0.046)	(0.053)
Central vs Desk Formula	0.556	0.587	0.776	0.560
	(0.056)	(0.086)	(0.070)	(0.070)

Notes: Pairwise rank divergence between implied targeting rankings, computed by applying each group's AMCE vector as a scoring function to 5,070 prediction-sample households and comparing the resulting rankings. Bootstrap standard errors (10,000 replications) in parentheses. All four measures are normalized to $[0, 1]$; top-half mismatch reports the share of one group's top-half selections absent from the other's.

Table 4 reports pairwise differences in Spearman ρ distance with bootstrap confidence intervals and p-values, enabling formal comparison of whether any two pairs of actors differ significantly in their divergence from one another.

Table 4: Pairwise differences in Spearman ρ distance (Δ) with bootstrap inference

Comparison	Delta	95% CI	p-value
[Beneficiary vs Field] vs [Beneficiary vs Central]	-0.0569	[-0.2433, 0.1283]	0.5828
[Beneficiary vs Field] vs [Beneficiary vs Desk Formula]	-0.6077**	[-0.6841, -0.1903]	0.0044
[Beneficiary vs Field] vs [Field vs Central]	0.0131	[-0.1818, 0.2445]	0.8152
[Beneficiary vs Field] vs [Field vs Desk Formula]	-0.5620***	[-0.6537, -0.1939]	0.0006
[Beneficiary vs Field] vs [Central vs Desk Formula]	-0.4092*	[-0.5186, -0.0672]	0.0140
[Beneficiary vs Central] vs [Beneficiary vs Desk Formula]	-0.5508**	[-0.6295, -0.1598]	0.0036
[Beneficiary vs Central] vs [Field vs Central]	0.0700	[-0.1096, 0.2732]	0.4216
[Beneficiary vs Central] vs [Field vs Desk Formula]	-0.5052**	[-0.5963, -0.1745]	0.0014
[Beneficiary vs Central] vs [Central vs Desk Formula]	-0.3524	[-0.5145, 0.0505]	0.0996
[Beneficiary vs Desk Formula] vs [Field vs Central]	0.6208***	[0.2710, 0.6921]	0.0000
[Beneficiary vs Desk Formula] vs [Field vs Desk Formula]	0.0456	[-0.1801, 0.2109]	0.7810
[Beneficiary vs Desk Formula] vs [Central vs Desk Formula]	0.1984	[-0.0597, 0.3924]	0.1442
[Field vs Central] vs [Field vs Desk Formula]	-0.5752***	[-0.6499, -0.2637]	0.0002
[Field vs Central] vs [Central vs Desk Formula]	-0.4224*	[-0.5688, -0.0372]	0.0288
[Field vs Desk Formula] vs [Central vs Desk Formula]	0.1528	[-0.0587, 0.3544]	0.1748

Notes: Each row tests whether two pairwise Spearman ρ distances differ significantly. Bootstrap confidence intervals and p-values from 10,000 replications. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

The rank-based divergence measures reveal a clear pattern. First, all three human groups are relatively close to one another (Kendall tau 0.26–0.31; Spearman rho 0.17–0.24), while each is far from the desk formula (Kendall tau 0.56–0.70; Spearman rho 0.59–0.79). The pairwise delta tests for Spearman ρ confirm that these human–formula gaps are highly significant ($p < 0.01$ for all comparisons), while the within-human distances are not statistically distinguishable from one another. Point estimates for the remaining three measures are broadly consistent with this pattern (Table 3).

Second, among the three human groups, the desk formula lies closest to central office staff across all four measures (e.g., Kendall tau of 0.556 vs 0.649 for field and 0.699 for beneficiaries). Point estimates for all four divergence measures place the Central-vs-Formula distance below the Field-vs-Formula and Beneficiary-vs-Formula distances, though the pairwise delta tests for Spearman ρ do not reject equality at conventional levels ($p = 0.17$ and $p = 0.14$, respectively). This pattern is consistent with **Expectation 1** (donor-facing compliance): the implemented targeting strategy most closely reflects the preferences of central staff, who are most directly exposed to donor accountability pressures. As noted above, an alternative reading is that central staff preferences align with the desk formula because of greater exposure to PMT methodology rather than donor compliance incentives per se. The conjoint design mitigates this concern—profiles were unlabeled and did not reference targeting criteria—but we cannot fully rule out that institutional familiarity with the formula's logic shapes how central staff evaluate vulnerability. Under either interpretation, however, the substantive finding holds:

institutional targeting outcomes are closest to the preferences of the organizational layer with the greatest proximity to formal program rules, whether that proximity operates through accountability channels or informational ones.

Expectations 2 (local legitimacy) and **3** (field-level pragmatism) are not supported: the desk formula diverges most—not least—from beneficiary and field staff preferences. **Expectation 4** (internal compromise) finds limited support: while central and field staff differ modestly, the desk formula does not lie between them but rather beyond central staff in the direction of greater divergence from both field and beneficiary positions. **Expectation 5** (latent donor targeting) is partially consistent with the data, since the desk formula diverges substantially from all observed human groups, though its relative proximity to central staff is more consistent with institutional mediation than with pure donor imposition. **Expectation 6** (no meaningful trade-off) is firmly rejected.[4]

6.4 Targeting consequences of preference divergence

The rank-based divergence measures establish that the desk formula and human groups disagree systematically over who should be prioritized. We now quantify the practical targeting consequences of this disagreement by asking: if a program operates under a fixed budget, how many of the households selected by the desk formula would also be selected by each human group?

