

EVALUATION REPORT:

Thematic Evaluation of the Joint
UNDP–UNHCR Area-Based Return
Support (ABRS) Pilot Project in Bab
Al-Nairab, Aleppo

*A Demonstrative Case of Localized
Human Security Solutions for
Displacement-Affected
Communities in Syria*

28 August 2026

Submitted for UNDP Syria
and UNHCR Syria by
RMTeam International

Table of Contents

Table of Figures	2
ABBREVIATIONS AND ACRONYMS	3
1. EXECUTIVE SUMMARY	5
1.1 Overview	5
1.2 Key Findings by Criterion	6
1.3 Main Conclusions	8
1.4 Priority Recommendations	8
2. INTRODUCTION	10
2.1 Background and Context	10
2.2 Evaluation Purpose and Objectives	11
2.3 Scope and Limitations	11
3. PROGRAM DESCRIPTION	12
3.1 Reconstructed Theory of Change	12
3.2 Pillars and Intervention Logic	13
3.3 Stakeholder Ecosystem	13
4. METHODOLOGY	14
4.1 Design	14
4.2 Sampling and Data Collection: Planned versus Achieved	14
4.3 Analysis Approach	17
4.4 Limitations and Mitigation	18
4.5 Ethical Considerations	19
5. FINDINGS	20
5.1 Relevance	20
5.2 Coherence	22
5.3 Effectiveness	24
5.4 Efficiency	32
5.5 Coordination and Partnerships	34
5.6 Methodological Soundness	35
5.7 Sustainability and Scale-Up Potential	37

5.8 Impact	39
5.9 Cross-Cutting Issues.....	41
6. CONCLUSIONS.....	43
6.1 Strategic Interpretation.....	43
6.2 Positioning ABRS as a Model	44
6.3 Critical Success Factors and Constraints.....	44
7. RECOMMENDATIONS.....	46
7.1 Strategic Recommendations	46
7.2 Operational Recommendations.....	46
7.3 Programming Recommendations (including scale-up implications)	47
8. LESSONS LEARNED.....	49

Table of Figures

Figure 1: Survey Respondent Profile (sub-area, population group, gender, and age).....	16
Figure 2: Beneficiary Exposure by Pillar	17
Figure 3: Top Reported Current Challenges.....	21
Figure 4: Themes in open-ended suggestions for future programming (n=204).....	21
Figure 5: Perceived Change in Core Outcomes - Access to Services, Safety & Dignity, Social Cohesion	25
Figure 6: Beneficiary Satisfaction by Pillar (0-10) - Legal Assistance, LPP, QIP	26
Figure 7: Geographic Variation in QIP Satisfaction and Feedback Awareness.....	27
Figure 8: Outcome Gaps by Population Group - Returnees vs. Host Community vs. IDPs	28
Figure 9 Perceived Representation in LPP Activities (C3, n=62).....	29
Figure 10: Quick Impact Project (QIP) Type Reported by Beneficiaries (n=110).....	30
Figure 11: Condition and Usage of QIP-Linked Infrastructure (n=110).....	32
Figure 12: Awareness of Feedback/Complaints Mechanism, by Gender	42

ABBREVIATIONS AND ACRONYMS

Acronym	Full Term
ABC	Area-Based Coordination
ABRS	Area-Based Return Support
AHCT	Area Humanitarian Country Team
CSO	Civil Society Organization
DAC	Development Assistance Committee (OECD)
FGD	Focus Group Discussion
GBV	Gender-Based Violence
HDP	Humanitarian–Development–Peace (Nexus)
HLP	Housing, Land and Property
IDP	Internally Displaced Person
JSB	Japan Supplementary Budget
KII	Key Informant Interview
LPP	Local Participatory Platform
MEAL	Monitoring, Evaluation, Accountability and Learning
MSC	Most Significant Change
NCE	No-Cost Extension

Acronym	Full Term
OECD	Organisation for Economic Co-operation and Development
PSEA	Prevention from Sexual Exploitation and Abuse
QIP	Quick Impact Project
SARC	Syrian Arab Red Crescent
ToC	Theory of Change
ToR	Terms of Reference
UNDP	United Nations Development Program
UNEG	United Nations Evaluation Group
UNHCR	United Nations High Commissioner for Refugees
UNTFHS	United Nations Trust Fund for Human Security
VTC	Vocational Training Centre

1. EXECUTIVE SUMMARY

1.1 Overview

The Area-Based Return Support (ABRS) pilot in Bab Al-Nairab, Aleppo, is a joint UNDP–UNHCR initiative (February 2024–October 2025, with implementation extended via No-Cost Extension) funded by the UN Trust Fund for Human Security and the Japan Supplementary Budget. Conceived as a demonstrative case of localized, human-security-centred programming within the Humanitarian–Development–Peace (HDP) Nexus and the Durable Solutions Framework, the pilot combined three mutually reinforcing pillars: (i) housing, land and property (HLP) legal assistance; (ii) Local Participatory Platforms (LPPs) for community-driven recovery planning; and (iii) Quick Impact Projects (QIPs) rehabilitating basic infrastructure and services. It targeted over 4,400 direct beneficiaries – returnees, internally displaced persons (IDPs) and host-community members – in a densely populated, conflict- and earthquake-affected district of Aleppo City.

This evaluation applies a tailored set of OECD-DAC criteria (relevance, coherence, effectiveness, efficiency, sustainability, impact) plus additional criteria of coordination and partnerships and methodological soundness, viewed through a human security lens (safety, dignity, access to rights and services, social cohesion). Consistent with the Inception Report, the evaluation applies **contribution analysis, not attribution**: in the absence of a counterfactual or baseline, findings describe the pilot's plausible contribution to observed changes alongside other contextual factors, rather than claiming the pilot caused them.

Evidence was triangulated across a household survey (n=204), approximately 30 key informant interviews across three distinct tracks – 15 general/field-level informants in Bab Al-Nairab, 9 national- and coordination-level UNDP/UNHCR and UN-agency staff, and 6 senior/strategic-level informants (UNDP/UNHCR headquarters, a donor representative, and the UN Global Solutions Hub), nine focus group discussions (of ten planned), two Most Significant Change narratives (of three to five planned), three structured site-observation checklists, and an independent desk review of approximately 180 program documents. Where data streams fell short of the Inception Report's targets (FGDs, MSC, and the gender balance of the household survey sample), this is flagged as a limitation affecting the strength of specific findings, not treated as a gap the report papers over.

Confidence ratings: Throughout this report, each finding carries a High, Medium-High, Medium or Low confidence rating reflecting the strength and independence of the evidence behind it, not the size or importance of the finding itself: High reflects convergent evidence from at least three independent source types; Medium-High reflects two independent source types, or one especially strong, well-corroborated source; Medium reflects a single source or evidence with a noted limitation; and Low reflects little or no usable evidence despite the evaluation actively probing for it. The full criteria are set out in Section 4.3.

1.2 Key Findings by Criterion

Relevance is strongly evidenced and convergent across every data source: sanitation, electricity, HLP/civil documentation, and livelihoods consistently emerge as the area's priority needs, and the pilot's three pillars map closely into them. The evaluation finds the ABRS model's core value proposition - an integrated, area-based response to a set of interlocking human-security deficits rather than a single-sector intervention - to be well founded. A specific relevance gap is that housing/shelter repair, repeatedly named by beneficiaries and implementers alike as a priority need, sits outside the pilot's scope by design.

Coherence with the wider UN architecture in Syria is strong and has strengthened over the pilot's life: ABRS has been adopted by the Humanitarian and UN Country Teams as a national framework for area-based, localized solutions and is being replicated in eight further governorates. At the operational level, however, senior informants across multiple institutional vantage points independently cautioned against overstating how "integrated" delivery actually was in practice: UNDP and UNHCR retained separate systems, mandates and business models, with the substantive achievement being alignment of priorities rather than merged delivery.

Effectiveness is mixed and shows a marked contrast between output achievement and outcome depth. Reported outputs are substantial (3,998 people reached through legal assistance (plus 1,500+ remote consultations); approximately 1 km of sewage network rehabilitated; 81 LPP members; a rehabilitated vocational training centre), and the household survey shows broadly positive perceived outcomes (67% report improved access to services, 72% report improved safety and dignity, 62% report improved inter-group relations). However, these gains are markedly uneven: survey and qualitative data converge on coverage variation (perceptions of the pilot's performance differ sharply by street/sub-area rather than by demographic group) as the single most consequential effectiveness finding, and returnees - a core target group - report comparatively less positive outcomes than host-community respondents on nearly every survey indicator, a finding the evaluation treats as requiring explicit management attention rather than an incidental subgroup difference.

Efficiency findings are genuinely mixed rather than uniformly positive. The sewage rehabilitation works were delivered on time and within budget. However, the evaluation's field observations identified workmanship and safety concerns at both sanitation sites, with distinct issues at each site, as well as at the training centre. Two injuries were also reported at one sanitation site in connection with excavation pits during the works. The implementing team indicated that some of these issues may relate to infrastructure outside the project's rehabilitated segments.

At the same time, a government report confirmed that no incidents occurred in the segments rehabilitated by UNDP and that the works were completed to a satisfactory standard. It also noted that the issues identified at the UNHCR-led vocational training centre were addressed within the warranty period. Taken together, the evidence points less to an absence of quality assurance than to a gap in contractor progress-reporting and transparency. Existing quality-assurance and warranty mechanisms appear to have provided channels for addressing the issues identified.

Coordination and partnerships are the criteria with the widest spread of assessment across respondents - from strongly positive to candidly mixed - and this divergence is itself a finding: informants closest to day-to-day delivery describe specific frictions (differing UNDP/UNHCR business models, staff rotation, HQ-field splits) that more distant, strategic-level informants do not see or do not raise.

Methodological soundness is a genuine strength of the ABRS model's community-engagement pathway, demonstrated through a traceable Problem Tree, Solutions Matrix, and validated portfolio methodology. This is considered a genuine strength, rather than simply a well-documented process, because the evidence chain is independently traceable end-to-end, from 114 logged community-identified issues to a validated and resourced recovery portfolio, with authentic, rather than managed, community ownership corroborated across multiple data sources (Section 5.6).

The evaluation nonetheless finds a structural gap between an ambitious human-security Theory of Change and an M&E system built almost entirely around output and reach indicators, which cannot on their own demonstrate the higher-order changes the pilot seeks to contribute to. This gap is best understood in light of the pilot's limited scale and 12–18-month duration, which made a focus on output and reach indicators appropriate at this stage rather than indicative of a design flaw. Demonstrating the higher-order human-security outcomes the model ultimately aims to contribute to will instead require a harmonized, outcome-level MEAL framework applied across a portfolio of ABRS sites, rather than relying on the indicators of any single pilot.

Sustainability and scale-up prospects are strong at the strategic/policy level (national adoption, replication in eight governorates) but genuinely uncertain at the Bab Al-Nairab-specific level: the local implementer's own final report records that no government counterpart team had been appointed to carry the 90-project community recovery plan forward, though coordination with Government counterparts on the plan has continued since project close, with the Government's International Cooperation Directorate increasingly adopting area-based, community-prioritized approaches, and multiple senior informants converge independently on short, fragmented, single-year funding cycles and Syria's highly centralised, still-forming governance architecture as structural constraints on scale-up that are not primarily about ABRS's own design.

Impact, assessed through a contribution lens, shows credible but modest and unevenly distributed early contributions to safety, dignity and social cohesion, concentrated where interventions were actually completed. The evaluation's single most analytically important impact-adjacent observation, offered independently by a senior informant, is that population labels (IDP/returnee/host) "disappear" in practice once people are integrating into shared daily life - a finding that speaks to the pilot's underlying human-security logic more directly than any output count.

Cross-cutting issues show genuine substantive engagement on protection and conflict sensitivity, but uneven results on gender (LPP representation of women, at approximately 51%, fell short of the proposal's 65% target; the evaluation's own household survey sample is 79% male against an Inception Report target of 50–60% women) and thin results on disability inclusion (representation figures exist, but almost no qualitative evidence on lived experience or accessibility was captured across the KII tracks). Set against these gaps, a legal-assistance caseload that is 61% female and markedly higher

feedback-mechanism awareness among women than men (77% vs. 47%) point, per Section 5.9, to meaningful attention to gender equality within the program.

1.3 Main Conclusions

The ABRS pilot should be understood, and evaluated, as a **proof of concept for an area-based, human-security-centred model of return support**, not as a comprehensive recovery response to Bab Al-Nairab's needs - a distinction the pilot's own designers, implementers and senior stakeholders converge on independently and explicitly. Judged against that more accurate benchmark, the pilot performed credibly: it demonstrated that legal assistance, participatory governance and infrastructure rehabilitation can be delivered as a single, area-based package; it generated a genuinely strong and traceable community-engagement methodology; and it achieved meaningful strategic traction, becoming the basis for a national framework now scaling to eight governorates. Its most significant constraints are not primarily about whether the model is sound, but about the gap between a human-security ambition and (i) a monitoring system that measures only outputs, (ii) uneven, geographically and demographically patchy delivery within the pilot area itself, (iii) contractor progress-reporting transparency gap, for which existing quality-assurance and warranty channels appear to have provided a means of resolution, and (iv) a funding and still-forming governance environment in Syria that constrains sustainability and scale-up regardless of the model's own merits, notwithstanding continued Government coordination on the Community Recovery Plan since project close.

1.4 Priority Recommendations

1. **(Strategic, UNDP/UNHCR Country Offices and Solutions Working Group)** Before further scale-up, invest in a harmonized outcome-level MEAL framework for ABRS as a national framework - not only Bab Al-Nairab - that measures human-security and durable-solutions change directly, rather than relying on output/reach counts alone.
2. **(Strategic, UNDP/UNHCR with donors)** Pursue multi-year, pooled or blended financing architecture for ABRS sites; the evaluation's own evidence base and multiple independent senior informants converge on 12-month, single-donor funding cycles as a structural constraint that no amount of implementation quality can fully offset.
3. **(Operational, UNDP/UNHCR Aleppo field office)** Reinforce existing warranty and institutional-ownership arrangements for rehabilitated infrastructure - ensuring local authorities and community committees are properly equipped, briefed and resourced to assume operations-and-maintenance responsibility at handover - before further QIPs are implemented.
4. **(Operational, implementing team and local authorities)** Formally confirm a government/local-authority counterpart team to carry the validated 90-project community recovery plan forward; this was explicitly flagged as unresolved in the implementer's own final report and is a precondition for the participatory-planning pillar's sustainability claim.
5. **(Programming, next-phase design)** Extend future ABRS-model programming to cover housing/shelter repair - repeatedly identified as a priority need outside current scope - and

strengthen the accountability/feedback-mechanism function, given that only 53% of survey respondents knew where to raise a complaint.

6. **(Programming, cross-cutting)** Strengthen and better mainstream gender- and disability-inclusion data collection in future phases.

