REQUEST FOR PROPOSAL (RFP) FOR AN ENDLINE EVALUATION OF THE AKSHITA PROGRAM'S INTEGRATED MODEL – DAMOH DISTRICT

Issued by: The Antara Foundation
RFP Issue Date: 9 September 2026
Deadline to Receive Questions from Prospective Applicants: 12 September 2026
Deadline for Response to Questions (RFQ): 16 September 2026
Proposal Submission Deadline: 25 September 2026
Selection of Study Partner: 8 October 2026

1. ABOUT THE ANTARA FOUNDATION

The Antara Foundation (TAF) envisions a world where women and children do not suffer from preventable health conditions. We believe that every mother and every child deserves an equal start to a healthy life. We support the public health system in delivering solutions at scale to improve maternal and child health outcomes by partnering with the government and communities. Interventions of The Antara Foundation follow two closely integrated paths: helping the government system deliver higher-quality health care to the community, and supporting the community in mobilising and seeking better health outcomes.

TAF works to improve maternal, newborn, and child health and nutrition (MNCHN) outcomes across nine districts of Madhya Pradesh by strengthening last-mile public health delivery, anchored by its AAA Platform (coordinating ASHAs, ANMs, and Anganwadi Workers), supervisory strengthening, and facility enhancement.

For more information visit antarafoundation.org.

2. PURPOSE OF THIS RFP

TAF invites proposals from qualified research agencies (hereafter referred to as the study partner) to conduct an endline evaluation of the Akshita Program's integrated model in Damoh district, Madhya Pradesh.

In Damoh, TAF works across five blocks (Batiyagarh, Hatta, Jabera, Patera, and Tendukheda), delivering both supply-side and demand-side interventions. Initiated in 2022, supply-side interventions include the AAA Platform[1], facility-based care strengthening, and systems strengthening. Demand-side, or community-facing demand-generation work comprised two different models: Participatory Learning and Action (PLA), rolled out in November 2024 in the Jabera and Batiyagarh blocks, and Mothers and Enablers Groups (MEGs), rolled out in September 2025 in the Patera, Hatta, and Tendukheda blocks.

The baseline itself was conducted in September–October 2024, prior to both community models were rolled out. This design nests two comparisons: a cluster-randomised allocation of the demand-side layer across clusters within Damoh (integrated arm vs. supply-only arm), combined with a quasi-experimental, non-randomised comparison against Panna district (with no TAF interventions). TAF is now commissioning an endline assessment, expected to be fielded approximately two years after baseline

[1] AAA: ASHA–ANM–Anganwadi Worker coordination platform, TAF flagship supply-side intervention model.

(around September–October 2026), to complete this impact evaluation. Full background, baseline findings, methodology, and measurement domains are provided in Annex 1.

This RFP seeks a study partner with demonstrated experience in:

- Designing and conducting quasi-experimental / c-RCT impact evaluations in MNCHN, or closely related public health domains, in India
- Difference-in-differences (DiD) or comparable causal-inference analysis
- Large-scale CAPI-based household surveys in rural/tribal geographies, including managing migration-related fieldwork challenges
- Qualitative research grounded in implementation science or process evaluation
- Institutional Review Board (IRB) approval, mandatory (see Section 8)

This is an open, competitive RFP. Agencies that have previously worked with TAF are welcome to bid alongside other qualified agencies. Researchers/organisations applying must be registered in India.

3. SCOPE OF WORK

The study partner is expected to execute the endline round consistent with the baseline's evaluation design (see Annex 1, Sections II–IV for detailed objectives, evaluation design, scope of work, and measurement domains), flagging any deviations required by changed ground conditions.

- TAF will facilitate access to the baseline survey tool, the January 2025 baseline report, the de-identified baseline dataset, and government permissions; these materials, along with any other tools and de-identified datasets, will be shared only with the selected study partner. The study partner will be responsible for adapting the instruments, and for fielding, analysing, and reporting the endline round.
- **Key deliverables** will include an inception report, IRB approval documentation, cleaned quantitative and qualitative datasets, a draft endline report, a final endline report, and a presentation of findings to TAF (see Annex 1, Section VI).