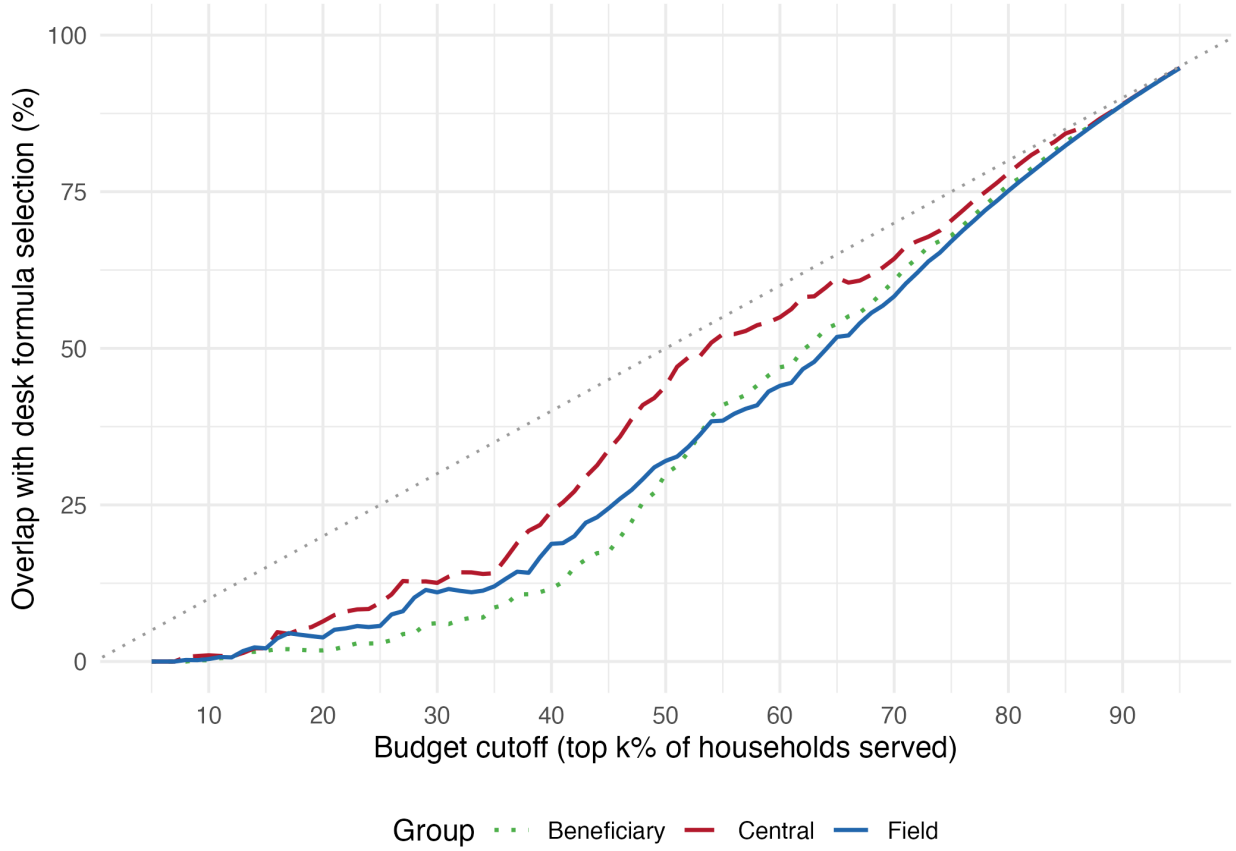
To construct this exercise, we score each of the 5,070 households in the prediction sample using the AMCE-derived preference weights for each group, rank them from most to least preferred, and compare the resulting selection sets at each budget cutoff from 5 to 95 percent of the population.

Figure 2 plots the share of formula-selected households that each human group would also select, across budget cutoffs. If the desk formula and a human group agreed perfectly, the overlap curve would reach 100 percent at every cutoff. If their rankings were independent, overlap would fall along the diagonal (equal to the cutoff itself). Across the full range of cutoffs, all human groups fall below the random-agreement diagonal, indicating that the formula and human rankings are not merely noisy versions of the same ordering but reflect systematically different prioritization criteria. At a 50 percent budget cutoff, only 30 to 44 percent of formula-selected households would also be selected by the human groups, with central staff showing the highest overlap (44%) and beneficiaries the lowest (30%). At the more restrictive 25 percent cutoff—closer to the budget constraints faced by many

[4]Appendix A reports complementary tests using Mahalanobis distances between AMCE coefficient vectors with associated Wald tests, which yield broadly consistent conclusions.

humanitarian programs—overlap drops below 10 percent for all groups. Even central staff, whose preferences most closely track the desk formula, would contest more than 9 in 10 targeting decisions at this cutoff.

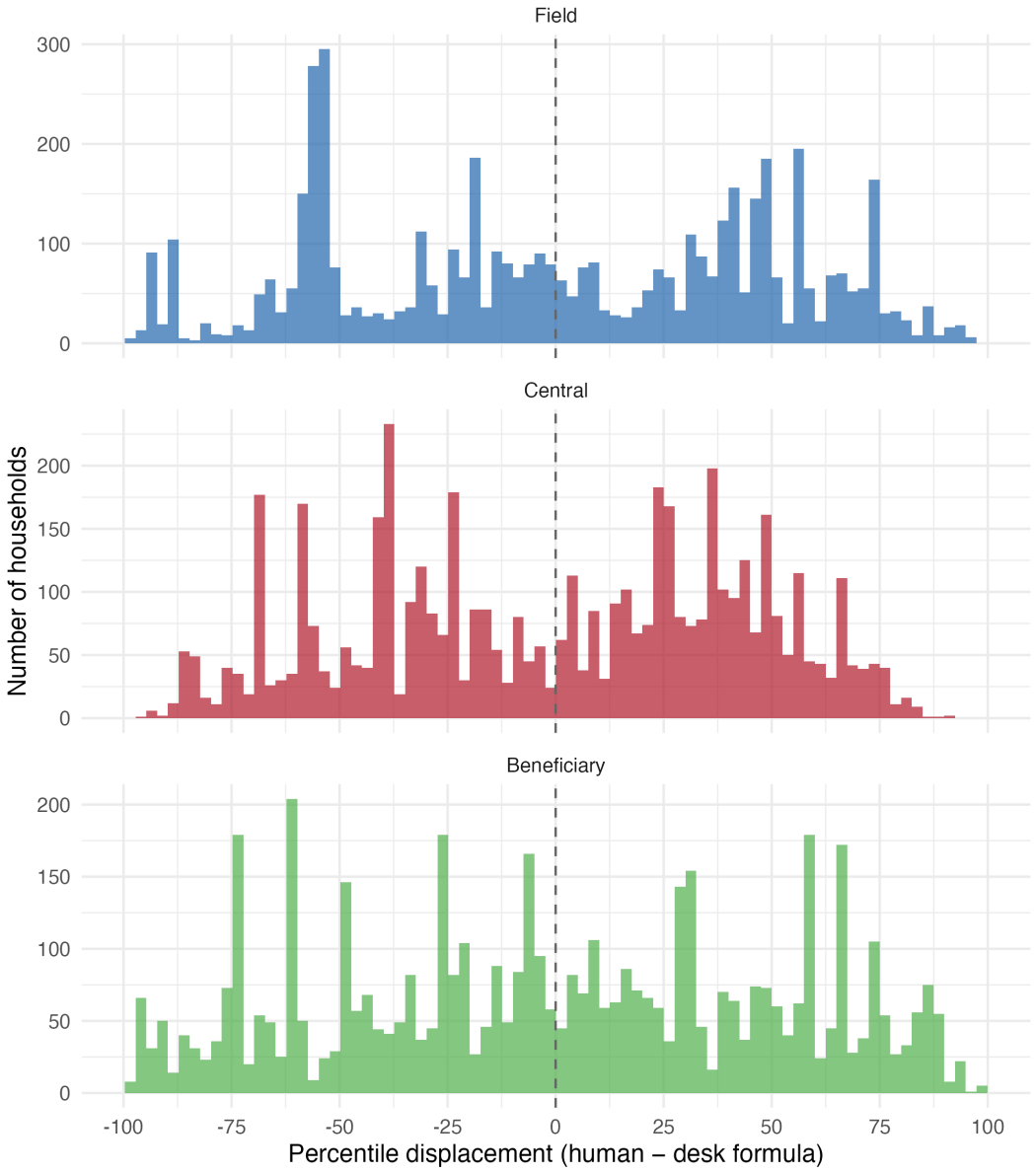
Figure 2: Selection overlap between the desk formula and human groups across budget cutoffs