2. INTRODUCTION

2.1 Background and Context

Syria remains one of the world's most protracted and complex displacement crises, shaped by over fourteen years of conflict, recurrent shocks (including a devastating February 2023 earthquake that struck Aleppo), and systemic human insecurity. The collapse of the previous government in December 2024 and the subsequent transitional political process have marked a potential turning point for durable solutions, while also introducing new uncertainty into the institutional and governance environment in which return and reintegration programming operates. At the time of the ToR (August 2025), more than 7.4 million Syrians remained internally displaced and an estimated 6.4 million continued to live as refugees outside the country, while approximately 850,000 refugees had returned to Syria and 1.7 million IDPs had returned to areas of origin or relocated elsewhere in the country. This report updates that baseline with the most current figures available at the time of drafting: as of May 2026, an estimated 5.5 million people remain internally displaced inside Syria, and approximately 4.7 million Syrian refugees remain registered in neighbouring countries and North Africa (1.79 million in Egypt, Iraq, Jordan and Lebanon; 2.87 million in Türkiye; and over 43,000 across North Africa). Since the December 2024 political transition, an estimated 1.9 million IDPs and 1.7 million refugees have returned to Syria - both figures continuing to rise from the ToR's August 2025 reference point, with refugee returns in particular roughly doubling over that period. Some 15.6 million people - over 61% of Syria's 25.6 million population - are still estimated to require humanitarian assistance in 2026, and UNHCR and partners report that returns are occurring against continued structural constraints (damaged housing, limited services, documentation barriers, and constrained livelihoods) rather than as evidence that underlying conditions have been resolved. This evaluation should be read against this evolving national picture: return and reintegration needs remain very large in scale even as the overall displacement caseload gradually declines (UNHCR, *Syria Operational Update*, May 2026; UNHCR Operational Data Portal, Syria Regional Refugee Response situation page, accessed July 2026).

Bab Al-Nairab, a historic, densely populated district of Jabal Samaan in Aleppo City, was selected as an ABRS pilot location following the 2023 earthquake and the area's cumulative exposure to conflict-related damage, displacement and return pressure. The desk review's baseline sources describe the sector as spanning approximately 2.035 km² across seven neighbourhoods with an estimated population of approximately 131,894 (27,521 households), of whom only around a quarter of original residents had returned at the time of baseline data collection; the area also hosts several hundred displaced families from Idlib governorate.

The ABRS model itself is a UN-wide organizing framework - not a discrete UNDP or UNHCR project. It was institutionalized in 2022 under the RC/HC's leadership, and is technically co-led by UNDP and UNHCR within the Return and Reintegration Technical Working Group. Since 2023–2024 it has been adopted as a UN-wide approach, operationalized through Area Humanitarian Country Teams (AHCTs) across Syria; ABRS has been piloted in nine locations to date, including Bab Al-Nairab, the majority implemented concurrently rather than in strict sequence. The Bab Al-Nairab pilot, funded through a

joint UNDP–UNHCR proposal to UNTFHS/JSB submitted in December 2023, is the first of these nine pilots to undergo independent evaluation.

2.2 Evaluation Purpose and Objectives

Per the ToR and the approved Inception Report, this is an independent, learning-focused thematic evaluation - not an audit or compliance review. Its overall purpose is to assess the performance, operational effectiveness and early results of the ABRS pilot as a demonstrative case of ABRS implementation in Syria, examining relevance, coherence, impact, coordination among AHCT partners, and methodological soundness within the HDP Nexus, Durable Solutions and Human Security frameworks. Its five specific objectives, as stated in the ToR, are to: (1) assess impact, performance and operational effectiveness, including implementation progress and early outcomes; (2) evaluate relevance and coherence with local recovery priorities, conflict sensitivity and social cohesion; (3) examine the functionality and added value of coordination mechanisms; (4) review the methodological soundness of the ABRS model; and (5) generate lessons learned and strategic recommendations for adaptation, refinement and scale-up.

Consistent with the Inception Report, this evaluation does not assess "return promotion" as such, but rather how the ABRS approach contributes to enabling conditions for sustainable reintegration in areas of return - for those who have voluntarily decided to return, alongside IDPs and host-community members - in a fragile, evolving context. Any conclusions on scale-up or replication are framed cautiously, recognizing that transferability is contingent on context-specific political, institutional, security and resource conditions that lie substantially outside the pilot's control.

2.3 Scope and Limitations

The evaluation covers ABRS pilot activities in Bab Al-Nairab from inception (February 2024) through October 2025, geographically centred on Bab Al-Nairab with consideration of linkages to the wider Aleppo context, and thematically covering all three ABRS pillars and the cross-cutting issues of human rights, gender equality, disability inclusion and (to the extent evidenced) environmental considerations. A full account of methodological limitations is provided in Section 4.4; in summary, the principal scope limitations affecting this draft are: (i) a household survey sample that is heavily male-skewed (79%) relative to the Inception Report's 50–60% female target, which limits confidence in gender-disaggregated quantitative findings; (ii) FGD and MSC data collection that fell short of planned targets (nine of ten FGDs; two of three-to-five MSC narratives); (iii) small subsample sizes for several disaggregated survey indicators (e.g., legal assistance recipients, n=19; IDP respondents, n=18; respondents with disabilities, n=16), which the evaluation treats as indicative rather than statistically robust.

3. PROGRAM DESCRIPTION

3.1 Reconstructed Theory of Change

The ABRS Theory of Change, as set out in program documentation, states: "If an area-based, people-centred and participatory approach is implemented through coordinated multi-stakeholder action, strengthening protection, community empowerment and access to basic social services, then communities in Bab al-Nairab will experience improved human security, social cohesion and resilience, enabling sustainable early recovery and progress toward durable solutions." The pillar-level logic set out below elaborates this overarching statement for evaluative purposes - disaggregating it into the three pillars' specific causal steps, assumptions and risk conditions - **reconstructed** from the ToR, the Inception Report, and the program logic implicit in the three pillars and their reported results chain, consistent with the Inception Report's own commitment to "validating the program's theory of change" as part of contribution analysis.

If displacement-affected communities in an area of return receive (a) legal assistance to resolve HLP and civil-documentation barriers, (b) a functioning, inclusive platform through which they can identify and prioritize their own recovery needs, and (c) rehabilitation of the basic infrastructure and services those platforms prioritize,

then returnees, IDPs and host-community members will experience improved safety, dignity, access to rights and services, and social cohesion (human security),

because legal certainty reduces exploitation and enables access to services; genuine participation builds trust between communities and both local authorities and international actors, and produces better-targeted investments than externally designed ones; and visible, jointly prioritized infrastructure improvements create shared benefit across displacement-status lines rather than reinforcing perceptions of favouritism -

provided that local authorities remain willing and able to engage in good faith, funding is sufficient and predictable enough to complete (not merely start) infrastructure works, gatekeeping by informal power-holders (e.g., mukhtars) does not distort platform representativeness, and the security and political environment permits sustained access and community mobilization.

The evaluation finds this reconstructed logic to be broadly well evidenced at the level of its first-order assumptions (Section 5.1, 5.3), but only partially tested at the level of its human-security outcome claims. The desk review, corroborated by the evaluation's own review of the M&E framework (Section 5.6), found that the pilot's own indicator set does not, on its own, fully operationalize "human security" beyond output and reach counts. This is a scope limitation appropriate to a single pilot, not a design flaw. The finding concerns the M&E framework's fitness for future, larger-scale measurement (see Recommendation R1), not a criticism of the project's own indicator design. Nor does it extend to whether human-security change occurred: triangulated evidence elsewhere in the evaluation (Finding 5.8.1) independently indicates a credible, if modest, contribution to human-security conditions, drawing on sources outside the project's own indicator set.

3.2 Pillars and Intervention Logic

Pillar 1 - HLP legal assistance and civil documentation. Delivered primarily through SARC-run community centres under a long-standing UNHCR partnership arrangement, complemented by a separate UNDP-led legal sub-project. Reported reach: 3,998 direct legal-assistance beneficiaries plus over 1,500 remote consultations (Final Narrative Report). Services include civil documentation, marriage/paternity registration, HLP counselling, and referral support.

Pillar 2 - Local Participatory Platforms (LPPs). Community structures through which residents - with particular emphasis on women and youth - identify recovery priorities, engage in decision-making, and interface with local authorities. The pilot's community-engagement pathway operationalized this through three rounds of "MRS" diagnostic workshops, a multi-axis co-design process, and a validation exercise, producing a portfolio of 90 prioritized community-recovery projects from 99 identified problems and 435 proposed solutions (desk review). Reported LPP membership: 81 members (41 women, 40 men, 3 persons with disabilities) per the Final Narrative Report.

Pillar 3 - Quick Impact Projects (QIPs). Rehabilitation of essential community infrastructure to deliver early "peace dividends." Delivered outputs include approximately 1 km of sewage network rehabilitation (reported to benefit 2,000 households), a rehabilitated vocational training centre (the "Al Khalil School" / Sixth Industrial School site), street lighting (including a community-designed solar lighting quick-win initiative), and 52 short-term cash-for-work jobs linked to infrastructure works.

3.3 Stakeholder Ecosystem

Consistent with the Inception Report's stakeholder mapping, the pilot's ecosystem comprises: **beneficiaries** (returnees, IDPs, host-community members); **LPP members and community leaders** (including women, youth, and persons with disabilities); **implementing partners** (SARC for legal aid; a locally contracted community-engagement unit; a construction contractor for the school/VTC rehabilitation); **UNDP and UNHCR** (as lead and co-lead agencies respectively, with UNHCR's comparative advantage in protection/legal/documentation and UNDP's in early recovery/governance/infrastructure); **government and local authorities** (Aleppo Governorate officials, municipal departments, technical directorates for electricity/water/sanitation, and local mukhtars as informal community gatekeepers); **donors** (UNTFHS/JSB, with an HSTF Window 1/Window 2 structure separating operational funding from evidence-building); and **coordination bodies**, most significantly a Solutions Working Group that merged former Returns/Reintegration and INFLUX working groups and now anchors ABRS and area-based programming - within *the return-and-reintegration priorities it reports upward to the Humanitarian Country Team (HCT) and UN Country Team (UNCT)* - within the 2026 Humanitarian Program Cycle and the forthcoming UN Sustainable Development Cooperation Framework.

A recurring, cross-source finding regarding this ecosystem - corroborated independently by field-level KIIs, FGDs, and the desk review - is the significant informal power exercised by local mukhtars over project and participant selection, discussed further in Sections 5.1, 5.3 and 5.9.

4. METHODOLOGY

4.1 Design

As set out in the approved Inception Report, the evaluation applied a mixed-methods design combining qualitative and quantitative approaches, explicitly oriented toward **contribution analysis rather than causal attribution**, given the absence of a counterfactual or a program baseline against which to measure change. The design combined three data streams: (i) document review (desk review); (ii) qualitative data collection (KIIs, FGDs, MSC narratives, structured observation); and (iii) a quantitative household survey. Triangulation was applied systematically across data sources, methods, stakeholder groups, and locations within Bab Al-Nairab, and was explicitly used to test human-security dimensions (safety, dignity, access to rights and services, social cohesion) by comparing evidence across sources, population groups and pillars.

4.2 Sampling and Data Collection: Planned versus Achieved

Method	Planned (Inception Report)	Achieved	Assessment
Household survey	200 (up to 220 with oversample); stratified by displacement status, gender (50–60% women target), age, disability (3–15% target), pillar	n=204; 57% host community/34% returnee/9% IDP; 79% male / 21% female ; 8% disability	Sample size target met. Gender stratification target not met - a material limitation for gender-disaggregated findings (Section 4.4). Disability representation (8%) within target range.
Key informant interviews	approximately 30, across program staff, local authorities, LPP members, NGOs/CSOs, donors, external experts	Approximately 30 across three complementary reporting tracks (15 general/field-level; 9 national/coordination-level; 6 senior/strategic-level)	Target broadly met
Focus group discussions	10, stratified by displacement status, gender, age, disability	9 (3 women-only, 4 men-only, 2 mixed; approximately 60 participants)	Slightly below target; evidence gap for one planned group.

Method	Planned (Inception Report)	Achieved	Assessment
MSC narratives	3–5	2	Below target range; MSC evidence in this report should be read as illustrative of two individual cases, not representative.
Observation visits	2 (Visit 1: infrastructure/QIPs; Visit 2: services/education, including UNHCR community centre, legal clinic, Bar Association)	3 structured checklists completed (Al-Salihin sanitation network; Al-Marjeh sanitation network; Bab Al-Nairab vocational training centre)	Infrastructure and VTC observation completed with good detail.
Desk review	All relevant documentation	approximately 180 documents reviewed independently ahead of fieldwork	Exceeded typical scope in volume and produced a genuinely useful, internally cross-referenced set of secondary evidence. However, the evaluation team can only confirm what it was given access to and has no independent means of verifying whether this document set is complete or authentic; desk-review findings are therefore treated as secondary and corroborating rather than as freestanding conclusions

The achieved household survey sample (n=204) reflects a broadly representative geographic spread across Bab Al-Nairab's sub-areas, though respondents are unevenly split by displacement status and, most notably, by gender. By sub-area, the largest shares of respondents were drawn from West (26%, n=53), South (23%, n=46), East (22%, n=45) and Bab Al-Nairab Central (20%, n=41), with smaller shares from North (7%, n=14) and other/unspecified locations (2%, n=5); this spread supports the geographic

disaggregation presented in Finding 5.3.3. By population group, host-community respondents make up the majority of the sample (57%, n=117), followed by returnees (34%, n=69) and IDPs (9%, n=18); the comparatively small IDP subsample is flagged throughout Section 5 as indicative rather than conclusive. By age, respondents span all adult brackets with a concentration in the 25-44 range (34% aged 25-34, n=69; 28% aged 35-44, n=57), alongside 19% aged 18-24 (n=39), 9% under 18 (n=18, reflecting household members interviewed by proxy or as young heads of household), 8% aged 45-59 (n=17) and 2% aged 60 and above (n=4). Eight percent of respondents (n=16) reported a disability, within the Inception Report's 3-15% target range. As noted above, the sample's gender composition (79% male, n=161; 21% female, n=43) fell well short of the Inception Report's 50-60% female target; this is treated as a material limitation for gender-disaggregated findings and is addressed explicitly in Section 4.4 and Section 5.9.

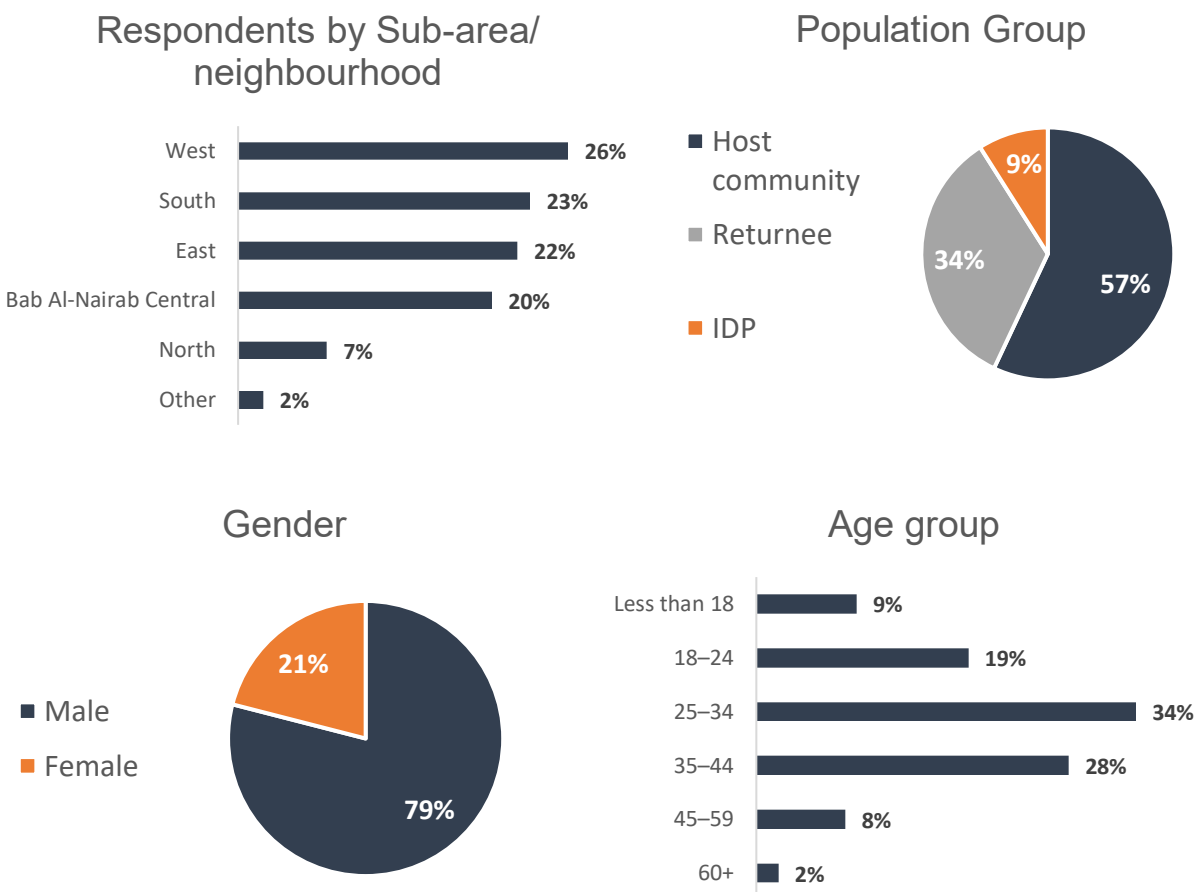


Figure 1: Survey Respondent Profile (sub-area, population group, gender, and age)

Figure 2 disaggregates the same sample by which ABRS pillar respondents report having engaged with, using a single derived "primary exposure" category to avoid double-counting across pillars. Just under half of respondents (48%, n=98) engaged only with a Quick Impact Project, a quarter (25%, n=50) engaged only with the Local Participatory Platform, and 8% (n=17) engaged only with HLP/legal

assistance. A further 6% (n=12) reported engagement with two or more pillars, and 13% (n=27) reported no recorded engagement with any of the three pillars; this last group's responses are nonetheless included in this evaluation's broader relevance and needs-based findings (Section 5.1) as a comparator against beneficiary experience.