Key Research Questions

The endline is anchored around the following overarching research questions:

- **RQ1a.** What is the incremental effect of the Akshita program's integrated approach (supply-side + PLA/MEG) over the supply-only approach, on knowledge, practices, and service access related to MNCHN?
- **RQ1b (secondary, exploratory).** Does the incremental effect differ between blocks receiving PLA versus blocks receiving MEG? This is exploratory, not causal.[2]
- **RQ2.** What is the overall effect of the Akshita program on (a) key MNCHN health outcomes and (b) knowledge, practices, and service access, relative to the comparison arm (Panna district, with no TAF interventions)?[3]
- **RQ3.** If and how did the integrated model's community engagement components (PLA and MEG) contribute to the effects seen above, and if so, how?
- **RQ4a.** How were the PLA and MEG components of the integrated model adopted, implemented, and experienced on the ground in their respective blocks, including implementation fidelity, reach, and variation in delivery?

[2] PLA vs. MEG assignment follows block rather than random cluster assignment, was not powered to detect differences, and the two interventions also differ in exposure duration by endline. Agencies must construct this exposure-duration difference as a covariate; see Section III.A.

[3] This is a quasi-experimental comparison; its causal interpretation depends on the parallel-trends assumption, which this design cannot fully test (see Section I.D).

- **RQ4b.** What were TAF's design models and implementation plans for the integrated model, how were these adapted along the way (e.g., the phased PLA-then-MEG rollout), and what lessons do these offer? (See Section III.C.)

4. TIMELINE

The indicative phase-wise timeline is below; the study partner should propose a detailed workplan against these phases as part of the technical proposal.

Table 1. Indicative Timeline

Phase	Indicative period
Inception, tool finalisation, IRB submission	October 2026
IRB approval, sampling frame update, enumerator training	October-November 2026
Quantitative and qualitative fieldwork (concurrent)	December 2026
Data cleaning and analysis	January 2027
Draft report and triangulation workshop	January-February 2027
Final report and presentation to TAF	February 2027

5. PRICING STRUCTURE AND BUDGET

This is a competitive bid process. TAF will not be disclosing an indicative budget ceiling upfront. Agencies are requested to propose a realistic, itemised costing for the study based on the scope of work described in this RFP and Annex 1, clearly separating quantitative and qualitative components.

Payments will be made based on acceptance of deliverables (Section 3 and Annex 1, Section VI), against the indicative milestones below; final percentages and amounts will be confirmed at the time of contract signing.

Inception Report accepted: 20%

- IRB Approval Documentation accepted: 10%
- Cleaned quantitative and qualitative datasets accepted: 20%
- Draft Endline Report accepted: 30%
- Final Endline Report and presentation to TAF accepted: 20%

6. PROPOSAL SUBMISSION REQUIREMENTS

Interested agencies should submit the following as separate documents:

- **A. Technical Proposal (≤15 pages):**
  - Understanding of the assignment including key RQs
  - Proposed methodology including the analytical approach:
    - For the DiD/regression estimation, including the covariate-adjustment strategy (Annex 1, Section I.D)
    - To address recall-bias mitigation, non-response tracking, and baseline–endline comparability (Annex 1, Section I.D)
    - For inclusion of themes around MNCHN-related agency in the endline (Annex 1, Section III.B)

        - Qualitative data collection and analysis, including triangulation with quantitative findings (Annex 1, Section III.C)
    - Workplan and timeline
    - Team composition and roles (including CVs of key personnel in the annexures)
    - Quality assurance plan
        - Data collection management plan, including tool finalisation, enumerator training and certification, the consenting process, and field supervision arrangements
        - Fieldwork risk mitigation and contingency plan (e.g., adverse weather, access constraints, staff attrition, local disruptions)
- **B. Capability Statement (≤2 pages):** highlighting relevant past experience, with references from at least two comparable completed evaluations in the last five years.
- **C. Financial Proposal (≤3 pages), itemised by phase/deliverable:**
    - Personnel: Daily rates and number of days for key staff (biostatisticians/econometricians, qualitative researchers, field team leads, etc.)
    - Fieldwork/data collection: Enumerator wages, qualitative researcher costs, training costs, quantitative and qualitative components shown separately
    - Travel & logistics: Accommodation, local transport, and domestic travel
    - Data management and analysis costs
    - Overheads/management fee: Any administrative percentage charged by the agency
    - Taxation clarity: Base costs and any GST shown separately
- **D.** registration and legal documents to be enclosed:
    - Copy of registration certificate of the company/firm
    - Copy of PAN and GST
    - Bank account details
    - Declaration on clean track (Proforma enclosed)
    - Signed undertaking to transfer all study data and materials to TAF upon contract completion (Proforma enclosed)