Notes: Share of desk-formula-selected households that each human group would also select, at budget cutoffs from 5% to 95% of the 5,070-household prediction sample. Selection sets are determined by ranking households using AMCE-derived preference weights and choosing the top k%. The dotted diagonal indicates overlap expected under independent rankings. Central staff (dashed red) show the highest overlap at every cutoff; beneficiaries (dotted green) the lowest.

Figure 3 shows the distribution of percentile displacement—the change in a household's percentile rank when moving from the desk formula's ranking to each human group's ranking. The distributions are wide and roughly uniform, spanning nearly the full range from -100 to $+100$ percentile points. Across all three human groups, more than 70 percent of households shift by 25 or more percentile points relative to the desk formula, and over 86 percent shift by at least 10 percentile points. The near-uniform spread indicates that disagreement is not concentrated among marginal cases near the eligibility threshold; households throughout the distribution are substantially reordered.

Figure 3: Distribution of percentile displacement between the desk formula and each human group



Notes: Each panel shows the distribution of percentile displacement—the change in a household's percentile rank when moving from the desk formula's ranking to the indicated human group's ranking—across 5,070 prediction-sample households. Positive values indicate households ranked higher by the human group; negative values indicate the reverse. The dashed vertical line marks zero displacement. Distributions span nearly the full $[-100, +100]$ range, indicating pervasive reordering rather than disagreement concentrated near the eligibility threshold.

Characterizing the households in the disagreement zone—those selected by the desk formula but not by human groups—reveals a consistent pattern across cutoffs and groups. The desk

formula disproportionately selects large families (five or more children), households in Zahle and Nabatieh, and households with no medical conditions. Human groups, by contrast, would instead prioritize small families with elderly members, households headed by individuals with medical conditions or disabilities, and residents of Beirut. At the 50 percent cutoff, households with elderly members constitute 33 percent of the full sample but are nearly absent from the set of households that the desk formula selects and humans would not (Table 5). Conversely, large families with five children are overrepresented in this contested set by approximately 30 percentage points relative to the full sample. These patterns confirm that the divergence between the desk formula and human preferences is not driven by noise at the margin but by a fundamental difference in what each treats as the primary signal of vulnerability: the desk formula weights demographic predictors of consumption, while human respondents weight observable markers of acute need.

Table 5: Targeting disagreement: share of desk-formula-selected households not selected by each group

Budget cutoff	Group	Disagreement (n)	Disagreement (%)
25%	Field	1196	94.3%
25%	Central	1149	90.6%
25%	Beneficiary	1231	97.1%
50%	Field	1723	68.0%
50%	Central	1420	56.0%
50%	Beneficiary	1775	70.0%
75%	Field	1253	33.0%
75%	Central	1125	29.6%
75%	Beneficiary	1217	32.0%

Notes: At each budget cutoff, the table reports the count and share of households selected by the desk formula that would not be selected by each human group's AMCE-derived ranking of 5,070 prediction-sample households.

7 Implications and Discussion

Donors in foreign aid and humanitarian contexts primarily monitor for compliance with technical targeting models and formal eligibility rules, rather than for alignment with local perceptions of fairness (Easterly, 2006; Fox, 2007; Martinez-Bravo, 2017). Divergences between implementer and beneficiary preferences and formal targeting criteria thus offer indirect evidence about the degree of institutional autonomy and the balance institutions strike between donor compliance and local legitimacy.

Technical targeting methods that fail to capture salient vulnerability dimensions also risk

systematically excluding highly at-risk households. Both humanitarian staff and refugee outreach volunteers placed greater weight on attributes such as disability, serious medical conditions, and inadequate household assets, while the desk formula emphasized demographic correlates of expenditure. This divergence illustrates the limits of purely econometric targeting approaches in contexts where multidimensional vulnerability is salient. Refugees' perceptions of vulnerability are not independent of the targeting system itself. In a population that is almost homogeneously poor and heavily dependent on assistance, households systematically excluded by the formula may be perceived as more vulnerable precisely because they are excluded.

The analysis also speaks to broader patterns of institutional behavior under competing accountability pressures. The desk formula aligns most closely with central office staff—the organizational layer most directly exposed to donor oversight—while diverging sharply from field staff and beneficiary preferences. The IO literature offers three accounts of why this pattern might arise: effective donor monitoring transmitted through proximate principals, professional socialization in which central staff internalize econometric targeting as best practice (Barnett and Finnemore, 1999), and bureaucratic drift in which operational routines serve administrative efficiency and audit defensibility rather than either principal (Nielson and Tierney, 2003). The hierarchy clearly filters information directionally: upward accountability pressures are amplified, while field knowledge about local conceptions of vulnerability is attenuated, consistent with Nielson and Tierney (2003)'s prediction that pressure on IO agents operates through proximate principals in the delegation chain. Our data cannot distinguish whether this filtering reflects formal monitoring, socialization, or institutional incentives.

More broadly, our results illustrate how observed targeting outcomes can reveal the balance institutions strike between compliance and legitimacy when formal accountability mechanisms diverge—an approach that may generalize beyond humanitarian cash transfers to other settings where multiple principals exert competing pressures on implementing agents. That said, our findings derive from a single program in one country, and the relative strength of compliance and legitimacy pressures likely varies across institutional contexts, donor relationships, and displacement settings. Whether the patterns we document—particularly the proximity of implemented targeting to central-office preferences—hold in programs with different governance structures or weaker donor oversight remains an open empirical question.