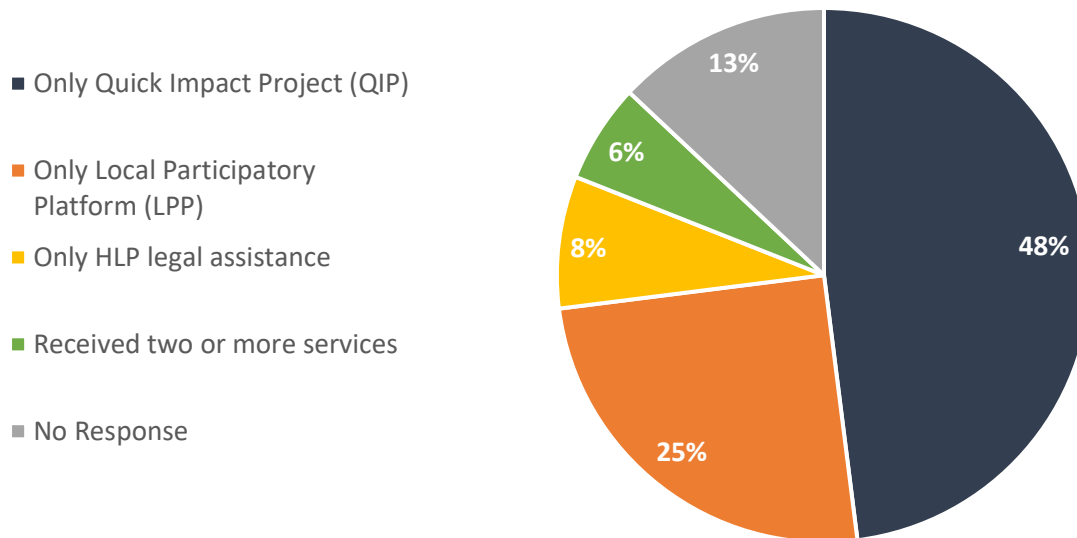


Figure 2: Beneficiary Exposure by Pillar

4.3 Analysis Approach

Quantitative data were analysed descriptively (frequencies, means, cross-tabulations) and disaggregated by gender, displacement status, age, disability and geography, subject to the small-sample caveats noted above. Qualitative data from KIIs, FGDs and MSC narratives were analysed thematically against the evaluation criteria. Contribution analysis was applied throughout: findings describe the pilot's plausible contribution to observed change, situated against other contextual factors (notably the political transition of December 2024, the 2023 earthquake, and pre-existing structural conditions), rather than asserting causality. Where a finding rests on a single source or an unverified recollection of an interviewee, this is explicitly flagged.

Confidence ratings:

Each finding is assigned a confidence rating reflecting the strength and independence of the evidence behind it, not the size or importance of the finding itself:

- **High confidence:** convergent evidence from at least three independent source types (for example, the household survey, KIIs, FGDs, desk review, or direct observation).
- **Medium-High confidence:** convergence from two independent source types, or a single, strong, well-corroborated source.
- **Medium confidence:** a finding resting on a single source, or on evidence with an identified limitation, such as a small subsample, unverified recall, or an attribution claim that could not be independently confirmed.
- **Low confidence:** little or no usable evidence on the topic despite the evaluation actively probing for it, or evidence too sparse, self-reported, or internally inconsistent to support even a single-source finding; the absence or weakness of evidence is itself flagged as the finding.

4.4 Limitations and Mitigation

Beyond the planned-versus-achieved gaps in Section 4.2, this evaluation identifies the following limitations, several of which were anticipated in the Inception Report and several of which emerged during data collection and desk review:

- **Desk review reliability and completeness.** The independent desk review draws entirely on program documentation made available to the evaluation team. The team has no way of confirming that this document set is complete, and no independent means of verifying the authenticity of what was shared. *Mitigation applied in this report:* the evaluation's primary evidence base is its own directly collected data (household survey, KIIs, FGDs, MSC narratives, structured observation). Desk-review findings are used carefully: where possible they are corroborated against at least one primary data source before being stated as a finding; where a finding rests on the desk review alone, this is stated explicitly rather than presented with the same weight as a triangulated, multi-source finding.
- **Gender representativeness of the household survey.** At 79% male, the survey sample substantially under-represents women relative to both the Inception Report's own target and to a robust basis for gender-disaggregated claims about beneficiary experience. This shortfall reflects the sampling frame available to the team: the beneficiary list shared by the project team was limited in size, and reaching respondents required directly contacting individuals on that list and obtaining their consent to be interviewed, which constrained the team's ability to purposively balance gender in the way achieved for the FGDs. *Mitigation applied in this report:* gender-related findings draw more heavily on FGD data (which achieved closer gender balance through purposive stratification) and are explicitly flagged as lower-confidence where they rely primarily on the household survey's female subsample (n=43).
- **Small disaggregated subsamples.** Legal-assistance (n=19), IDP (n=18), and disability (n=16) subsamples in the household survey are too small to support firm quantitative claims. *Mitigation applied:* these findings are presented as indicative and cross-checked against qualitative sources, consistent with the household survey's own key-findings analysis.

- **Recall-based, unverified figures from senior KIIs.** Several quantitative claims offered by senior/thematic informants (e.g., project budget, VTC enrolment, LPP gender composition) were given from interview recall rather than documentary record, and in some cases are corroborated by only one respondent. *Mitigation applied:* such figures are presented with an explicit verification caveat rather than as confirmed data.
- **Access and security constraints, and the political transition.** Consistent with the Inception Report's anticipated risk, the December 2024 political transition disrupted coordination, staffing, and administrative processes for the program itself during part of the review period, which the evaluation treats as important context for interpreting efficiency and timeline findings rather than as a limitation of the evaluation's own access (no material evaluation access constraints attributable to the transition were reported in the underlying data products).

4.5 Ethical Considerations

The evaluation adhered to UNEG's Ethical Guidelines for Evaluation and UNDP/UNHCR's respective evaluation and protection policies, as set out in the Inception Report: informed consent and voluntary participation; confidentiality and anonymity (no comments in this report are attributed to named individuals, notwithstanding that some source working documents retained names for internal team traceability); do-no-harm and conflict-sensitive framing of questions; gender-balanced and accessible data collection arrangements (including female enumerators for women's interviews and accessible venues for persons with disabilities); and secure, access-restricted data storage.

5. FINDINGS

5.1 Relevance

Primary evaluation question: To what extent is the ABRS pilot responsive to the needs and priorities of returnees, IDPs, host communities and local authorities in Bab Al-Nairab, and does it reflect HDP Nexus, Human Security and Durable Solutions principles?

Finding 5.1.1 - The pilot's design is strongly and independently evidenced against documented community needs. [High confidence]

Three independent baseline evidence streams reviewed in the desk review - an infrastructure/demographic baseline survey, community FGDs, and a dedicated conflict-sensitivity and protection risk analysis (February 2024) - converge on degraded sewage, intermittent electricity, HLP informality, and protection risks (GBV, child labour, weapons prevalence) as priority concerns. This is corroborated by every primary data source collected directly for this evaluation: household survey respondents cite basic services (79%), livelihoods (55%), education (41%) and health (41%) as their top current challenges (respondents could name up to three challenges, so figures do not sum to 100%); general/field-level KII respondents converge on sanitation, electricity, water, and HLP/civil documentation as foundational needs; and FGD participants across all nine groups cite livelihoods, sanitation completion, electricity and housing repair as the most frequently unmet needs. The alignment between documented baseline needs and the pilot's eventual three pillars is close and well triangulated. *"Sanitation was the most prominent need - electricity and water were also urgent, but sanitation took priority."* (General KII, sanitation company engineer)

The survey's own open-ended closing question (n=204) independently echoes this needs picture from a different angle: training/vocational skills and education (29% of responses) and livelihoods/income (28%) are the two most frequently volunteered priorities for future programming, both raised unprompted rather than selected from a fixed list – a further, independent form of corroboration for the relevance finding above (see Figure 12 below).

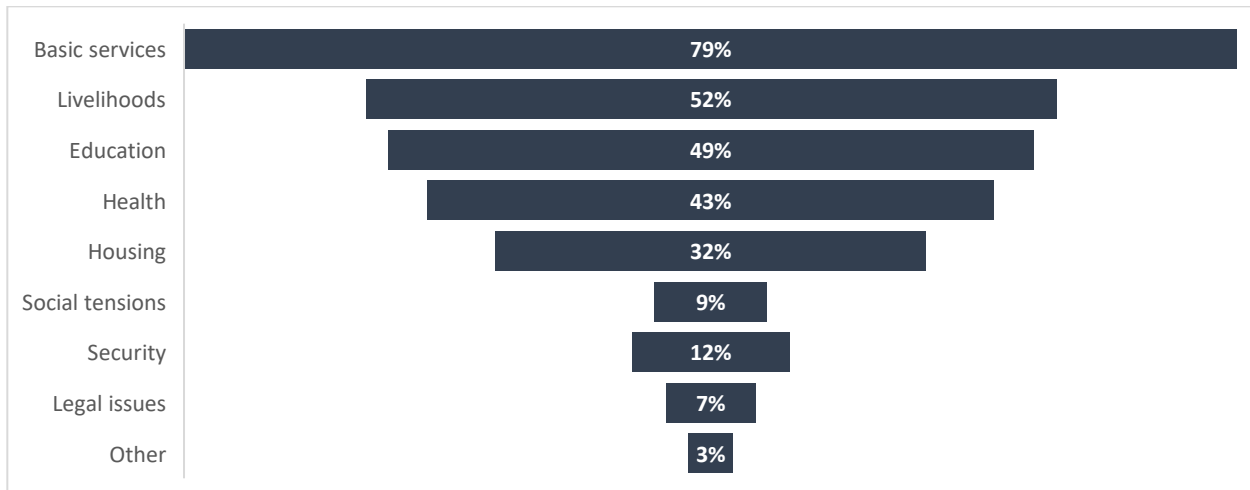


Figure 3: Top Reported Current Challenges

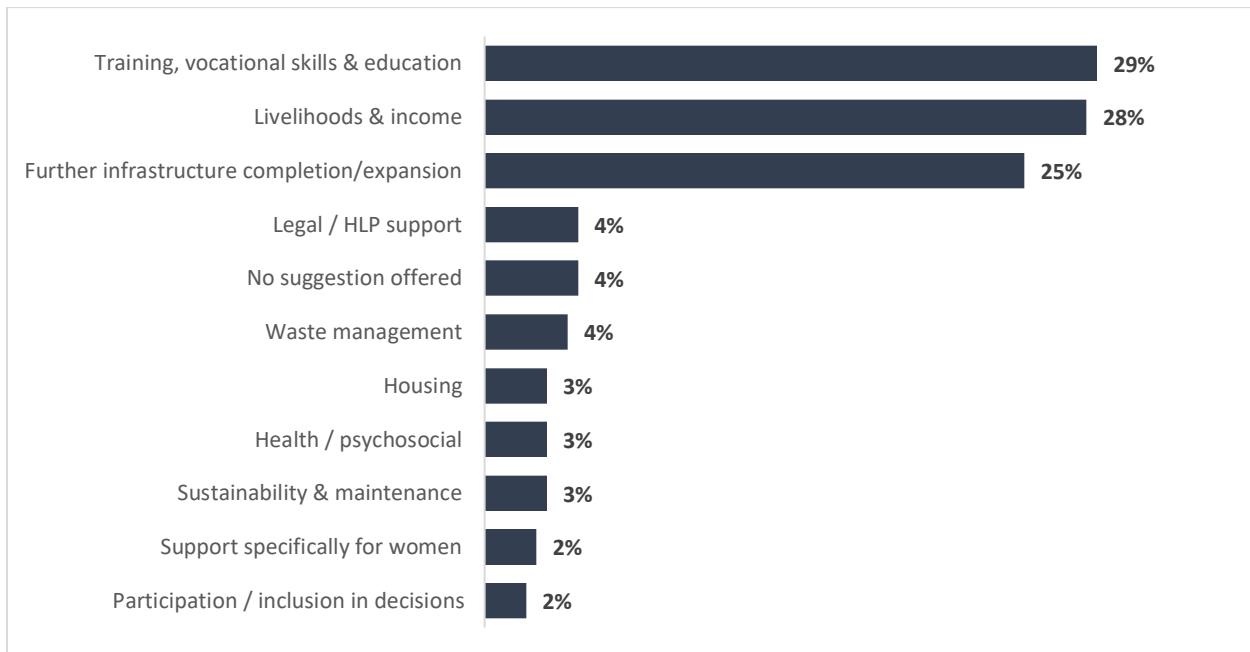


Figure 4: Themes in open-ended suggestions for future programming (n=204)

Finding 5.1.2 - Housing/shelter repair is a well-corroborated relevance gap, sitting outside the pilot's scope by design. [High confidence] Three independent field-level KII informants across local-authority and NGO-implementer categories, the FGD data (housing repair funds cited as a top unmet need across groups), and the household survey (housing cited by 37% of respondents as a current top-three challenge) all converge on this point. The evaluation reads this as a scope limitation consistent with the pilot's stated focus (legal aid, participation, and QIPs, not shelter reconstruction) rather than an implementation failure, but it is a materially important input for any future programming or scale-up design.