7. EVALUATION CRITERIA

Proposals will be evaluated based on:

Table 2. Evaluation Criteria and Weights

#	Evaluation Criteria	Description	Weight (%)
1	Technical approach	Quality and appropriateness of the proposed approach, including responses to the methodological requirements in Annex 1, Section I.D (recall bias, non-response, comparability)	35
2	Team composition & relevant experience	Qualifications of key personnel and relevant past experience/references	25
3	Qualitative data collection and analysis, including triangulation with quantitative findings	Quality of the proposed qualitative design, including its ability to document the PLA and MEG models and implementation experience (RQ4a), its approach to documenting design models, implementation plans, and adaptations (RQ4b), and its integration with the quantitative DiD findings (Annex 1, Section III.C)	15

#	Evaluation Criteria	Description	Weight (%)
4	Work plan & timeline feasibility	Realistic schedule and milestone clarity	10
5	Financial proposal	Cost-effectiveness and clarity of the itemised budget	15

Evaluation will be conducted through a structured review of both technical and financial proposals. The highest-scoring proposal will be selected as the preferred bidder, subject to due diligence. TAF is not obligated to select the lowest-cost proposal.

Technical proposals must score at least 70% of available technical marks (Criteria 1–4 in Table 2) to qualify for financial evaluation. Only technical proposals meeting this threshold will have their financial proposals opened; the combined technical and financial scores then determine the final ranking.

8. ETHICS, CONFIDENTIALITY, AND OWNERSHIP

- The study partner must obtain IRB approval, registered with the ICMR (or an equivalent recognised body), prior to the start of data collection, and must obtain written informed consent, in Hindi or a locally appropriate language, from all respondents.
- The study partner must comply with the Digital Personal Data Protection Act, 2023 (DPDP Act) in the collection, storage, processing, sharing, and disposal of all personal data gathered during the study.
- Data collection quality assurance (CAPI, daily sync, back-checks, HFCs, and related checks) must follow the protocol set out in Section V.
- State- and district-level government permissions will be facilitated jointly by TAF and the study partner.
- All materials shared by TAF, including the baseline survey tool, report, and de-identified dataset, remain its intellectual property. This extends to all study outputs produced under this assignment, including cleaned datasets, analysis syntax/code, codebooks, qualitative transcripts and notes, and any other deliverables; these remain TAF's property and may not be reused or shared by the study partner beyond this assignment without TAF's written consent. The study partner must transfer all raw and cleaned datasets, analysis syntax/code, codebooks, qualitative transcripts, and study tools to TAF within 15 days of final report acceptance, and must submit a signed undertaking committing to this transfer as part of the proposal (see Section 6.D).
- The selected study partner must treat all documents as confidential.
- The study output will be finalised in consultation with TAF's programme and M&E teams.

9. SUBMISSION

All proposals and queries should be sent to TAF_MnE@antarafoundation.org and cc to puneet.naithani@antarafoundation.org with the subject line *"Damoh Endline Evaluation proposal || Organisation name"*.

Queries will be accepted until the deadline stated on the cover page; responses to queries (RFQ) will be shared with all prospective applicants by the date indicated.

10. GENERAL CONDITIONS

- TAF reserves the right to accept or reject a Bid without assigning any reasons for the same.
- Canvassing or collusion in any form or manner related to this RFP is strictly prohibited. Such acts would lead to automatic disqualification of the concerned Bid(s).