8 Conclusion

This study recovers the distributional preferences of actors at distinct levels of a humanitarian organization's hierarchy and compares them to the formal targeting rule the organization implements. Using a conjoint experiment fielded to central office staff, field implementers, and refugee outreach volunteers in Lebanon, we find that the desk formula used for cash transfer eligibility diverges substantially from the preferences of all human groups—but lies closest to central office staff, the organizational layer most directly exposed to donor oversight. Staff and refugees weight severe health conditions, disability, and acute economic distress; the formula weights demographic predictors of consumption. When translated into counterfactual targeting decisions, these divergences imply that the vast majority of the formula's selections would be contested by every human group.

These findings speak to a central question in the study of international organizations: how do competing accountability pressures shape institutional behavior when the demands of external principals and local constituencies diverge? The pattern we document—upward accountability dominating downward accountability, with beneficiary preferences exerting little observable influence on implemented policy—is consistent with delegation models in which oversight pressure is transmitted through proximate principals in the organizational hierarchy. Whether this pattern reflects effective donor monitoring, professional socialization into technocratic norms, or bureaucratic preferences for auditable procedures remains an open question that our single-case design cannot resolve.

Funding Declaration: The survey instrument was designed and the data collected as part of paid consultancy work by the authors for a UN agency (name omitted for anonymity), undertaken as an internal program improvement exercise within the agency's operations. The authors have had no relationship with the agency since 2022. The analyses presented in this article were conducted independently after the completion of that work and without financial support.

Competing Interests Declaration: The agency had no role in the analysis of the data, in the interpretation of the results, in the conclusions reached in this study, nor in the preparation of this study or in the decision to submit the article for publication. The authors declare no competing interests.

Data Availability: TBA

Ethical Statement: This study is a secondary analysis of de-identified data originally collected as part of an internal program improvement exercise. The data contain no personally identifiable information. This study was reviewed by the Emory University Institutional Review Board (Protocol 2026P000536).

References

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, 1206–1240.

Altındağ, O., S. D. 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, 102564.

Barnett, M. and M. Finnemore (2004): *Rules for the World: International Organizations in Global Politics*, Cornell University Press.

Barnett, M. N. and M. Finnemore (1999): "The Politics, Power, and Pathologies of International Organizations," *International Organization*, 53, 699–732.

Bernheim, B. D. and M. D. Whinston (1986): "Common Agency," *Econometrica*, 54, 923–942.

Brutger, R. and R. Clark (2023): "At What Cost? Power, Payments, and Public Support of International Organizations," *The Review of International Organizations*, 18, 431–465.

Bénabou, R. and J. Tirole (2006): "Incentives and Prosocial Behavior," *American Economic Review*, 96, 1652–1678.

Carnegie, A., R. T. Clark, and N. Zucker (2024): "Global Governance under Populism: The Challenge of Information Suppression," *World Politics*, 76, 639–666.

Clark, R. T. (2021): "Pool or Duel? Cooperation and Competition among International Organizations," *International Organization*, 75, 1133–1153.

Coady, D., M. Grosh, and J. Hoddinott (2004): *Targeting of Transfers in Developing Countries: Review of Lessons and Experience*, World Bank Publications.

de Bekker-Grob, E. W., B. Donkers, M. F. Jonker, and E. A. Stolk (2015): "Sample Size Requirements for Discrete-Choice Experiments in Healthcare: A Practical Guide," *The Patient – Patient-Centered Outcomes Research*, 8, 373–384.

Deardorff Miller, S. (2017): *UNHCR as a Surrogate State: Protracted Refugee Situations*, Global Institutions, London: Routledge.

Devarajan, S., S. Khemani, and M. Walton (2011): "Civil Society, Public Action and Accountability in Africa," *Policy Research Working Paper Series*.

Dietrich, S. (2016): "Donor Political Economies and the Pursuit of Aid Effectiveness," *International Organization*, 70, 65–102.

Easterly, W. (2006): *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*, New York: Penguin Press.

Fiszbein, A. and N. Schady (2009): *Conditional Cash Transfers: Reducing Present and Future Poverty*, World Bank Publications.

Fox, J. (2007): *Accountability Politics: Power and Voice in Rural Mexico*, Oxford: Oxford University Press.

Gentilini, U. (2022): *Cash Transfers in Pandemic Times: Evidence, Practices, and Implications from the Largest Scale Up in History*, World Bank Publications.

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.

Harvey, P. (2005): "Cash and vouchers in emergencies," *Humanitarian Policy Group Report*, 24.

Hawkins, D. G., D. A. Lake, D. L. Nielson, and M. J. Tierney, eds. (2006): *Delegation and Agency in International Organizations*, Political Economy of Institutions and Decisions, Cambridge University Press.

Kagan, M. (2011): "We Live in a Country of UNHCR: The UN Surrogate State and Refugee Policy in the Middle East," Research Paper 201, UNHCR Policy Development and Evaluation Service.

Kosec, K. and C. H. Mo (2023): "Does relative deprivation condition the effects of social protection programs on political support? Experimental evidence from Pakistan," *American Journal of Political Science*, 68, 832–849.

Lancsar, E. and J. Louviere (2008): "Conducting Discrete Choice Experiments to Inform Healthcare Decision Making: A User's Guide," *PharmacoEconomics*, 26, 661–677.

Levi, M. (1997): *Consent, Dissent, and Patriotism*, Cambridge: Cambridge University Press.

Mansuri, G. and V. Rao (2013): *Localizing Development: Does Participation Work?*, World Bank Publications.

Martens, B., U. Mummert, P. Murrell, and P. Seabright (2002): *The Institutional Economics of Foreign Aid*, Cambridge: Cambridge University Press.

Martinez-Bravo, M. (2017): "The Role of Local Officials in New Democracies: Evidence from Indonesia," *American Economic Review*, 107, 3180–3210.

Nielson, D. L. and M. J. Tierney (2003): "Delegation to International Organizations: Agency Theory and World Bank Environmental Reform," *International Organization*, 57, 241–276.

North, D. C. (1990): *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.

Olken, B. A. (2010): "Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia," *American Political Science Review*, 104, 243–267.

Platteau, J.-P. (2004): "Monitoring Elite Capture in Community-Driven Development," *Development and Change*, 35, 223–246.

Pritchett, L., M. Woolcock, and M. Andrews (2013): "Looking Like a State: Techniques of Persistent Failure in State Capability for Implementation," *Journal of Development Studies*, 49, 1–18.