Finding 5.1.3 - Perceived responsiveness varies sharply by vantage point, and this variation is itself informative. [Medium-High confidence] Agency- and coordination-level KII informants rated ABRS's responsiveness to needs highly, while informants closer to specific streets or sub-areas gave markedly lower, more granular estimates (e.g., one field-level informant estimated needs met at "no more than 10%" in a specific sub-area; another at "40%"; another at "60–70%"). These are individual, subjective impressions rather than measured indicators, but their consistent directionality - the further a respondent sits from day-to-day delivery, the more positive their relevance assessment - is a pattern that recurs across multiple criteria in this evaluation (see also 5.3, 5.4) and should inform how future monitoring data is interpreted. *"It didn't meet more than 10% of needs in the Firdaws area, for example. There was no medical point; as residents we asked for repairs but saw no real progress over the year - electricity is still absent."* (General KII, community leader)

Finding 5.1.4 - The human-security and Durable Solutions framing is well understood and owned at the strategic level, but not independently validated by beneficiary-level evidence of framework awareness. [Medium confidence] Senior/thematic KII respondents across UNDP, UNHCR, and the UN Global Solutions Hub independently describe this framing as substantively embedded in the pilot's design, not applied retrospectively by the evaluation. UNDP's regional Solutions Advisor situates ABRS within UNDP's institutional strategy on solutions to internal displacement, noting that UNDP is globally mandated as one of three "solutions champions" (with UNHCR and IOM) under the UN Secretary-General's Action Agenda on Internal Displacement. Moreover, the key informant from the UN Global Solutions Hub locates ABRS within the established HDP Nexus and interagency guidance (IASC, UN Sustainable Development Group), describing it as "a good example" of the nexus ingredients working together rather than a distinct or novel approach. Additionally, the UNHCR Senior Repatriation Officer affirms the framing was deliberate and clear from design: *"the framing of the engagement was clear... this is nexus programming. From the point of design, there has to be built-in flexibility."* Taken together, these three independent, cross-institutional accounts support the strategic-level ownership claim. However, no source available to this evaluation shows beneficiaries themselves using or recognizing this framing; beneficiary-level evidence (survey, FGD) speaks to concrete needs and services rather than to the conceptual frameworks used to justify them. This is expected and not a criticism of design, but the evaluation notes it so that "human security" and "Durable Solutions" claims in this report are understood as an evaluative lens applied by the evaluation team and senior stakeholders, not a framing validated bottom-up.

5.2 Coherence

Primary evaluation question: How coherent is the ABRS intervention with broader UN, governmental and partner strategies, and does it foster effective multi-stakeholder coordination?

Finding 5.2.1 - Strategic coherence with the UN architecture in Syria is strong and has strengthened over the pilot's life. [High confidence]

The desk review documents that ABRS has been absorbed into a single inter-agency Solutions Working Group, is being mainstreamed into the 2026 Humanitarian Program Cycle and the forthcoming UN Sustainable Development Cooperation Framework, and is being replicated in eight further

governorates. Senior KII respondents independently situate ABRS within UNDP's global institutional strategy on displacement solutions (one of three UN "solutions champions" alongside IOM and UNHCR under the Secretary-General's Action Agenda on Internal Displacement) and within jointly owned IASC/UN Sustainable Development Group guidance. This is treated as strong, multiply-corroborated evidence that the pilot has functioned as intended: as a testable model feeding a national framework, not a stand-alone project.

Finding 5.2.2 - "Integration" should not be overstated; the substantive coherence achievement is alignment of priorities, not merged agency delivery. [High confidence]

This is one of the most robust convergent findings in the entire evidence base: three separate senior informants, speaking independently and using different language, each cautioned against romanticizing UNDP–UNHCR collaboration as single integrated delivery. UNDP and UNHCR retained separate systems, funding modalities (UNHCR's program/credit modality versus UNDP's project-based, donor-funded modality), and implementation procedures throughout. The evaluation treats this convergence - from senior figures with every institutional incentive to describe the partnership in the most favourable terms - as a credible, load-bearing finding rather than a minor nuance. *"Agencies retained their respective systems, sources, partners and implementation procedures - which is normal. The real achievement was that they were increasingly directed towards a common area priority, rather than operating in parallel."* (Senior KII, UNHCR field management)

Finding 5.2.3 - Coordination mechanisms are well documented and functioning, with transition-related disruption as the main documented caveat. [Medium-High confidence]

Both the general/field-level and thematic KIIs describe multiple specific, named coordination mechanisms (WhatsApp coordination groups, joint field visits, the OCHA platform, a national International Cooperation Directorate linking donors to government technical directorates, the Solutions Working Group). Two field-level informants specifically flagged coordination disruption tied to the December 2024 political transition (administrative/territorial boundary redefinition; a period of administrative disorder that forced work to be routed through an implementing partner). This is corroborated at the thematic level by senior informants who independently name the transition (staff evacuation/security confinement, customs/regulatory disruption) as an efficiency constraint (see 5.4). *"I think the coordination was quite high, because the program's nature is participatory - that's clear from its name."* (General KII, local authority official)

Finding 5.2.4 - Local ownership and authority engagement is structurally embedded but unevenly distributed. [Medium-High confidence]

The desk review's own pre/post workshop surveys record a measured shift in participants' sense of responsibility (agreement that "I am responsible for contributing to solving my community's problems" rose from 76% to 100% in Round 1), and separately document participants self-organizing between formal sessions - genuine, not managed, ownership behaviour rather than an evaluator's inference (see also Finding 5.6.1). This is corroborated at the field level: most general-KII informants who addressed the question directly rated ownership as strong to complete, anchored most concretely in the

requirement that local government approve every activity. The oft-cited estimate that local authorities hold roughly 70% of ownership relative to the community came with a caveat rarely quoted alongside it - the same informant expected community ownership to run higher on area-based return work specifically (General KII, local authority official) - and should be read with that qualification rather than as a fixed ratio. A single dissent, from a legal professional who rated ownership "limited and constrained" (General KII, legal professional), is treated as a credible minority view rather than an outlier to be discarded. The senior/thematic KII track adds a more institutionally distant corroboration: a UNHCR field-management informant independently assessed local ownership motivation as "very strong", while flagging that its durability depends on whether current Syrian authorities remain in place, given the country's still-forming governance structure - the same tension already noted for site-specific sustainability at Finding 5.7.1. The community side of this triangulation comes from the FGD's neighbourhood committee, which independently corroborates genuine ownership from the beneficiary end, citing concrete instances where its own proposals shaped what was actually built (Finding 5.3.5). Two caveats temper the overall picture. First, the desk review's facilitation team itself flagged that institutional access repeatedly ran through a single local-authority gatekeeper - an efficient arrangement, but a concentration risk in its own right. Second, Local Participatory Platform Terms of Reference remain, in places, generic UNDP-wide templates not fully adapted to Bab Al-Nairab specifically - a reason to distinguish site-specific commitments from generic ABRS-wide language when assessing what was actually promised locally. A further, independent caveat on this same point comes from the national/coordination-level KIIs: several of these informants distinguished ownership from institutional capacity to sustain it, noting that local authorities' willingness to lead is real but is constrained by limited financial resources, insufficient staffing, weak logistical capacity, and limited technical expertise - a capacity gap, not a commitment gap (National/coordination-level KII).

5.3 Effectiveness

Primary evaluation question: To what extent were the pilot's planned outputs achieved, and what early outcomes are observable on human security, participation and social cohesion?

Finding 5.3.1 - Reported output achievement is substantial across all three pillars. [Medium-High confidence - output figures per Final Narrative Report, with the caveats in Section 4.4]

Under the legal-assistance pillar: 3,998 direct beneficiaries were reached, plus 1,500+ remote consultations. Separately, the program's results framework reports 4,102 against the Outcome 1 indicator ("# of key community members involved in developing protection and participatory strategies"), which aggregates 3,998 legal-assistance beneficiaries, 81 LPP members and 23 local-authority representatives; this is a distinct, outcome-level indicator rather than a reach count, and should not be summed with the reach figures above. Under the LPP pillar: 81 members (41 women, 40 men, 3 persons with disabilities), and a community-engagement pathway that produced a traceable chain from 99 identified problems to 435 proposed solutions to a validated 90-project recovery portfolio. Under the QIP pillar: approximately 1 km of sewage network rehabilitated (reported to benefit 2,000 households), a rehabilitated vocational training centre, and 52 cash-for-work jobs. These figures should be read together with the desk review's specific, unresolved concerns about the Outcome

2 beneficiary redefinition and the possible VTC/school double-counting issue (Section 3.2), which this evaluation could not independently resolve.

Finding 5.3.2 - Household survey data show broadly positive perceived early outcomes across the sampled population. [Medium-High confidence, subject to the gender-representativeness caveat in Section 4.4] 67% of survey respondents (n=204) report improved access to basic services (54% "improved," 12% "improved a lot"); 72% report improved safety and dignity; 62% report that relations between returnees, IDPs and host communities are "better," against only 4% reporting "worse" - a meaningful signal in an active return context.

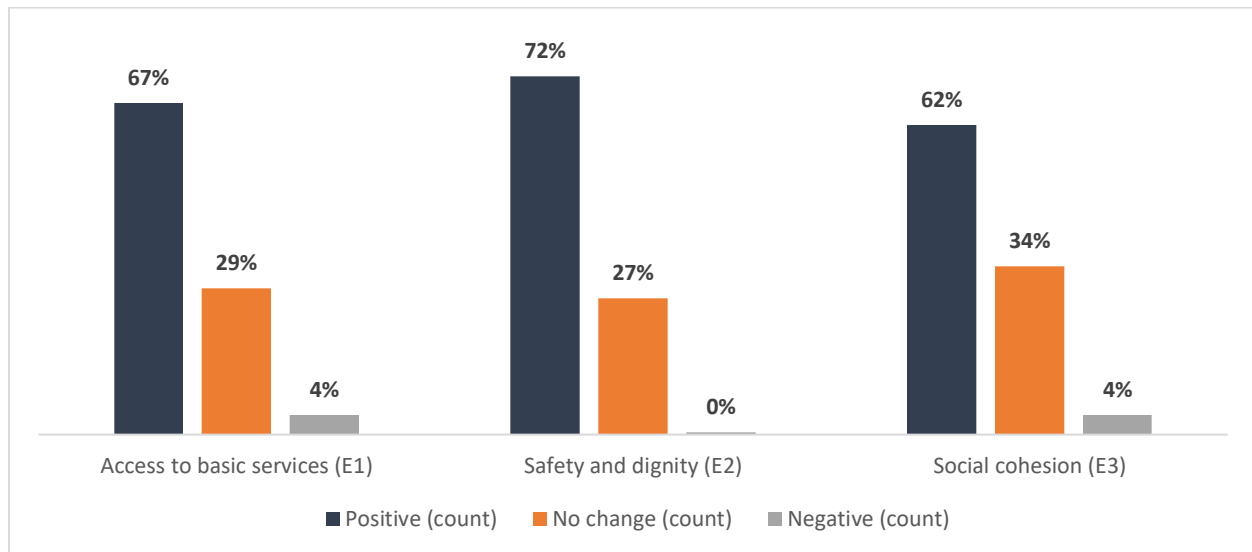


Figure 5: Perceived Change in Core Outcomes - Access to Services, Safety & Dignity, Social Cohesion

Satisfaction scores were consistently positive: legal-assistance satisfaction averaged 8.6/10 (n=19), LPP effectiveness 7.5/10 (n=62), and QIP satisfaction 7.5/10 (n=110). 79% of LPP respondents felt their views were heard and taken seriously, and 82% of QIP-linked facilities were reported in daily use. "Before the project we lived like it was the dark ages - there was no lamp post in our community. Now the health post and the roads are lit at night." (FGD2, resident man)

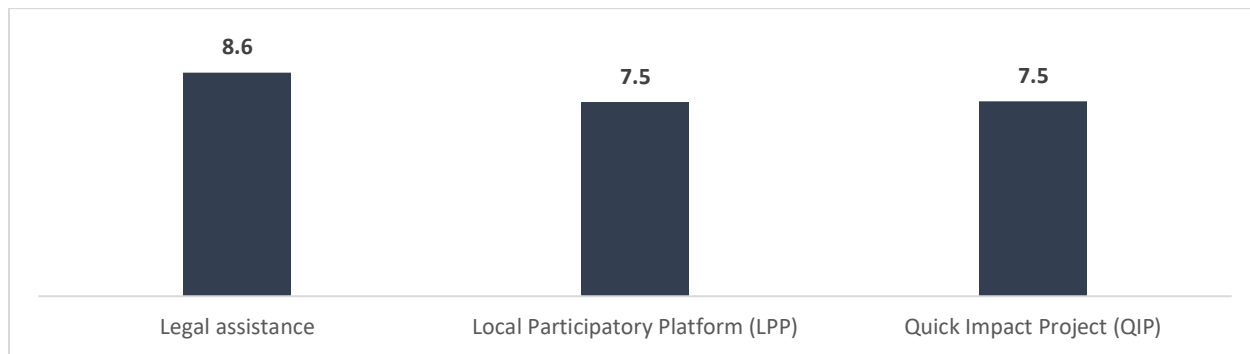


Figure 6: Beneficiary Satisfaction by Pillar (0-10) - Legal Assistance, LPP, QIP

Finding 5.3.3 - Coverage variation within Bab Al-Nairab is the single most consequential effectiveness finding, corroborated independently across every qualitative source. [High confidence]

FGD analysis identifies this as the most consequential finding in that dataset: perceptions of pilot's performance differs dramatically depending on whether a participant's specific street received a completed intervention, more than by any demographic factor. Two FGD groups reported project impact as minimal to non-existent, including one where a nine-month waste-collection service had been discontinued six weeks before the discussion - a municipal service outside this project's scope, cited here as context for the group's overall sentiment rather than as evidence of the project's own coverage gap. This is independently corroborated from the implementer side: a field-level KII informant gave a concrete, named example (a specific sub-area where needs were "no more than 10%" met), and the household survey's geographic disaggregation shows the South sub-area as a consistent laggard on every measure examined (lowest QIP satisfaction at 6.1/10 against a 7.5/10 mean; lowest feedback-mechanism awareness at 35% against a 53% mean; weakest social-cohesion gains at 39% against a 62% mean). No single sub-area leads on all three measures, however: Bab Al-Nairab Central records the highest QIP satisfaction (8.5/10) and, together with West, the strongest social-cohesion gains (76% and 75% respectively), while West leads specifically on feedback-mechanism awareness (70%) and East, not West, records the second-highest QIP satisfaction (7.9/10) - a reminder that "coverage variation" is a real, convergent pattern but not a simple single ranking of sub-areas. The evaluation treats this as a genuine, systemic finding - not an artifact of any single data source - with direct implications for how future ABRS phases sequence and communicate geographic targeting decisions. *"There have been no positive changes - on the contrary, some services like waste removal were stopped a month and a half ago, and there are many blockages in the sewage network. Most organizations have stopped supporting the area."* (FGD, resident man)

Furthermore, Al-Salihin observation checklist documents destroyed infrastructure in adjoining streets that fall outside the project's defined scope, falling within the responsibility of the relevant State institutions (municipal/sanitation directorates) rather than reflecting a shortcoming of the project, and illustrating the area-wide service-delivery needs that exceed any single QIP's scope (see Recommendation R5) - the same coverage-limitation pattern already identified through the household survey, FGD and KII data. On the other hand, the survey's open-ended suggestions add another independent corroborating source: further infrastructure completion or expansion - chiefly requests to

extend street lighting, sewage rehabilitation and road repair to streets not yet covered – was volunteered unprompted by 25% of all 204 respondents, the third most common theme in the entire open-ended dataset (Finding 5.1.1) and a direct, beneficiary-originated echo of the coverage-variation pattern documented above.