ANNEX 1: DETAILED TERMS OF REFERENCE

I. Background and Rationale

As introduced in Sections 1–2, TAF's Akshita programme began as a supply-side health-systems initiative. The programme comprises three distinct interventions. First is the flagship 'AAA Platform' focusing on improving coordination between and capacity of frontline workers (ASHAs, Anganwadi Workers and ANMs) to provide high-quality MNCHN services to the communities they serve by encouraging joint microplanning and high-risk case identification and management at least once a month in village during Village Health, Sanitation and Nutrition Days (VHSNDs). The second intervention is set of distinct activities focused on strengthening the health system by providing handholding and mentoring to health officials at the block and district levels — building their capacities to become effective managers capable of supporting and handholding the cadres under them. The third supply side intervention focuses on building the capacities nurses in high-burden health facilities and labour rooms to ensure quality intrapartum and post-partum care. As the supply-side programme matured, TAF added a community-level demand-generation layer (PLA and MEG), PLA rolled out from November 2024 in Jabera and Batiyagarh and MEG from September 2025 in Patera, Hatta, and Tendukheda, to address the social and access barriers that supply-side strengthening alone could not resolve — evolving into an integrated model. The integrated model has since been rolled out across the five Damoh blocks named in Section 2: Batiyagarh, Hatta, Jabera, Patera, and Tendukheda.

I.A Theory of Change

As described in Section I, TAF's Akshita programme operates through three tracks: a beneficiary/community track (PLA and MEG), acting on household awareness, social norms, and care-seeking behaviour; a provider track (frontline worker and facility capacity building), acting on the skill and capability of the people delivering care; and a health-system track (AAA Platform and supportive supervision), acting on coordination, data quality, and the consistency of service delivery infrastructure. The provider and health-system tracks both act on the supply side, service availability, quality, and consistency. The beneficiary/community track acts on the demand side, on household awareness and care-seeking behaviour. TAF's theory of change links the beneficiary/community track most directly to four indicators of primary interest (see Figure 1 below): 4+ ANC visits, ANC visit in the first trimester, timely initiation of breastfeeding, and exclusive breastfeeding. This mapping draws on the wider literature on participatory women's-group models in maternal and newborn health, which links community mobilisation to improved care-seeking and newborn-care behaviours.[4] The original baseline sampling and power calculation, however, were based on all nine indicators (Section III.A); these continue to be measured and analysed at endline.

[4] Prost A, Colbourn T, Seward N, Azad K, Coomarasamy A, Copas A, Houweling TAJ, et al. Women's groups practising participatory learning and action to improve maternal and newborn health in low-resource settings: a systematic review and meta-analysis. Lancet. 2013;381(9879):1736–1746.
Seward N, Neuman M, Colbourn T, Osrin D, Lewycka S, Azad K, et al. Effects of women's groups practising participatory learning and action on preventive and care-seeking behaviours to reduce neonatal mortality: a meta-analysis of cluster-randomised trials. PLoS Med. 2017;14(12):e1002467.