Stokes, S. C. (2005): "Perverse Accountability: A Formal Model of Machine Politics with Evidence from Argentina," *American Political Science Review*, 99, 315–325.

Tallberg, J., K. Bäckstrand, and J. A. Scholte, eds. (2018): *Legitimacy in Global Governance: Sources, Processes, and Consequences*, Oxford University Press.

UNHCR (2022): "Vulnerability Assessment of Syrian Refugees in Lebanon (VASyR)," Accessed from UNHCR reports.

A Formal Specification of the Conceptual Framework

This appendix collects the formal components of the conceptual framework introduced in Section 3.

Properties of u, R, and C

- $u(\cdot)$ is an increasing function representing the monetary equivalent of the divergence between I's ideal targeting strategy T_I and the implemented one T_t. $u(0) = \bar{U}$ and $u' > 0$, meaning that higher divergence from the ideal strategy lowers the monetary equivalent of the implemented strategy.
- $R(\cdot)$ is the revenue function denoting the funding that I receives from D. $R(\cdot)$ depends on how close the implemented targeting strategy is to D's ideal, with this proximity captured by d_D. We assume $R(0) = \bar{R}$ and $R'(\cdot) > 0$: under perfect compliance (when divergence between T_t and T_D is zero), I receives the maximum funding $\bar{R}$, and higher divergence reduces funding.
- $C(\cdot)$ represents the implementation costs of the targeting strategy and captures the legitimacy aspect of the problem. Lower legitimacy from the target public raises the cost of implementing a strategy for I. We assume $C(0) = 0$ and $C'(\cdot) < 0$: if I's targeting strategy (T_t) coincides with B's ideal (T_B), implementation is costless in legitimacy terms, and higher divergence raises costs.

Representing Targeting Strategies

For a given set $\mathcal{S}$ of potential beneficiaries consisting of N elements, we can predict the ideal targeting strategies of the three agents (and also the implemented one, T_t) using the estimated values of AMCEs from the experimental data. By observing the relative divergence of $\widehat{T}_t$ from $\widehat{T}_D$, $\widehat{T}_B$, and $\widehat{T}_I$, we can infer the underlying structure of $R(\cdot)$ and $C(\cdot)$.

Targeting strategies can be represented either as an $N \times 1$ vector of selection indicators ($T_j \in \{0,1\}^N$) or as a ranked ordering of beneficiaries, with the cutoff determined by the budget constraint.

Divergence Measures: Formal Definition

The choice of divergence measure is central to the empirical application of the framework. Since I is the optimizing agent, T_t will be located on the same plane as T_I, T_D, and T_B, meaning that the realized targeting strategy minimizes divergence from T_j (for all j) conditional on the divergence from T_k and T_l (for all $k, l \neq j$), where $j, k, l \in \{D, B, I\}$.