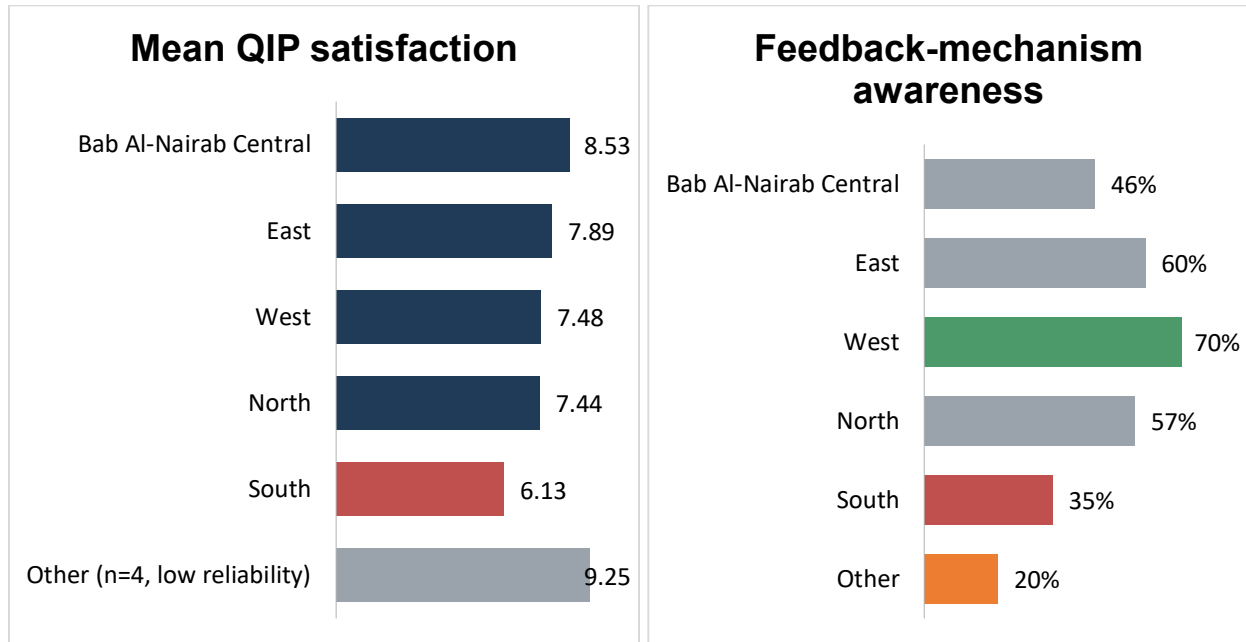


Figure 7: Geographic Variation in QIP Satisfaction and Feedback Awareness

Finding 5.3.4 - Returnees, a core target group, report comparatively less positive outcomes than host-community respondents on nearly every quantitative outcome indicator. [Medium-High confidence]

The household survey's key-findings analysis flags this explicitly as a concern warranting further investigation: returnees report lower service improvement (59% vs. 69–78% for host community/IDPs), lower safety improvement (65% vs. 73–94%), lower feedback-mechanism awareness (39% vs. 59–67%), and more reported facility problems. (Figures use the same "Improved"+"Improved a lot" positive-change definition applied consistently elsewhere in this report, e.g. Section 1.2 and Finding 5.3.2; the IDP subsample is small, n=18, and should be read as indicative.) This sits in tension with - and should be read alongside - FGD evidence of a distinct, separate finding: perceived favouritism toward returnees in aid and property-repair prioritization, raised independently in two FGD groups. The evaluation does not read these two findings as contradictory so much as evidence that different population groups experience different aspects of unevenness: returnees appear to fare worse on service/safety outcomes in the aggregate survey data, while some host-community and long-term-resident FGD participants nonetheless perceive returnees as being prioritized in specific, visible instances (e.g., property-related decisions). Both dynamics plausibly coexist and both merit explicit management attention; this is exactly the kind of divergence contribution analysis is designed to surface rather than average away. *"The relationship between returnees and residents has gotten worse because of the mukhtars - some of them discriminate, making it easier for a returnee's paperwork*

to move than a resident's, and I've seen this myself." (FGD1, returnee woman) This dynamic is independently corroborated from the implementer side: a general-KII respondent (UN agency) separately reported friction between host-community and returnee participants during implementation, reinforcing that this is a genuine, cross-source pattern rather than an artifact of a single data source.

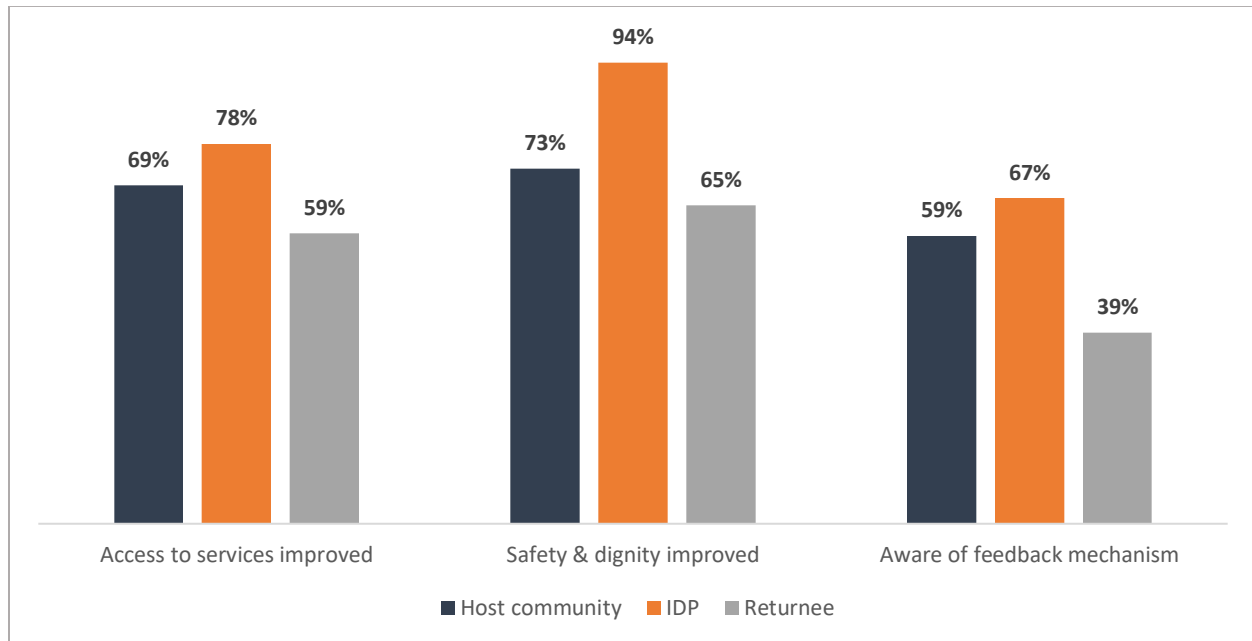


Figure 8: Outcome Gaps by Population Group - Returnees vs. Host Community vs. IDPs

Finding 5.3.5 - Local Participatory Platforms functioned as a genuine, if imperfect, prioritization mechanism, with well-documented gatekeeping risks. [Medium-High confidence]

LPP survey respondents (n=62) report high perceived representativeness for youth (92%) and women (81%), lower for returnees (65%) and IDPs (56%), and markedly lower for persons with disabilities (19%). Qualitatively, the LPP mechanism produced concrete, traceable examples of community input shaping implementation. *"We proposed to UNDP that Bab Al-Nairab's priority was the sewage network because the overflow was suffocating people in winter. Two months later they began rehabilitating the streets we had identified on the map."* (FGD9, neighbourhood committee member, resident)

Set against this, both the FGD data and field-level KIIs independently and specifically corroborate a pattern of mukhtar-mediated gatekeeping. *"Yes, there were problems with the mukhtar - he was biased and controlling. We spoke to the trainer to handle this, so they moved the workshop away from him."* *"The mukhtar won't approve projects unless he's personally convinced. We suffer from this - for example, we needed lighting around the schools, but the mukhtar opposed the project and personally decided which streets would be lit instead."* (General KII, community leader) The evaluation treats this convergence (an implementer's own account corroborating community complaints) as elevating mukhtar gatekeeping from a perception to a corroborated, cross-source pattern requiring continued,

further-strengthened mitigation in future LPP design, building on the program's existing efforts to move beyond reliance on traditional leaders - with mukhtars engaging as one stakeholder among many and facilitators remaining responsive to community concerns throughout - rather than an unaddressed design gap. Representation figures triangulate imperfectly across sources. A senior KII respondent (UNHCR) recalled approximately 48% women and 42% youth among roughly 90 LPP community representatives, broadly consistent with the reported LPP membership breakdown (41 women, 40 men, 3 persons with disabilities among 81 members - Finding 5.3.1).

Senior key informant (UNHCR headquarters), pressed directly on gender representation, and was candid that full representation was not achieved and that participation had to remain voluntary given social norms: *"Can I say that full representation exists within this program? No, I cannot. We tried to ensure that vulnerable groups were represented. But it has to be voluntary."* The FGD data add sharper texture on perceived influence, with one LPP-experienced participant estimating the community's actual decision-making influence at "about 10%" and another describing the mechanism in starker terms: *"Our impact is symbolic at best - it feels like the organizations bring us in to justify decisions they've already made, not to hear our opinions."* (FGD, IDP man) A separate FGD participant with a physical disability specifically linked low disability inclusion to a planning gap: *"I need help moving around the pavements and facilities, and this is a demand that everyone overlooks."* (FGD, man with a physical disability)

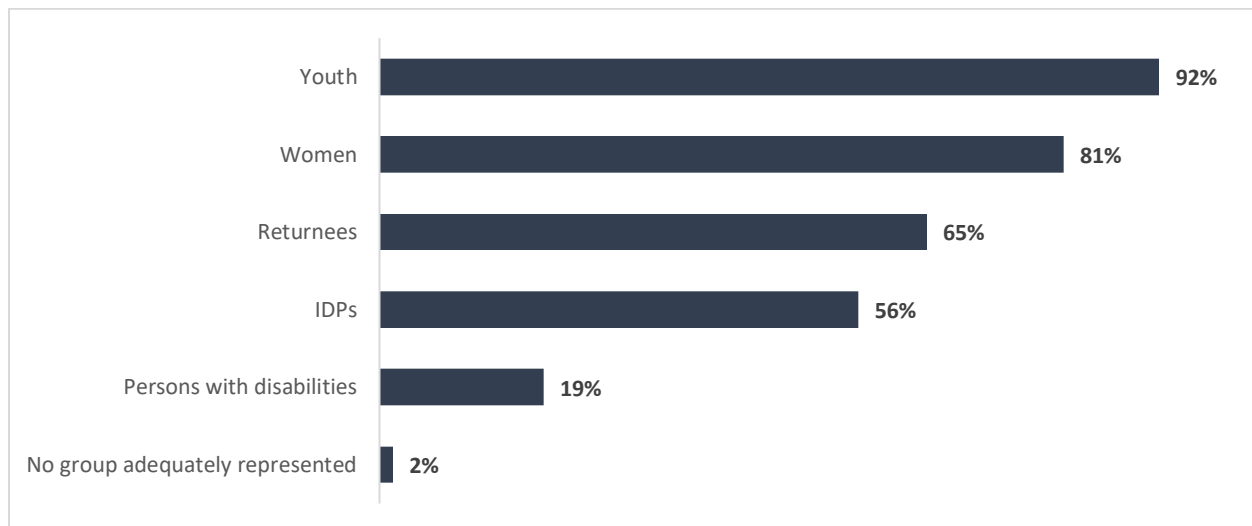


Figure 9 Perceived Representation in LPP Activities (C3, n=62)

Finding 5.3.6 - Legal assistance is the most consistently and strongly rated pillar across every data source. [High confidence]

Household survey respondents who received legal assistance (n=19) report high satisfaction (mean 8.6/10), high perceived confidentiality (89% "always"), and a 74% issue-resolution rate. FGD participants describe concrete outcomes (marriage/paternity registration, birth certificates for children

born abroad, housing-dispute guidance). *"The landlord I rent from refuses to acknowledge the lease. The agency helped me secure my right and supported me with \$600 when I returned."* (FGD, IDP man)

Field-level and thematic KII informants across every category converge on legal assistance as the pilot's most effective and valued component, while also noting continued unmet demand given long judicial timelines and the pilot's short duration relative to caseload. *"Beneficiaries increasingly obtained legal and official documents and gained the motivation to correct their legal status... this strengthened legal protection for the most vulnerable groups, particularly women, children, older people, and persons with disabilities."* (General KII, UN agency legal aid coordinator)

A smaller but consistent share of the survey's open-ended suggestions (4% of 204 respondents) volunteered HLP/legal support as a priority for future programming - a modest but independent echo of the unmet-demand pattern described above.

Finding 5.3.7 - Infrastructure functionality is generally good but not problem-free, and workmanship/safety issues identified through direct observation are a distinct concern from usage/satisfaction data. [Medium-High confidence]

82% of QIP-linked facilities are reported in daily use and 63% show no current problems (survey).

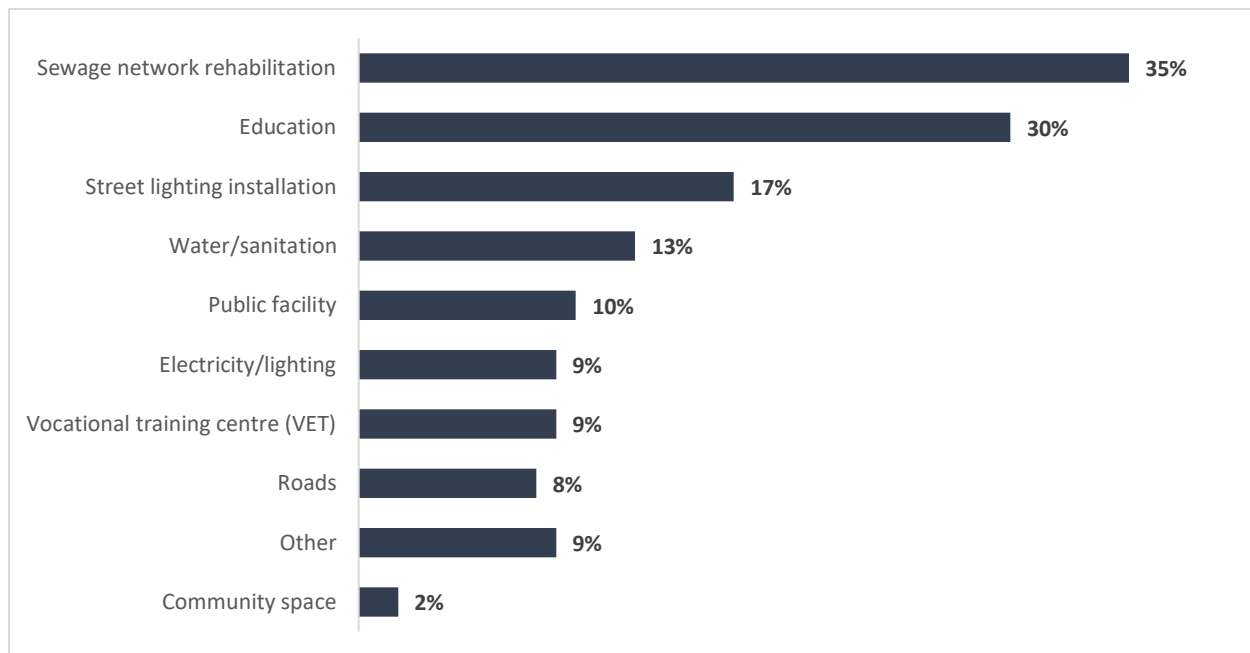


Figure 10: Quick Impact Project (QIP) Type Reported by Beneficiaries (n=110)

Structured observation at the two sanitation sites identified distinct, site-specific technical concerns rather than a single uniform issue. At Al-Salihin, observation found sections of exposed or damaged sewage line that residents flagged as a health and environmental-contamination risk, site cleaning that residents considered incomplete before pipe-laying and re-asphalting, and asphalt thickness residents considered below standard for road durability. At Al-Marjeh, observation identified a manhole/cement

installation error that traps rainwater and required residents to dig makeshift drainage channels themselves, a material-quality complaint (concrete versus reinforced-plastic piping causing recurring faults), and, most seriously, resident reports of **two individuals injured (bruises/fractures) after falling into sewage excavation pits during construction**. *"Residents complained about the length of the implementation period and reported that two people suffered bruises and fractures due to falling into sewage excavation pits during the project's implementation."* (Structured observation checklist, Al-Marjeh sanitation site) The implementing team has indicated that the specific pipework may not be part of the project's rehabilitated segments; this could not be independently verified, and the finding is retained as directly observed. These observations were made during fieldwork, while the one-year defect-liability period was still open; the Sewerage Directorate recorded no subsequent complaints on the project's rehabilitated sections, and upon the warranty's expiry on 7 July 2026, the Sewerage Company issued Final Acceptance Minutes confirming no outstanding defects. Both the sewage and VTC components carried formal warranty mechanisms, so concerns within the project's own scope were addressed through standard quality-assurance channels; issues beyond that scope should be directed to the responsible public authorities rather than read as a shortcoming of the project.