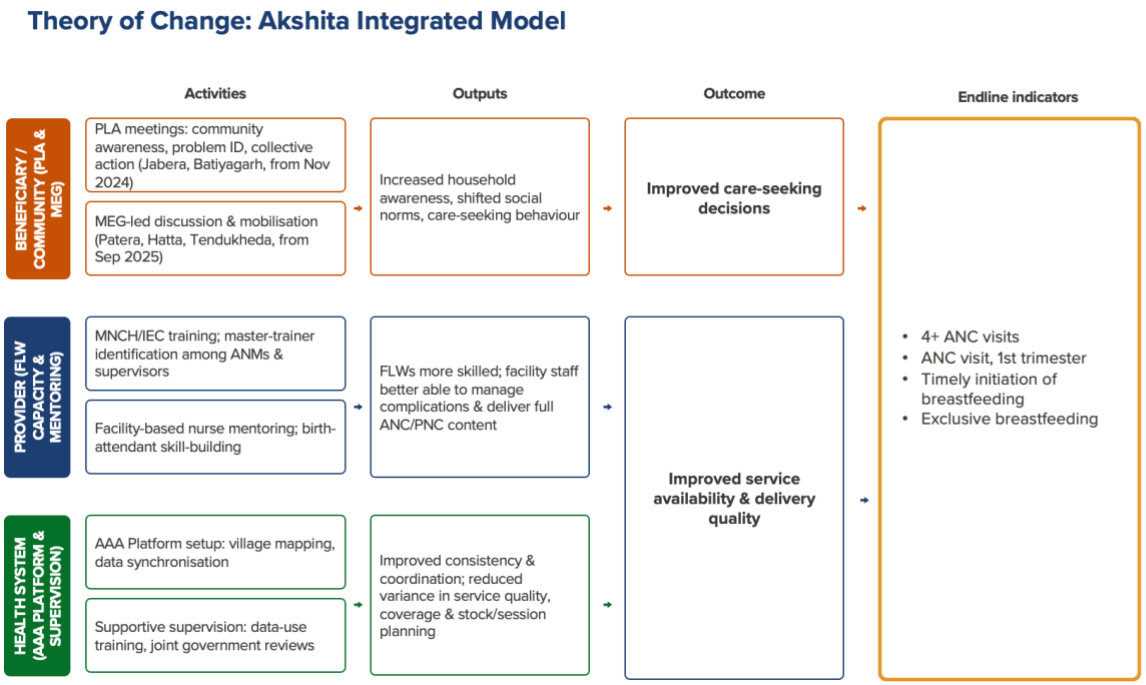

Figure 1. Theory of Change: Akshita Integrated Model, mapping the four indicators of primary interest to the beneficiary/community track.

How agencies should use this theory of change to propose a primary outcome and sampling approach, applying methods appropriate to each comparison's evidentiary strength, is set out in Section III.A. This theory of change is TAF's preliminary hypothesis, based on programme documentation and general evidence on community-based interventions; it has not been independently validated for this specific model. As part of the study, the agency is expected to develop a more detailed theory of change and to test, and revise where warranted, its underlying assumptions using both the quantitative and qualitative endline findings, rather than treating the pathways described here as established (see Section III.C).

I.B The Baseline Assessment (2024–25)

An impact evaluation of the integrated model was designed as a nested design combining two distinct comparisons with different inferential bases. Bidders should note precisely what is, and is not, randomised:

- **Randomised (c-RCT) component *(RQ1a).*** Within Damoh, the integrated model (PLA and MEG) and the supply-side model were implemented across 120 Health and Wellness Centres (HWCs). For the baseline, an HWC was considered as the unit of randomization. About 75% of the HWCs (90 clusters) were randomized to receive the integrated model and the remaining 25% of HWCs (30 clusters) received supply-side interventions only (see Figure 2 below and Table A1). The randomised clusters were stratified by block.

  Note: *This is the only randomised element of the design, and it is used to estimate the incremental effect of the pooled PLA/MEG layer over supply-only (RQ1a). Within the integrated arm, which of the two demand-side interventions a block received (PLA or MEG) was not itself randomised; it follows block, per Table A2 and RQ1b below.*

- **Quasi-experimental (non-randomised) component *(RQ2).*** Damoh's selection as the intervention site, and Panna's selection as the comparison district, were not randomised. Damoh was purposively selected for the reasons given in Section I; Panna was purposively matched via a composite index of six district-level indicators (see Section I.D). This comparison is used to estimate the overall program effect relative to no intervention (RQ2).

The three study arms are:

- **Integrated arm (Damoh):** supply-side interventions plus community-level (PLA in two blocks and MEG in 3 blocks) interventions

- **Supply-only arm (Damoh):** supply-side interventions only
- **Comparison arm (Panna district):** no TAF interventions; matched to Damoh on a composite index of six district-level indicators