Appendix Table 1: Average Marginal Component Effects (AMCEs) by Sample

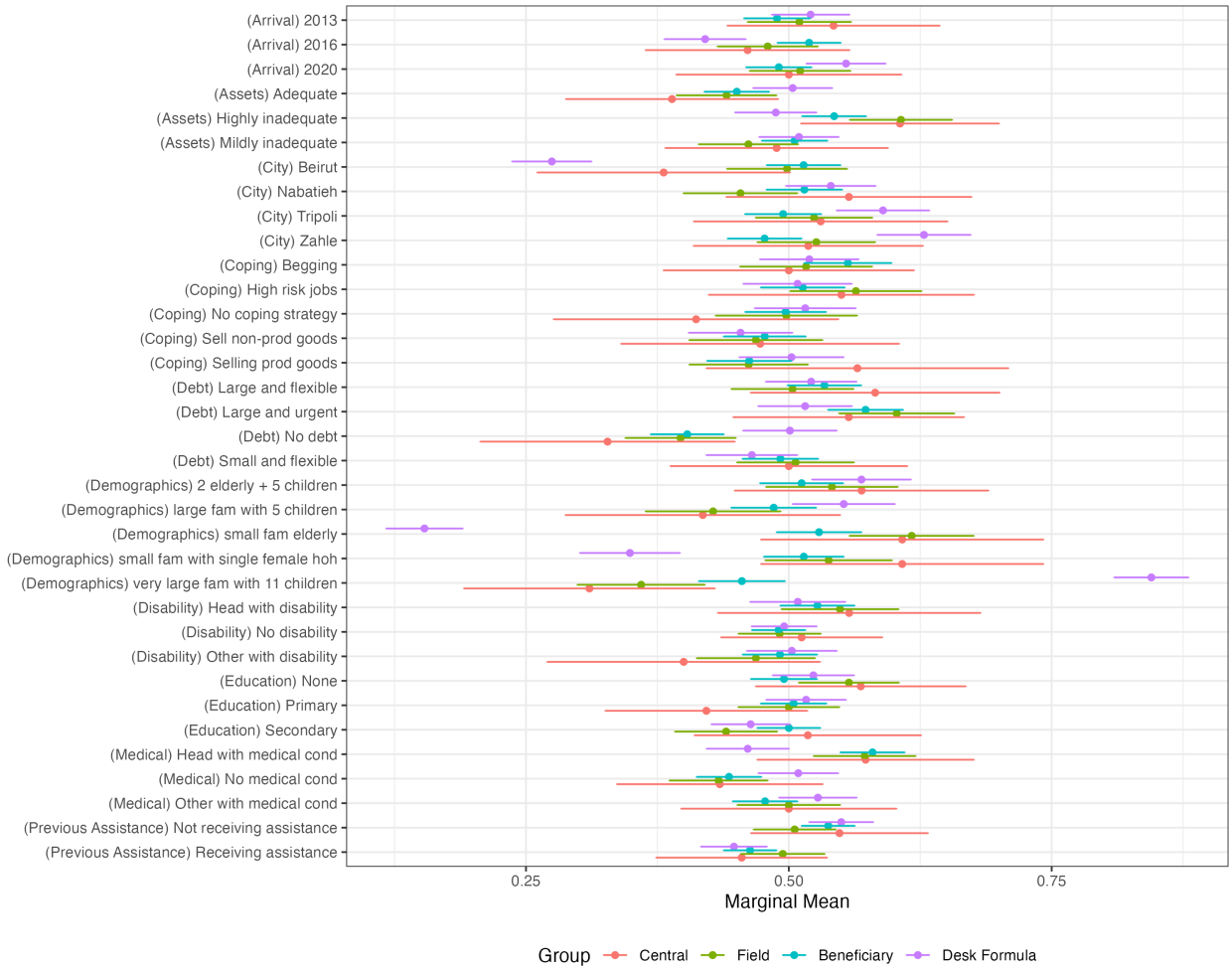
Attribute level	Central office staff	Field staff	Refugee respondents	Targeting model (PMT)
Arrival: 2016	-0.102	-0.031	0.026	-0.094
	(0.072)	(0.037)	(0.022)	(0.023)
Arrival: 2020	-0.054	0.007	0.003	0.031
	(0.085)	(0.036)	(0.022)	(0.023)
Assets: Mildly inadequate	0.115	0.030	0.052	0.019
	(0.054)	(0.034)	(0.022)	(0.023)
Assets: Highly inadequate	0.259	0.156	0.089	-0.015
	(0.075)	(0.032)	(0.022)	(0.023)
Nabatieh	0.191	-0.035	0.008	0.266
	(0.082)	(0.038)	(0.025)	(0.025)
Tripoli	0.164	0.019	-0.021	0.326
	(0.075)	(0.038)	(0.025)	(0.026)
Zahle	0.146	0.026	-0.030	0.392
	(0.070)	(0.040)	(0.024)	(0.026)
Coping: Begging	0.102	0.041	0.059	0.018
	(0.086)	(0.051)	(0.028)	(0.029)
Coping: High risk jobs	0.129	0.069	0.013	-0.013
	(0.095)	(0.045)	(0.028)	(0.029)
Coping: Sell nonprod. goods	0.060	-0.010	-0.017	-0.025
	(0.088)	(0.048)	(0.028)	(0.028)
Coping: Selling prod. goods	0.115	-0.030	-0.030	-0.003
	(0.102)	(0.044)	(0.028)	(0.030)
Debt: Small and flexible	0.200	0.098	0.088	-0.028
	(0.066)	(0.041)	(0.027)	(0.026)
Debt: Large and flexible	0.296	0.114	0.132	0.008
	(0.071)	(0.037)	(0.026)	(0.026)
Debt: Large and urgent	0.255	0.190	0.172	-0.013
	(0.074)	(0.037)	(0.026)	(0.027)
Demographics: large fam. with 5 children	-0.192	-0.097	-0.035	-0.019
	(0.079)	(0.041)	(0.027)	(0.032)
Demographics: small fam. with elderly	0.015	0.078	0.014	-0.441
	(0.101)	(0.039)	(0.029)	(0.028)
Demographics: small fam. with single female HoH	0.033	0.015	-0.002	-0.222
	(0.095)	(0.041)	(0.028)	(0.032)
Demographics: very large fam. with 11 children	-0.246	-0.171	-0.066	0.269
	(0.101)	(0.042)	(0.030)	(0.027)
Disability: Head with disability	0.030	0.073	0.038	-0.013
	(0.067)	(0.034)	(0.023)	(0.023)
Disability: Other with disability	-0.074	-0.007	0.003	-0.016
	(0.073)	(0.035)	(0.022)	(0.023)
Education: Primary	-0.141	-0.051	0.008	-0.035
	(0.076)	(0.033)	(0.022)	(0.023)
Education: Secondary	-0.066	-0.105	0.009	-0.097
	(0.083)	(0.032)	(0.021)	(0.023)
Medical: Head with medical cond	0.119	0.133	0.133	-0.037
	(0.084)	(0.034)	(0.022)	(0.023)
Medical: Other with medical cond	0.078	0.054	0.040	-0.011
	(0.070)	(0.034)	(0.022)	(0.023)
Previous Assistance	0.096	0.005	0.069	0.113
	(0.051)	(0.029)	(0.018)	(0.019)

Appendix Table 2: Mahalanobis distance and Wald test results for all pairwise comparisons

Pair 1	Pair 2	D_1^2	D_2^2	Difference	W (Wald χ_1^2)	p-value
Central - Benef	Benef - Model	101.800***	816.793***	-714.993	144.459	$p < 0.001$
Central - Benef	Central - Model	101.800***	327.633***	-225.832	40.615	$p < 0.001$
Central - Benef	Field - Benef	101.800***	55.237***	46.563	4.110	$p = 0.043$
Central - Benef	Field - Model	101.800***	650.539***	-548.738	100.060	$p < 0.001$
Central - Field	Benef - Model	46.496***	816.793***	-770.297	171.833	$p < 0.001$
Central - Field	Central - Benef	46.496***	101.800***	-55.304	9.511	$p = 0.002$
Central - Field	Central - Model	46.496***	327.633***	-281.137	62.321	$p < 0.001$
Central - Field	Field - Benef	46.496***	55.237***	-8.741	0.177	$p = 0.674$
Central - Field	Field - Model	46.496***	650.539***	-604.043	134.838	$p < 0.001$
Central - Model	Benef - Model	327.633***	816.793***	-489.161	61.404	$p < 0.001$
Central - Model	Field - Benef	327.633***	55.237***	272.396	48.450	$p < 0.001$
Central - Model	Field - Model	327.633***	650.539***	-322.906	31.342	$p < 0.001$
Field - Benef	Benef - Model	55.237***	816.793***	-761.557	154.356	$p < 0.001$
Field - Benef	Field - Model	55.237***	650.539***	-595.302	152.532	$p < 0.001$
Field - Model	Benef - Model	650.539***	816.793***	-166.255	6.261	$p = 0.012$

Note: The table reports Mahalanobis distance (D^2) and Wald test results for all pairwise comparisons among targeting preference profiles. Each D_i^2 measures the multivariate divergence between two actor preference vectors across all conjoint attributes. The Wald statistic (W) tests whether the difference between the two distances is statistically significant. Stars denote significance based on a χ_{25}^2 reference: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The Wald test follows a χ_1^2 distribution. *Reading:* Each row compares the divergence patterns between two actor pairs. For example, in the first row, the distance between *Central–Beneficiary* preferences ($D^2 = 101.8$) is contrasted with that between *Beneficiary–Model* preferences ($D^2 = 816.8$). The large, significant difference ($W = 144.5$, $p < 0.001$) indicates that beneficiary preferences diverge more strongly from the model than from the central office, consistent with a donor-facing compliance constraint.

Appendix Figure 1: Marginal Means Across Groups



Detailed Rank Divergence Matrices

Tables 3–6 report pairwise rank divergence measures with bootstrap 95% confidence intervals (1,000 replications) for each measure individually.