Observation at the UNHCR-rehabilitated vocational training centre (VET) found the external structure and interior finishing in good condition, but identified multiple incomplete or substandard elements (non-functional bathroom fixtures, an unrepaired drainage defect near the school entrance, unrepaired training-hall doors, poor-quality maintenance work in the basement level, undersized solar capacity, and students reportedly needing to walk approximately 20 metres with no covered walkway to reach bathrooms in winter rain - a dignity and accessibility concern in its own right), together with a reported lack of responsiveness from project staff to a site guard's technical concerns and an absence of visible project signage/information at the site. These observation-based findings corroborate, from an independent field-verification source, the desk review's separate finding of an undisclosed approximately 285% quantity overrun on this civil-works item - a contractor progress-reporting gap addressed further under Recommendation R5, rather than a functional defect in the completed works. An open snag list is being actively addressed by the contractor as part of its own contractual obligations, in accordance with UNHCR's standard quality-assurance and defects-liability procedures, under the standard one-year, 10%-retention defects-liability period, in coordination with the Aleppo Directorate of Education; the reported quantity variance has separately been referred for program-management review. Together, these point to a consistent, multi-source picture of gaps at the point of contractor progress reporting - in tension with reports that described the project as "on schedule" with "no challenges" - though remediation is now under way through standard warranty channels.

Additionally, an economic-harm example was documented at Al-Marjeh: residents reported that excavation and soil stockpiling in front of a resident's blacksmith workshop forced the business to close for approximately a month and a half during construction - a concrete, if localized, illustration of the livelihoods costs lengthy construction periods can impose on directly affected households. On the positive side, observation at the vocational training centre found the rehabilitated basketball and football courts fully painted and prepared.

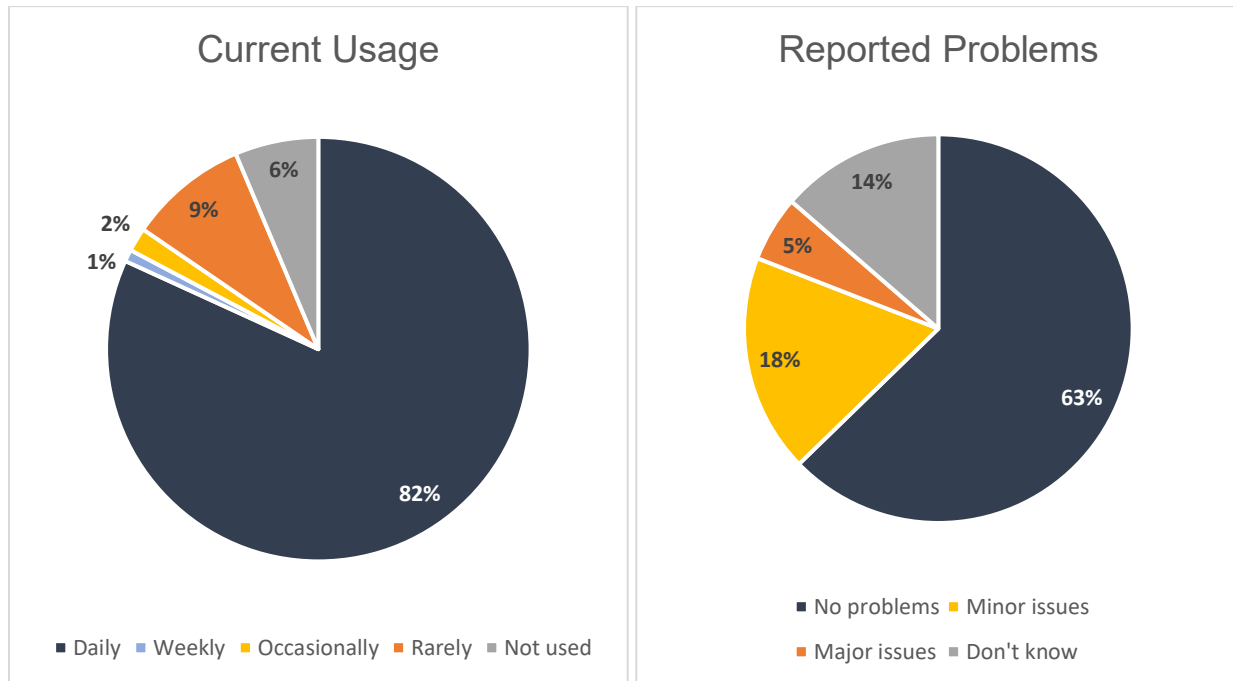


Figure 11: Condition and Usage of QIP-Linked Infrastructure (n=110)

5.4 Efficiency

Primary evaluation question: Were resources used optimally, and were coordination mechanisms cost-effective and operationally streamlined?

Finding 5.4.1 - Efficiency performance is genuinely mixed, not uniformly positive, and differs sharply between the sewage and school/VTC infrastructure lines. [Medium-High confidence]

The sewage rehabilitation was delivered within a tight three-month window with its full budget spent and no extension, despite a disruptive mid-project change in the receiving utility's management - a positive efficiency finding corroborated by this evaluation's own field-level KII evidence (Finding 5.4.3). By contrast, the UNHCR-financed VET rehabilitation shows a documented approximately 285% quantity overrun on its largest civil-works item that was not narratively disclosed as a deviation in progress reporting, no US-dollar cost figures anywhere in its procurement documentation (complicating benchmarking against exchange-rate volatility), and a separate legal sub-project that closed with 38% of its budget unspent alongside contradictory internal statements about whether a no-cost extension was granted. This evaluation's own field observation independently corroborates workmanship and quality-assurance concerns at the same VTC site (Finding 5.3.7), which strengthens confidence in the general picture even though the specific financial figures themselves remain a desk-review-only, unverified data point. The evaluation recommends these be treated and reported as

distinct efficiency profiles rather than summarized into a single, blended efficiency rating for "infrastructure."¹

Finding 5.4.2 - The extension of the implementation period was a legitimate, well-documented protection-driven trade-off, not primarily a planning issue. [Medium-High confidence]

Both the desk review and senior KII evidence independently and consistently attribute the No-Cost Extension to the VTC's original site being reassessed as unsafe (armed-group presence), requiring mine clearance and relocation to a Ministry of Education-cleared site. A senior informant explicitly characterized this as *"the correct protection decision"* (Senior KII, UNHCR field management) with an operational cost. The evaluation concurs that this specific delay reflects legitimate protection mainstreaming rather than poor efficiency, while noting (Finding 5.4.1) that separate, unrelated efficiency issues at the same site should not be conflated with or excused by this finding.

Finding 5.4.3 - Multiple, independently identified operational bottlenecks affected timeliness, most linked to the external operating environment rather than program design. [Medium-High confidence]

Field-level and thematic KII informants converge on: import/logistics delays for materials sourced internationally, compounded by customs disruption during the political transition; centralization of legal case processing in Damascus slowing turnaround; and administrative/coordination disruption following the transition. These are corroborated by the desk review's independent finding that the transition caused staff evacuation/security confinement and disrupted UNHCR's internal administrative/emergency-contracting processes. This is corroborated by a third, independent source: national/coordination-level informants separately identified changes in government counterparts and delays caused by centralized decision-making as live operational bottlenecks during implementation itself, consistent with - and predating - the unresolved government-counterpart gap documented at project close (Finding 5.7.1) (National/coordination-level KII).

Finding 5.4.4 - Differing UNDP-UNHCR business models caused some start-up friction alongside clear complementarity; one informant's view that joint working reduced efficiency is a minority perspective, not a standalone finding. [Medium-High confidence on the business-model/complementarity point; the dissenting view is a single source and is treated as a minority perspective, not a rated finding]

One field-level informant - drawing on prior direct experience implementing with UNDP alone elsewhere in Aleppo - assessed joint implementation as reducing efficiency: *"Honestly, the effect was negative. We had worked with UNDP alone in Old Aleppo, and as a single agency it gave more flexibility. When you have a partner, you get one viewpoint and then another - sometimes the*

¹ These specific figures are a desk-review documentary finding, drawn from procurement and financial records made available to the evaluation team, which could not independently verify the completeness of that document set.

difference is constructive, but sometimes it's the opposite. I think relying on a single agency would have been better, to save time and effort." (General KII, local authority official) The one senior informant who gave a direct verdict on this exact question took the opposite view, rating the joint architecture's effect on efficiency as "overwhelmingly positive" at the management level: "*I don't think there was any issue that was impactful in the negative way.*" (Senior KII, UNHCR field management) A second, narrower point is independently corroborated rather than contradicted: a genuine UNHCR/UNDP business-model mismatch - program-based versus project-based funding - complicated synchronizing when implementation could begin. "UNHCR operates through a program modality whereas UNDP is project-based. Synchronizing when we were able to actually implement was a lesson that was learned." (Senior KII, UNHCR headquarters) This is a bounded start-up friction, not evidence of a sustained efficiency drag - the same informant separately assessed the pilot as cost-effective overall (Finding 5.4.5). The evaluation treats the minority view as a genuine outlier worth probing further, given the informant's unusual direct comparative basis, rather than a broadly corroborated finding.

Finding 5.4.5 - Reported budget figures differ across documents for an identifiable, explainable reason: source-specific retention deductions. [High confidence - this is a documentary finding, not an interpretive one]

The original program proposal cites a total budget of USD 795,518; the signed UN-to-UN transfer agreement (USD 395,891.44) plus the UNHCR budget revision (USD 377,882) sum to USD 773,773 - a gap of approximately USD 21,745 that reflects the percentage retained by the UN Trust Fund for Human Security Secretariat, per its established procedures, to cover Secretariat coordination costs with the Joint Steering Board (JSB); the amounts actually transferred to each agency are those in the signed UN-to-UN Agreements. Separately, one senior KII informant cited an approximate total project budget of USD 800,000 from recall, which is broadly consistent with these figures. That same informant (UNHCR headquarters) separately framed the pilot as cost-effective, citing a "multiplying effect" from in-kind contributions - local-authority machinery and UN-Habitat's urban-renewal expertise - that is not captured in any of the budget figures above; this qualitative cost-effectiveness claim could not be independently verified by this evaluation.

Separately, on the UNHCR-financed VET component, UNHCR topped up the allocated budget by USD 50,380 above the USD 344,000 donor allocation - absorbing the cost of tender results coming in above budget rather than reducing scope - with a 5% retention (USD 18,600.23) held for six months pending close-out of the defects-liability period.

5.5 Coordination and Partnerships

Primary evaluation question: How effective were inter-agency and inter-sectoral coordination mechanisms, and did partnerships enable collective outcomes?

Finding 5.5.1 - Coordination and partnerships is the criterion with the widest spread of assessment across the evidence base, and this spread is itself the finding. [High confidence]

At the strategic/donor level, coordination is described in strongly positive, largely uncaveated terms. At the level of informants closest to actual delivery, assessment is positive overall but includes specific, concrete, named frictions: staff rotation in both agencies; an HQ/field organizational split; and - most substantively - a genuine business-model mismatch between UNHCR's program/credit modality and UNDP's project-based, donor-funded modality that complicated synchronizing funding and implementation timing. The evaluation's clearest methodological observation on this criterion is that senior/strategic and field/delivery informants are not contradicting each other so much as describing coordination at different altitudes, and that a future evaluation or monitoring report should resist collapsing this into a single "coordination was good" rating.

Finding 5.5.2 - Partnership breadth beyond UNDP and UNHCR is real but under-leveraged relative to its potential. [Medium confidence]

Coordination with UN-Habitat (urban-recovery planning expertise), local authorities and service directorates, and government coordination structures (an International Cooperation Directorate) is documented and credited with genuine complementary value by multiple informants. However, a senior informant specifically named WFP and UNICEF as agencies that *"should have been more involved"* (Senior KII, UNHCR headquarters), suggesting the AHCT/S architecture has not yet fully leveraged the wider humanitarian-development ecosystem present in Aleppo. At the time of implementation, responsibility for operationalizing the ABRS model rested with the Area Humanitarian Country Teams (AHCTs) referenced above. This is independently corroborated by one of the nine national/coordination-level informants interviewed for this evaluation, who offered a more critical assessment of the same architecture, arguing that although coordination mechanisms existed, they did not sufficiently engage humanitarian coordination structures, local civil society organizations, or community-based organizations - a specific, credible minority view rather than a general complaint (National/coordination-level KII).

Finding 5.5.3 - Concrete, specific improvement areas were volunteered by the most operationally engaged senior informant, and merit direct uptake. [Medium-High confidence]

These include: an explicit joint implementation plan agreed from project outset (not only at kickoff); a shared action tracker with clear sequencing and responsibilities; a common area-level results framework capturing outcomes, not only activities; more systematic joint monitoring; and more predictable, flexible multi-agency financing. The specificity and internal consistency of these recommendations, from an informant with direct operational responsibility, gives them particular weight for the Recommendations section.

5.6 Methodological Soundness

Primary evaluation question: How conceptually and methodologically sound is the ABRS model as a practical tool for operationalizing localized solutions, and what are its strengths and gaps for adaptation elsewhere?

Finding 5.6.1 - The community-engagement methodology is a genuine, well-documented strength of the ABRS model. [Medium-High confidence - a desk-review documentary finding, independently corroborated by this evaluation's own MSC data]

The desk review traces a clear evidence chain from 114 logged community-identified issues to a validated, resourced project portfolio, calling this "methodologically the strongest-documented part of the pilot" - an assessment this evaluation finds credible, though it rests on the desk review's documents rather than this evaluation's own data collection. The same documentation shows signs of authentic, not managed, ownership: participants self-organized coordination outside formal sessions, and the validation step recorded live disagreement and methodology harmonization rather than pro-forma consensus. This is independently corroborated by this evaluation's own MSC data: one respondent, reflecting on his cash-for-work role in the lighting works, urged that residents be treated as genuine co-planners rather than only as labour, arguing this produces better-targeted, more sustainable projects and stronger community ownership (MSC narrative 1, host-community respondent). The national/coordination-level informants separately described a recurring phenomenon of "assessment fatigue," in which community members had grown weary of repeated needs assessments and explicitly wanted to become active contributors to recovery rather than being assessed again - direct evidence of authentic community demand for the participatory model this finding describes, from a source independent of the desk review and the MSC data above (National/coordination-level KII).

Finding 5.6.2 - There is a structural gap between the pilot's human-security ambition and what its M&E framework actually measures. [High confidence]

The desk review finds, and this evaluation concurs, that the indicator structure is applied consistently but consists almost entirely of output/reach counts (people served, platforms attended) rather than outcome-level measures of the human-security resilience or durable-solutions change the Theory of Change describes - terms not currently operationalized into any indicator in the project's results framework. Where the results framework does include outcome-level indicators, these partially contribute to the Theory of Change without fully demonstrating the human-security and durable-solutions outcomes it envisions, reflecting this pilot's scope, duration and implementation conditions. No document confirms a budgeted mid-point review took place, and monitoring tracks physical progress with no visible link to burn-rate or cost-per-beneficiary data. This is the evaluation's single most important methodological-soundness finding: given the pilot's short duration and small scale, this gap is best read as a scale-up requirement (see Recommendation R1) rather than a design flaw in this project's own indicators; it does not, however, limit this evaluation's own ability to make outcome-level claims, since the evaluation draws independently on the household survey, FGDs, KIIs and MSC narratives for that purpose (Findings 5.3.2, 5.8.1).

Finding 5.6.3 - The participatory-planning tool's transferability is evidenced by documented replication of specific components to date; views on further adaptation are treated as lessons learned and good practices. [Medium-High confidence in the documented adaptation pattern; Medium confidence, single-source, in the caveats on adapting it further]

This is not the model's first transfer: its governance structure (the Law 107 volunteer-committee model) was adapted from a 2022-2025 precedent in an adjacent district (Kadi Askar), and the Local Participatory Platform's own Terms of Reference reference an earlier pilot in Latakia, Damascus and Homs - consistent with the ongoing replication into eight further governorates corroborated at Finding 5.7.1. Two senior informants added single-source caveats on how, not whether, to adapt it further: one noted the tool's social-cohesion mapping needs linkage to actual investment opportunities, and that Syria's communal composition means it *"may not translate directly"* elsewhere (*Senior KII, UNDP regional advisor*); the other named community-representation and government-convening arrangements as most replicable, while flagging the missing feedback loop from community consultation to government prioritization as a scale-up gap (*Senior KII, UN global coordinator*).

5.7 Sustainability and Scale-Up Potential

Primary evaluation question: What systems, capacities or partnerships are in place to sustain ABRS's benefits, and what are the enabling or limiting factors for adaptation and scale-up?

Finding 5.7.1 - Strategic-level sustainability and scale-up prospects are strong; site-specific sustainability is genuinely uncertain. [High confidence]

At the strategic level, ABRS has been formally adopted as Syria's national localized-solutions framework and is scaling to eight further locations, with mainstreaming into major UN planning instruments under way. At the Bab Al-Nairab-specific level, however, three risks are documented by implementers themselves, not inferred by the evaluation: (i) the local implementing unit's own final report states that no alternate government team had been appointed, following a governorate-level leadership transition, to carry the 90-project community recovery plan forward into an actual government reconstruction plan- though coordination with Government counterparts on the plan has continued since project close, with the Government's International Cooperation Directorate increasingly adopting area-based, community-prioritized approaches; (ii) the sewage rehabilitation's own final report states that additional funding is "urgently needed" to complete the wider network, meaning only a partial segment was addressed; and (iii) post-handover vocational training start date at the rehabilitated school/VTC has slipped past its scheduled February 2026 date, though this reflects an active, ongoing handover process rather than an absence of follow-up: UNHCR continued to monitor the school's condition post-handover and, in coordination with the Aleppo Directorate of Education, scheduled the contractor's corrective works for the summer holiday period specifically so as not to disrupt students, under the standard one-year defects-liability period. *"Are the authorities you are dealing with today going to be there tomorrow? For decisions to be not only sustainable but to stand the test of time, the infrastructure of governance needs to be strong and sustainable. We are still in the unknown."* (*Senior KII, UNHCR field management*)

Finding 5.7.2 - Participatory-planning capacity, not physical infrastructure, is what stakeholders across every level most consistently expect to endure. [High confidence]

This is one of the most convergent findings in the entire evidence base: field-level KII informants, thematic KII informants, and the desk review's own assessment of the community-engagement pathway's documentary strength all independently identify the participatory-planning relationships and governance capacity - not the rehabilitated infrastructure itself - as the pilot's most durable legacy. *"ABRS is a blueprint on which we can build upon", while explicitly rejecting any suggestion that Bab Al-Nairab has reached a stable end-state ("by no means is this the perfect community where things are back to normal").* (Senior KII, UNHCR headquarters)

Finding 5.7.3 - Funding architecture is the single most convergent structural constraint on sustainability and scale-up identified anywhere in this evaluation. [High confidence]

Independently and without prompting from one another, multiple senior informants across UNDP, UNHCR, and donor institutions converge on the same structural diagnosis: short, single-year, humanitarian-style funding cycles are poorly matched to the multi-year, multi-sectoral nature of area-based recovery. This is treated as a robust, load-bearing finding precisely because it emerged independently across institutionally distinct respondents rather than being volunteered in response to a leading question. *"This cannot be donor-driven alone."* (Senior KII, UNDP regional advisor)

Finding 5.7.4 - Syria's current governance architecture presents a genuine, convergently identified constraint on scale-up that is substantially exogenous to ABRS's own design. [Medium-High confidence]

Two senior informants from different institutions (a UNDP Syria-based informant and a global-level UN informant), reasoning from different starting points, independently reach the same structural conclusion: highly centralized, still-forming decision-making (decision-making remaining highly centralized, with limited devolution to sub-national level) constrains the decentralized, municipal-level engagement the ABRS model requires for scale-up. The evaluation treats this convergence as a robust finding for the sustainability narrative, while noting a genuine divergence among senior informants on how ready the Syrian government actually is for a scale-up conversation - a divergence plausibly reflecting the fast-moving nature of the political context rather than a factual disagreement to be resolved. This constraint is compounded by the government's own evolving return policy: one senior informant referenced the government's *"no camp, no tent"* policy framing directly as the backdrop against which scale-up conversations are now occurring (Senior KII, UN global coordinator).

Finding 5.7.5 - Fine-grained community-level demand for continuity is strong and specific, and centres on maintenance more than on further construction. [High confidence]

FGD recommendations converge strongly on establishing funded, locally staffed maintenance arrangements - not further one-off construction - as the top sustainability priority, citing specific, concrete failure examples within the project's scope (a solar-lighting inverter with no local spare-parts or repair capacity; recurring sewage blockages with no servicing schedule). Community members also raised a discontinued waste-collection service, which falls outside this project's direct scope and

attribution; it is noted here as context for the community's broader expectation of continuity and clearer communication of intervention boundaries (see Recommendation R4), rather than as evidence for this finding. This is corroborated by MSC narrative evidence, in which both documented Most Significant Change stories independently and unprompted identify dedicated maintenance budgets and rapid-response repair mechanisms as the single condition needed to sustain their described change. *"If the inverter on the solar panels fails, we go back to the darkness we used to live in, because there is no maintenance and no spare parts."* (FGD, IDP man) *"There must be dedicated, permanent resources and budget for maintaining the lighting and sewage networks, not just during project implementation... [and] a system with clear mechanisms for managing and maintaining public facilities, encouraging residents to report any faults."* (MSC narrative 1, host-community respondent)

5.8 Impact

Primary evaluation question: To what extent has the pilot contributed to improved safety, social cohesion and reintegration, and are there unintended effects?²

Finding 5.8.1 - The pilot has made a credible, if modest and unevenly distributed, contribution to human-security conditions- constituting evidence toward the ABRS Theory of Change's broader human-security outcomes, albeit necessarily limited by this pilot's scope and duration. [Medium-High confidence]

The household survey shows a broadly positive pattern of self-reported change in safety/dignity (72% improved) and social cohesion (62% better), and the most senior, most carefully hedged informant in the evidence base - explicitly declining to overclaim - describes the pilot as having made *"a practical and credible contribution to conditions for safety, dignity, and social cohesion, rather than claiming that it solved the probably deeper social or structural tensions in the area."* (Senior KII, UNHCR field management) This aggregate positive picture, which carries greater evidentiary weight given the household survey's much larger base (n=204), carries a smaller, complementary qualitative caveat from the field-level KIIs that the survey's top-line percentages do not capture: two general-KII informants who separately reported improved social cohesion also explicitly stated, in the same answer, that their own sense of safety specifically had not changed - a reminder that cohesion and safety can move independently of one another, and that this evaluation's positive cohesion finding should not be read as automatically implying a matching improvement in safety at the same sites (Field KII).

² Consistent with the evaluation's contribution-analysis approach, findings in this section describe plausible contribution to observed change, not causal attribution.

Finding 5.8.2 - The most analytically significant impact-related observation highlights identity and integration beyond service delivery. [Medium confidence - single strong source]

A senior informant observed that population labels (IDP/returnee/host community) *"disappear"* in practice once people are integrating into shared daily life, because everyone faces *"the same needs of the day"*, with the ABRS approach specifically valuable because it resolves the tension between mandate-based agency programming (organized around these labels) and lived community reality (*Senior KII, UNHCR field management*). While this observation comes from a single source, its analytical depth and direct relevance to the evaluation's human-security and Durable Solutions framing lead the evaluation to treat it as a genuine, citable insight - while flagging, transparently, that no other respondent made this precise empirical observation, though it should still be tested further, e.g. a validation workshop or session. A normative echo of this same underlying logic - though not an identical empirical claim, and so treated here as convergent support rather than direct replication - comes independently from the national/coordination-level KIIs: one of these informants recommended that the ABRS approach remain area-based rather than status-based specifically so that interventions respond to the needs of all residents rather than focusing on any single population group, a design principle that only makes sense if status-based labels matter less operationally than shared geography and need - the same reasoning this finding draws from lived experience (National/coordination-level KII).

Finding 5.8.3 - Unintended effects were reported and were mostly, but not exclusively, positive. [Medium confidence]

Positive effects reported by field-level KII informants include shop reopenings, increased return movement, and increased social trust. Negative or ambiguous unintended effects include noise/nuisance from workshops sited within residential areas, erosion of community trust and credibility where promised interventions were delayed or seen as too limited relative to need. The physical safety incidents documented through direct observation (Finding 5.3.7) occurred at Al-Marjeh, one of the project's own sanitation sites. The implementing team has indicated that the specific pipework involved may not be part of the project's rehabilitated segments, though this could not be independently verified. If confirmed, the relevant unintended effect would be reputational: communities do not always distinguish between infrastructure under different contracts or by different contractors within the same site, so faults in adjacent segments can still be attributed to the project. The finding is retained as directly observed. The evaluation notes that "unintended effects" in this evidence base skew toward process-level community-relations effects rather than the kind of major unintended consequences (e.g., exacerbating displacement-status tensions) that a more adversarial context might have produced - itself a modestly positive finding about the pilot's conflict-sensitivity design, though one that should not be over-read given the modest scale of the evidence base.

5.9 Cross-Cutting Issues

Human Rights. [Medium-High confidence] The pilot shows genuine substantive rights-based framing, not boilerplate: HLP and civil-documentation assistance is explicitly and consistently linked by senior and field-level informants alike to beneficiaries' ability to access services, register family events, resolve property disputes, avoid exploitation, and establish a stable basis for reintegration. Legal-assistance beneficiaries in the household survey report high confidentiality (89% "always") and resolution rates (74%). At the same time, systemic HLP/legal reform and civil-documentation backlogs are explicitly and consistently described, by the same informants, as remaining unresolved at the national level regardless of individual-level gains achieved through the pilot - an important scope clarification for how "rights" claims in this evaluation should be read.

Gender Equality. [Medium-High confidence on program-level gender integration; Medium confidence specifically on survey-based gender claims, given the sample's gender skew] Legal-assistance beneficiaries skew female (the desk review reports roughly 61% in its reviewed disaggregation, exceeding program expectations), and FGD evidence documents a genuine trajectory of increasing women's participation over the life of the community-engagement process (from no women attending initial meetings to roughly half of participants by the end, independently corroborated by two separate KIIs - a general/field-level informant and a national/coordination-level informant). However, LPP membership (41 of 81, approximately 51% women) fell short of the proposal's 65% target, and - most importantly for this evaluation's own confidence in gender-related claims - **the household survey sample itself is 79% male**, a substantial deviation from the Inception Report's 50-60% female target that materially limits how much weight quantitative, survey-based gender claims in this report can bear. A senior informant's own candid, unprompted acknowledgment is treated as a credible, reliable data point precisely because it was volunteered rather than extracted: *"Can I say that full representation exists within this program? No, I cannot. We tried to ensure that vulnerable groups were represented. But it has to be voluntary."* (Senior KII, UNHCR headquarters) Evidence gap identified: no respondent across any KII track addressed whether observable improvements in decision-making access for women have occurred, as distinct from participation/attendance figures at the input level. A verification cross-check of the household survey's gender disaggregation (undertaken after the first draft of this report) surfaced one further point that had not previously been reported: among the small female sub-sample (n=43), awareness of the feedback/complaints mechanism is markedly higher than among men - 77% of women (33 of 43) versus 47% of men (75 of 161) know where to give feedback or complaints. Read together with the survey's own 79%-male skew, this could indicate either a genuine gender difference in mechanism awareness or a composition effect (e.g., women respondents concentrated among more-engaged households); the small female base size (n=43) means this should be treated as indicative rather than conclusive, and is flagged here for triangulation at the validation workshop rather than as a standalone finding. Separately, one general-KII informant recalled an explicit 50% participation target having been applied for women and other vulnerable groups during community-engagement activities - a distinct, field-level data point on target-setting practice, additional to the LPP's 65% target and the survey's 50-60% target discussed above, though it could not be corroborated against a documentary source (General KII). Taken together, the LPP (51% women), legal-assistance (61% women) and

feedback-mechanism-awareness (77% women) figures point to meaningful attention to gender equality within the program, alongside the specific gaps identified above.

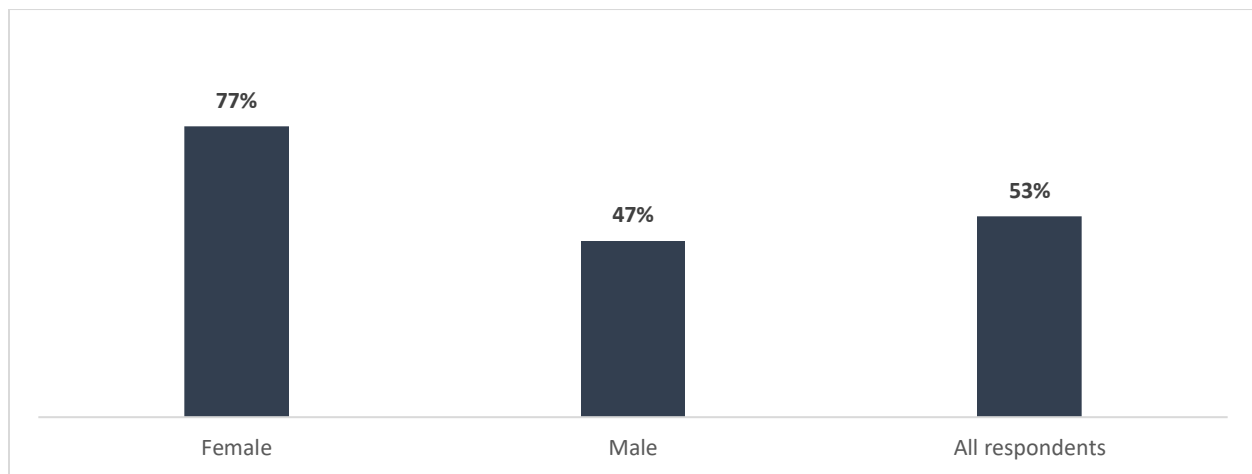


Figure 12: Awareness of Feedback/Complaints Mechanism, by Gender

Disability Inclusion. [Low confidence - this is itself a finding] This is the thinnest cross-cutting theme across every data source reviewed for this evaluation, despite being explicitly probed in multiple interview guides. Household survey disability representation (8%, n=16) sits within the Inception Report's 3–15% target range, and LPP disability representation is estimated at 5–8 of approximately 90 representatives (a senior informant's estimate, explicitly imprecise) against a reported target of 3%, actually achieved figures around 0.3% in some legal-assistance disaggregation per the desk review - a further internal inconsistency. Qualitatively, one FGD respondent with a physical disability directly and pointedly raised unmet accessibility needs: *"As a vulnerable person with a disability - even counting me as a resident - my needs are different from everyone else's: I need help moving around the pavements and facilities, and this is a demand that everyone overlooks."* (FGD, man with a physical disability) Observation checklists included accessibility-for-persons-with-disabilities as a structured field at every site visited, but the underlying checklist responses on this dimension were not extractable in a form that allows this evaluation to report a specific finding.