Appendix Table 3: Kendall τ distance (bootstrap 95% CI)

Group	Beneficiary	Field	Central	Desk Formula
Beneficiary	—	0.274 [0.227, 0.453]	0.315 [0.254, 0.495]	0.699 [0.539, 0.759]
Field	0.274 [0.227, 0.453]	—	0.261 [0.204, 0.429]	0.649 [0.541, 0.720]
Central	0.315 [0.254, 0.495]	0.261 [0.204, 0.429]	—	0.556 [0.425, 0.640]
Desk Formula	0.699 [0.539, 0.759]	0.649 [0.541, 0.720]	0.556 [0.425, 0.640]	—

Appendix Table 4: Spearman ρ distance (bootstrap 95% CI)

Group	Beneficiary	Field	Central	Desk Formula
Beneficiary	—	0.178 [0.128, 0.433]	0.235 [0.159, 0.498]	0.786 [0.561, 0.861]
Field	0.178 [0.128, 0.433]	—	0.165 [0.107, 0.398]	0.740 [0.573, 0.824]
Central	0.235 [0.159, 0.498]	0.165 [0.107, 0.398]	—	0.587 [0.389, 0.720]
Desk Formula	0.786 [0.561, 0.861]	0.740 [0.573, 0.824]	0.587 [0.389, 0.720]	—

Appendix Table 5: Spearman footrule distance (bootstrap 95% CI)

Group	Beneficiary	Field	Central	Desk Formula
Beneficiary	—	0.393 [0.320, 0.621]	0.434 [0.354, 0.668]	0.873 [0.709, 0.929]
Field	0.393 [0.320, 0.621]	—	0.366 [0.291, 0.593]	0.868 [0.739, 0.917]
Central	0.434 [0.354, 0.668]	0.366 [0.291, 0.593]	—	0.776 [0.593, 0.862]
Desk Formula	0.873 [0.709, 0.929]	0.868 [0.739, 0.917]	0.776 [0.593, 0.862]	—

Appendix Table 6: Top-half mismatch rate (bootstrap 95% CI)

Group	Beneficiary	Field	Central	Desk Formula
Beneficiary	—	0.234 [0.188, 0.471]	0.270 [0.225, 0.508]	0.700 [0.518, 0.787]
Field	0.234 [0.188, 0.471]	—	0.248 [0.183, 0.431]	0.680 [0.548, 0.756]
Central	0.270 [0.225, 0.508]	0.248 [0.183, 0.431]	—	0.560 [0.400, 0.677]
Desk Formula	0.700 [0.518, 0.787]	0.680 [0.548, 0.756]	0.560 [0.400, 0.677]	—

AMCE Vector Divergence Tests Using Mahalanobis Distance

As a complement to the rank-based divergence measures reported in the main text, this appendix presents tests based on comparing the multivariate vectors of estimated AMCEs

across groups. While rank-based measures capture disagreement in terms of actual targeting decisions—which households are prioritized and in what order—the Mahalanobis distance approach summarizes divergence in the space of estimated attribute weights, accounting for estimation uncertainty and covariance.

Methodology

We compute the squared Mahalanobis distance between each pair of estimated AMCE vectors:

$$D_{ij}^2 = (\boldsymbol{\beta}_i - \boldsymbol{\beta}_j)' \boldsymbol{\Sigma}_{ij}^{-1} (\boldsymbol{\beta}_i - \boldsymbol{\beta}_j),$$

where $\boldsymbol{\beta}_i$ and $\boldsymbol{\beta}_j$ denote the AMCE vectors for groups i and j, and $\boldsymbol{\Sigma}_{ij}$ is the estimated covariance matrix of their difference, constructed from the joint block-diagonal covariance of all four models. This metric rescales the raw Euclidean distance by the precision and correlation structure of the estimates. Values near zero imply statistically indistinguishable AMCEs; larger values indicate greater divergence.

To assess whether two divergences are statistically distinguishable, we employ a Wald test on the difference of squared Mahalanobis distances:

$$H_0 : \; D_{ab}^2 = D_{cd}^2, \qquad W = \frac{h(\hat{\boldsymbol{\theta}})^2}{\nabla h(\hat{\boldsymbol{\theta}})' \widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}}) \nabla h(\hat{\boldsymbol{\theta}})} \overset{H_0}{\sim} \chi_1^2.$$

Results

Appendix Table 7 presents the Wald tests. Both tests under Expectation 1 (donor-facing compliance) are significant and in the expected direction: the divergence between the desk formula and field preferences exceeds that between the desk formula and central staff ($W = 23.615$, $p < 0.001$), and the divergence between the desk formula and beneficiaries exceeds that between the desk formula and central staff ($W = 59.058$, $p < 0.001$). Expectation 2 (local legitimacy) and Expectation 4 (internal compromise) are not supported. Expectation 3 (field-level pragmatism) yields mixed evidence. Expectation 6 (no meaningful trade-off) is rejected. Overall, the AMCE-based divergence tests indicate that the desk formula lies nearest to central-office preferences, consistent with donor-facing compliance pressures dominating legitimacy considerations. These conclusions are broadly consistent with the rank-based results reported in the main text.

Appendix Table 7: Mahalanobis Divergence Tests Corresponding to Each Hypothesis (with expected directional signs)

Expectations		**Expected Pattern of Divergence**	**Test Statistic / p-value**
1.	Legitimacy constraint binding	$D^2_{\text{Central,Formula}} > D^2_{\text{Beneficiary,Formula}}$	*sign reversed,* $W = 61.404$; $p < 0.001$
		$D^2_{\text{Field,Formula}} > D^2_{\text{Beneficiary,Formula}}$	*sign reversed,* $W = 6.261$; $p = 0.012$
2.	Donor-facing compliance constraint dominates	$D^2_{\text{Field,Formula}} > D^2_{\text{Central,Formula}}$	$W = 31.342$; $p < 0.001$
		$D^2_{\text{Beneficiary,Formula}} > D^2_{\text{Central,Formula}}$	$W = 61.404$; $p < 0.001$
3.	Operational or field-level pragmatism dominates	$D^2_{\text{Central,Formula}} > D^2_{\text{Field,Formula}}$	*sign reversed,* $W = 31.342$; $p < 0.001$
		$D^2_{\text{Beneficiary,Formula}} > D^2_{\text{Field,Formula}}$	$W = 6.261$; $p = 0.012$
4.	Internal compromise between donor-facing and field preferences	$D^2_{\text{Central,Formula}} \approx D^2_{\text{Field,Formula}}$	*sign reversed,* $W = 31.342$; $p < 0.001$
		$D^2_{\text{Central,Field}} > D^2_{\text{Central,Formula}}$	*sign reversed,* $W = 62.321$; $p < 0.001$
5.	Possible unobserved donor constraint:	None of the above, and $(T_D \notin \{T_C, T_F, T_B\})$	**Expectation 1** is satisfied
6.	No meaningful accountability trade-off	All pairwise D^2_{ij} small and not significantly different	All D^2_{ij} values reject see Appendix Table TKTK Column TKTK