Environmental Considerations. [Low confidence] No dedicated environmental-assessment document, environmental indicator, or environment-specific interview content was identified anywhere in the evidence base made available to this evaluation, beyond the technical (non-environmental) concern raised by a field-level engineer that any expanded sanitation network requires downstream treatment or collection infrastructure to avoid future flooding at outfall points. *"We need to start with sanitation and plan for it before people return, and we also need to study where the network ends - there needs to be a treatment plant or collection point."* (General, sanitation company engineer)

Protection, Safeguarding and Data Protection. [High confidence] Protection mainstreaming shows genuine substantive engagement (e.g., the VTC site relocation decision made specifically on protection grounds; conflict-sensitivity analysis shaping program design from the outset).

6. CONCLUSIONS

6.1 Strategic Interpretation

The evidence assembled for this evaluation supports a clear strategic interpretation: **The ABRS in Bab Al-Nairab succeeded as a proof of concept for an area-based, human-security-centred model of return support, and this - rather than comprehensive service delivery to Bab Al-Nairab's full range of needs - is the correct standard against which its performance should be judged.** This interpretation is not the evaluation team's own construction; it is the interpretation independently and repeatedly volunteered by the pilot's own designers, implementers, and senior oversight stakeholders across every KII reviewed, and it is consistent with the pilot's modest budget (on the order of USD 775,000–800,000, depending on source) relative to a target area of well over 100,000 people. This distinction was put most directly by one of the national/coordination-level informants interviewed for this evaluation, who described the pilot as having delivered "100% of what we intended to deliver," while explicitly acknowledging that this represented only a small proportion of the overall needs in Bab Al-Nairab (National/coordination-level KII).

Judged against that standard, three conclusions follow. First, **the model itself is sound:** the combination of legal assistance, participatory governance, and infrastructure rehabilitation, delivered through an area-based rather than sector-based or status-based lens, addresses a genuinely interlocking set of human-security deficits that a conventional single-sector program would not have addressed as coherently. The evaluation's evidence for this is unusually strong and convergent across every data source, from the desk review's baseline triangulation to the household survey's positive outcome indicators to senior stakeholders' independent strategic endorsement and is directly grounded in the findings section.

Second, **the pilot's execution reveals a consistent pattern that recurs across nearly every evaluation criterion:** the further a stakeholder or data source sits from day-to-day delivery, the more uniformly positive their assessment; the closer to delivery, the more specific, textured, and mixed the picture becomes. This is not a contradiction to be resolved by picking one altitude as "correct" - both are valid and both are necessary for an accurate picture - but it does mean that senior-level, strategic endorsement of ABRS (which is genuine and well-founded) should not be read as evidence that implementation-level challenges documented in this report (coverage variation, infrastructure quality-assurance gaps, gatekeeping risks, uneven returnee outcomes) are minor or resolved. This same pattern was volunteered independently by the national/coordination-level KIIs, whose respondents distinguished strategic efficiency (coordination, institutional learning, and long-term planning value, which they rated highly) from operational efficiency (implementation speed and complexity, which they rated as slower and more constrained than anticipated) - a third, independent articulation of the altitude effect described above (National/coordination-level KII).

Third, **the most consequential constraints on ABRS's future are substantially exogenous to the pilot's own design:** Syria's funding architecture (short, fragmented, single-donor cycles) and governance architecture (centralized, still-forming, subject to rapid political change) were independently and

convergently identified by multiple senior stakeholders - speaking from institutionally distinct vantage points and without prompting from one another - as the primary structural constraints on scale-up. This does not exempt UNDP and UNHCR from responsibility for addressing what is within their control (the M&E gap in Finding 5.6.2; the infrastructure quality-assurance gaps in Finding 5.3.7; the coordination frictions in Finding 5.5.1), but it does mean that scale-up recommendations must be realistic about what a well-executed ABRS model can and cannot overcome on its own.

6.2 Positioning ABRS as a Model

The evaluation concurs with the strategic case, made independently by multiple senior stakeholders, that ABRS's greatest value may lie in its function as a transferable methodology - particularly its community-engagement pathway (Finding 5.6.1) and its institutional arrangement for bringing government into a convening role alongside community-identified priorities - rather than in the specific physical outputs delivered in Bab Al-Nairab. This positioning is credible and well evidenced. It should, however, be communicated to donors and government counterparts with the same honesty about scale and limitation that the pilot's own senior stakeholders showed in their interviews: as one senior informant put it, the risk of treating ABRS as a proof of concept is that if it is merely "sprinkled" thinly across many sites without addressing the funding and governance constraints identified in Section 5.7, it will fail to demonstrate the very scalability it is meant to prove. This framing is independently corroborated by several national/coordination-level informants, who separately described the pilot as catalytic, explaining that its primary objective was to establish a methodology capable of attracting future investment rather than to address all needs directly - the same positioning argument reached independently from a distinct institutional vantage point (National/coordination-level KII). A field-level informant's recommendation adds a concrete, dissenting counterpoint to this same risk from an entirely independent vantage point: rather than expanding to additional governorates, this informant argued for concentrating future programming on a single, clearly defined geographic area, explicitly to avoid diluting resources and impact - the same underlying concern about thin, ineffective replication expressed by the senior informant above, reached independently from the opposite direction (expand less, not more) (General KII).

6.3 Critical Success Factors and Constraints

Success factors (evidenced): a genuinely area-based rather than sector-based or status-based design logic; a methodologically strong, traceable community-engagement pathway; complementary UNDP/UNHCR mandates (protection/legal and early-recovery/infrastructure) that, while not merged in delivery, achieved real alignment of priorities; strong strategic-level coordination architecture (Solutions Working Group, AHCT); and a legal-assistance pillar that performed strongly and consistently across every data source.

Constraints (evidenced): an M&E system that cannot itself demonstrate the human-security outcomes the program claims to pursue; uneven, geographically patchy delivery within the pilot area that tracks completed interventions more than any other factor; documented infrastructure quality-assurance and transparency gaps, including safety incidents during construction; unresolved local institutional

continuity (no confirmed government counterpart team for the community recovery plan); a funding architecture mismatched to the multi-year nature of area-based recovery; and a governance environment in transition that constrains decentralized engagement regardless of program quality.

7. RECOMMENDATIONS

7.1 Strategic Recommendations

R1. UNDP and UNHCR Country Offices, with the Solutions Working Group, should commission and resource a harmonized, outcome-level MEAL framework for ABRS as a national framework before further scale-up decisions are finalized. *Linked to Finding 5.6.2.* The current indicator set measures reach and outputs, not the human-security or Durable Solutions change the model claims to pursue; this is a requirement for the next phase of ABRS replication and scale-up, not a precondition the original pilot itself should have met.

R2. UNDP, UNHCR and UNTFHS/JSB should jointly pursue multi-year, pooled, or blended financing architecture for ABRS sites, rather than continuing to rely on 12-month, single-donor funding cycles. *Linked to Finding 5.7.3.* This is the single most convergent structural finding in this evaluation, independently identified by informants across UNDP, UNHCR and the donor institution itself. Options to explore include multi-year HSTF windows, blended development finance, and - as one senior informant specifically suggested - packaging ABRS with other UNDP tools (stabilization, governance, HLP) that donors may be more willing to fund as a bundled, multi-year line than as a standalone humanitarian-style grant.

R3. UNDP and UNHCR should engage government counterparts on a realistic, phased scale-up conversation that is explicit about current governance constraints, rather than treating scale-up readiness as settled. *Linked to Finding 5.7.4.* Given the convergent, independent finding that Syria's current governance architecture is highly centralized and still forming, scale-up planning should build in an explicit assumption-testing step on government readiness and decision-making capacity at each proposed new site, to assess context-specific readiness and adapt the approach before replication or scale-up.

7.2 Operational Recommendations

R4. UNDP/UNHCR and implementing partners should reinforce existing warranty and institutional-ownership arrangements: ensuring local authorities and community committees are properly equipped, briefed and resourced to assume operations-and-maintenance responsibility at handover, in line with the defects-liability period already in place, and mitigate the risk of end-user misuse of rehabilitated infrastructure or equipment, as observed during this evaluation. Because communities may not always distinguish between infrastructure delivered under the pilot's different components, agencies, the Government or other actors, problems affecting interconnected infrastructure - such as sewage segments outside the pilot's rehabilitation scope - may still be attributed to the pilot as a whole (Finding 5.8.3). *Linked to Findings 5.3.7 and 5.7.5.* This is the most consistently and specifically requested community-level recommendation across FGDs and MSC narratives alike, and directly addresses the risk that early, genuine gains (Finding 5.3.2) erode for lack of spare parts, technicians, or a servicing schedule - by strengthening handover and warranty-period follow-through with the responsible institutions, not by creating a parallel structure.

R5. UNDP and UNHCR should strengthen quality-assurance and reporting-transparency requirements for infrastructure contractors, specifically requiring narrative disclosure of quantity/cost variances above a defined threshold and independent verification of "no challenges" progress-report language.

Linked to Finding 5.3.7 and 5.4.1. The convergence of an undisclosed approximately 285% quantity variance on the UNHCR-financed VET civil-works item, distinct from the UNDP-financed sewage component, an unresolved snag list, and independently observed workmanship and safety defects (including two reported construction-related injuries) at the same site is a pattern that current contractor self-reporting did not surface; the evaluation team's own field observation did. The implementing team has indicated the safety defects relate to sections outside this project's rehabilitated scope; this has not been independently verified, and the recommendation's basis - the gap between reported and observed conditions - stands regardless.

R6. UNDP, UNHCR and local implementing partners should confirm, formally and in writing, a government or local-authority counterpart team to carry the validated 90-project community recovery plan forward, and report on this at the next reporting milestone. *Linked to Finding 5.7.1.* This was flagged as unresolved in the implementing unit's own final report and is a direct precondition for the sustainability claim attached to the participatory-planning pillar.

R7. UNDP and UNHCR should reduce reliance on a single local-authority gatekeeper or mukhtar for community and institutional access, and continue strengthening the public-announcement, self-nomination-based LPP participant selection process already in place, to reduce gatekeeping risk.

Linked to Finding 5.3.5. The perception of mukhtar-mediated gatekeeping is corroborated independently from both the community (FGD) and implementer (KII) sides; notwithstanding the self-nomination safeguards already built into the design; this residual, perception-based risk is a specific, actionable basis for continued process reinforcement, rather than a generic inclusion recommendation.

R8. UNDP and UNHCR should strengthen the accountability/complaints mechanism and its visibility.

Linked to Finding 5.3.2 (household survey). Only 53% of survey respondents knew where to give feedback or complaints; this figure fell to as low as 11% among respondents with no recorded program service, indicating the awareness gap is worst precisely among those least engaged with the program - a specific, addressable finding for community outreach and signage (also flagged by observation as absent at the VTC site).

7.3 Programming Recommendations (including scale-up implications)

R9. Future ABRS-model programming should explicitly add or link to a housing/shelter-repair component, or clearly and transparently communicate its exclusion from scope. *Linked to Finding 5.1.2.* This is the most consistently identified, cross-category relevance gap in the evidence base; if genuinely out of scope for resource reasons, this should be stated proactively to manage community expectations rather than left as an implicit gap communities discover themselves.

R10. Future program design should build in from the outset the specific coordination improvements volunteered by the most operationally engaged senior stakeholder: an explicit joint implementation plan agreed at project outset; a shared, jointly owned action tracker; a common area-level outcome

results framework (not only an activity framework); more systematic joint monitoring; and predictable, flexible multi-agency financing arrangements. *Linked to Finding 5.5.3.*

R11. Future phases and scale-up sites should strengthen and better mainstream gender- and disability-inclusion data collection, including ensuring survey samples meet planned gender-stratification targets, collecting disaggregated outcome (not only participation) data for women, and adding a dedicated, adequately resourced qualitative data-collection stream for persons with disabilities. *Linked to the Cross-Cutting Issues findings in Section 5.9.* This is as much a recommendation to UNDP/UNHCR's program M&E function as to any future evaluation.

R12. Scale-up sites should build a structured, resourced sustainability handover protocol into project design from the outset - covering infrastructure operations and maintenance, government counterpart confirmation, and continuity of LPP structures - rather than addressing sustainability as an end-of-project activity. *Linked to Finding 5.7.1.*

8. LESSONS LEARNED

1. **An area-based, not sector-based or status-based, design logic is a genuine methodological strength for return contexts where population labels blur in lived practice.** The finding that IDP/returnee/host-community distinctions "disappear" once people are integrating (5.8.2) suggests future programming should be designed and monitored around geography and shared need, while still disaggregating data by status for equity monitoring - the two are not in tension, but require deliberate design.
2. **Community-engagement methodology quality does not automatically translate into evidence of outcome-level change; both need to be resourced deliberately.** The MRS, co-design, validation pathway (5.6.1) is a genuinely strong, replicable process. The M&E gap (5.6.2) is a separate, correctable gap - best read as a scale-up requirement rather than a problem with this pilot itself - that should not be allowed to undercut confidence in the participatory methodology itself.
3. **"Joint" agency implementation delivers real value through alignment of priorities, not merged delivery systems - future joint programs should plan for this explicitly rather than aspire to full integration.** The convergent finding in 5.2.2 suggests future joint UN program design documents should describe joint implementation accurately (aligned, complementary delivery under separate systems) rather than aspirationally, to set realistic expectations for donors and evaluators alike.
4. **Coverage variation within a single pilot area can be as consequential to community trust as coverage across areas - sequencing and communication matter as much as eventual completion.** Where interventions are phased, proactively communicating what is planned for later phases (5.3.3) may materially reduce the credibility and trust costs documented in FGD and KII data alike.
5. **Local informal power structures (mukhtars) are a predictable and specific risk to participatory processes in this context, and should be designed against from the outset,** a risk the program already designed against at the selection stage through a public, transparent participant-announcement process in which mukhtars held no formal nomination role. The residual risk lies instead in the participatory planning discussions themselves, where communities perceived mukhtars as influential; this should be further strengthened through measures such as rotating or multiple community-liaison roles (5.3.5).
6. **Infrastructure contractor self-reporting is not a reliable sole source of quality assurance in this operating environment; alongside the site supervision, defect-liability periods and post-warranty final acceptance processes already in place; independent field observation nonetheless surfaced defects and a safety incident that contractor progress reports did not.** Future ABRS-model infrastructure components should build in independent, structured observation over time as a standing quality-assurance function, rather than something that surfaces only through a periodic external evaluation exercise (5.3.7).
7. **Funding architecture, not implementation quality, is the binding constraint on multi-year area-based recovery in the current Syria context.** This lesson, while outside implementers' direct

control, should inform how UNDP, UNHCR and donors jointly design the next generation of ABRS financing instruments from the outset of scale-up planning (5.7.3).