Note: The table summarizes results from Mahalanobis divergence tests used to evaluate the framework's six hypothesized accountability regimes. Each hypothesis specifies an expected directional pattern in the squared Mahalanobis distances (D^2_{ij}) between estimated targeting preference vectors for different actors—beneficiaries (B), field staff (F), central management (C), and the desk formula (DF). The test statistic (W) and associated p-value indicate whether the observed divergence pattern is consistent with each hypothesized regime. Sign reversals denote cases where the observed direction of divergence was opposite the hypothesized one.

Reading: Each row presents a test of whether the observed distance between two preference profiles supports or contradicts a specific accountability hypothesis. For example, under **H2**, donor-facing compliance is expected to dominate if the divergence between field staff and the desk formula ($D^2_{\text{Field,Formula}}$) exceeds that between central management and the desk formula ($D^2_{\text{Central,Formula}}$). A significant positive test statistic for this comparison indicates evidence consistent with a donor-driven compliance constraint.

Alternative Model: Dynamic Institutional Survival under Continuous Accountability Constraints

As an alternative to the static common agency framework presented in the main text, this appendix presents a dynamic model in which the institution's targeting and scale decisions evolve over time. This formulation endogenizes scale and funding as functions of past legitimacy and compliance, capturing the path-dependent nature of institutional survival.

Players and Sequence of Actions

Three actors interact in each period t:

- The **institution** (I) chooses a targeting strategy T_t and a scale Scale_t, seeking to sustain operations and funding.
- The **local beneficiaries** (B) form perceptions of legitimacy based on the institution's alignment with local expectations.
- The **donor** (D) monitors adherence to its *own preferred targeting strategy* and allocates funding based on observed technical compliance.

The institution's strategy T_t can vary continuously between two benchmarks:

- T_B: Full alignment with local preferences,
- T_D: Full alignment with donor technical requirements.

The sequence of actions in each period t is as follows:

1. The institution receives funding F_t. Initial conditions are given by $\text{Scale}_0 > 0, \quad F_0 = b \times \text{Scale}_0$.
2. It chooses:
   - A targeting strategy T_t,
   - A scale of operations Scale_t.
3. Legitimacy and compliance are realized based on T_t:

$$\text{Legitimacy}_t = 1 - \gamma|T_t - T_B|, \quad \text{Compliance}_t = 1 - \theta|T_t - T_D|$$

   where $\gamma > 0$ controls how sharply legitimacy declines as the institution deviates from local expectations, and $\theta > 0$ controls how sharply compliance declines as the institution deviates from donor requirements.

   The institution incurs two types of immediate costs:

- **Operational costs**, subject to economies of scale:

$$\text{Operational Cost}_t = k_{\text{op}}(\text{Scale}_t), \quad \text{with } k'_{\text{op}} > 0,\ k''_{\text{op}} < 0$$

- **Legitimacy-related costs**, proportional to the size of operations:

$$\text{Legitimacy Cost}_t = k_{\text{leg}} \times (1 - \text{Legitimacy}_t) \times \text{Scale}_t$$

4. Within-period surplus:

$$\sigma_t = F_t - \text{Operational Cost}_t - \text{Legitimacy Cost}_t$$

5. The institution proposes a next-period scale:

$$\text{Scale}'_{t+1} = \text{Scale}_t \times (1 + \sigma_t)$$

6. Donor funding for the next period:

$$F_{t+1} = \text{Scale}'_{t+1} \times [b + \beta\,(\text{Compliance}_t - C_{\text{base}})]$$

where $b > 0$ is baseline funding per unit scale, $\beta > 0$ is donor sensitivity to compliance, and C_{base} is the donor's minimum expected compliance threshold.

Equilibrium Concept

An equilibrium strategy T_t^* satisfies: continuous survival (legitimacy and compliance remain within $[0, 1]$), adaptive adjustment of T_t and Scale_t to maintain viability over time, and tolerance for fluctuations in surplus as long as collapse is avoided.

Comparative Statics

Key predictions: higher γ encourages alignment with T_B due to rising legitimacy costs; higher θ encourages alignment with T_D due to stronger compliance incentives; larger Scale_t magnifies absolute legitimacy costs; and greater scale enables drift toward T_D as operational cost per unit falls while legitimacy cost per unit remains constant. The marginal incentive ratio:

$$\frac{dF_{t+1}/dT_t}{d(\text{Legitimacy Cost}_t)/dT_t} = \frac{-\text{Scale}'_{t+1} \times \beta\theta \times \text{sign}(T_t - T_D)}{k_{\text{leg}} \times \gamma \times \text{Scale}_t \times \text{sign}(T_t - T_B)}$$

captures the institution's incentive to reallocate targeting toward donor compliance or local preferences.

Surplus Maximization Extension

As a further extension, the institution may seek to maximize surplus subject to a hard legitimacy constraint $\text{Legitimacy}_t \geq \bar{L}$. The first-order condition at interior optima yields $dF_t/dT_t = d(\text{Legitimacy Cost}_t)/dT_t$. When the constraint binds, the institution chooses the targeting strategy achieving exactly $\text{Legitimacy}_t = \bar{L}$ and accepts lower surplus. This formalizes how growth-oriented institutions may be constrained by reputational floor effects.

Focus Groups

From June to September 2020, we conducted 9 focus groups with the prospective participants of the survey. All focus groups were conducted online due to COVID19 restrictions.

1. **Focus Group with the Outreach Volunteers (OVs):** In September 2020, we held a focus group discussion with agency outreach volunteers. 10 volunteers from North, BML, Bekaa, and South Lebanon participated in the meeting.

2. **Focus Group with UN agency field staff:** From July until mid-September, we conducted 4 focus groups with UN agency field staff working at various thematic focal points (e.g. basic assistance, protection, and registration) across different regions in Lebanon. 16 field staff in total participated in the focus groups.

3. **Focus Groups with UN agency central offices:** From July until mid-September, we conducted 4 focus groups with central office staff at the agencies' Beirut offices. 20 people in total participated in the meetings.