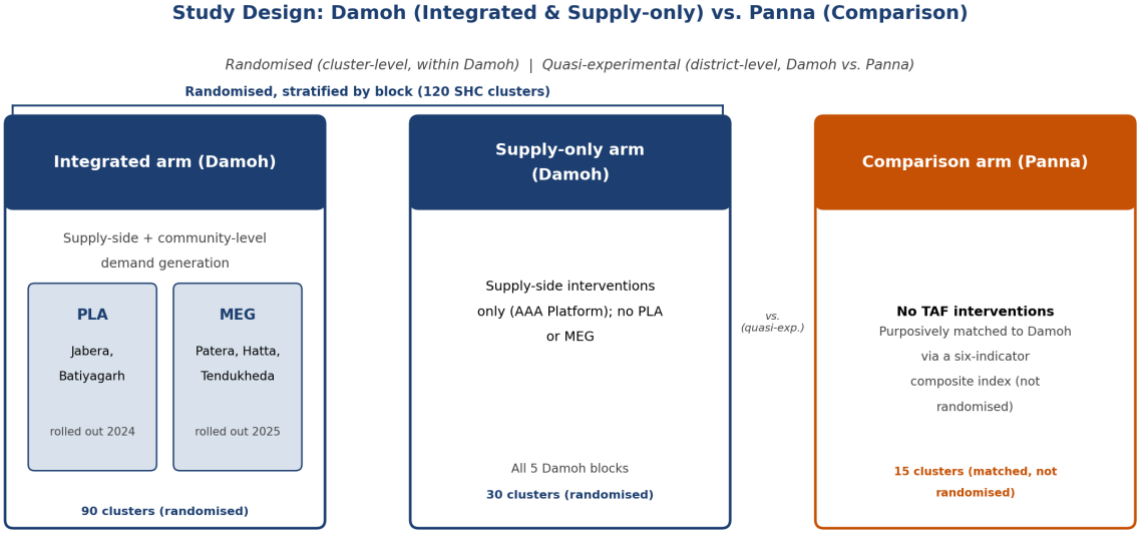

Figure 2. Study design: Damoh's two randomised arms (integrated, split by block into PLA (Jabera, Batiyagarh) and MEG (Patera, Hatta, Tendukheda), and supply-only) compared with the separately matched Panna comparison arm.

Key parameters of the baseline design are summarised below.

Table A1. Baseline Design Parameters

Parameter	Baseline detail
Unit of randomisation	HWC cluster, covering ~5,000 rural population / 5–6 villages
Clusters (Damoh, 5 blocks)	120 HWCs; 90 randomly allocated to integrated arm, 30 to supply-only arm
Clusters achieved (vs. estimated minimum)	17 integrated, 17 supply-only, 15 comparison (49 total) -- the design called for a minimum of 15 clusters per arm; 2 extra clusters were added in each Damoh arm (not Panna) to offset migration-driven respondent shortfall
Respondents	Women with children 0–5 months and 6–23 months, (fresh cross-sectional sample each round, not longitudinal; see Section III.A) systematic random sample within cluster
Achieved sample	3,677 respondents (target 3,600); 6,339 approached to reach this due to non-availability/non-response (see Section I.D for non-response drivers)
Data collection mode	Face-to-face CAPI (MQUAD platform), Hindi, by trained female data collectors
Analysis	Descriptive, bivariate (t-test/z-test), multivariate logistic regression; (baseline analysis only; endline's primary analysis is DiD, with ANCOVA-style covariate adjustment where indicated, using linear or logistic regression as appropriate and accounting for clustering, see Section III.A). Analysis was done on STATA 16.

I.C Key Baseline Findings

Selected findings relevant to endline design:

- Uptake of MNCHN care services was better in Damoh (intervention district) than Panna (comparison) on at least 15 indicators (statistically significant, $p<0.05$), likely attributable to over a

year of supply-side interventions in Damoh at the time of baseline. Two other measured indicators moved in the opposite direction -- postnatal care within 48 hours of delivery, and attendance at Village Health and Sanitation and Nutrition Day (VHSND) -- were both notably higher in Panna, so the comparison should not be read as Damoh leading on every indicator measured.

- Antenatal care testing coverage, early identification of pregnancy complications, and child growth monitoring remained gaps even in the intervention area.
- Disadvantaged groups, including women with no formal education, Scheduled Tribe households, and lower wealth quintiles, showed consistently poorer outcomes across both districts.
- Women's empowerment (self-confidence, self-esteem, mobility) was measured as a composite construct and was positively associated with uptake of MNCHN services, but decision-making across most domains, other than contraceptive use, remained low.

I.D Methodological Considerations Carried Forward from Baseline

The baseline assessment was rigorously conducted, but encountered a few methodological issues that bidding agencies are expected to explicitly address in their technical proposals for the endline:

- **Recall and reporting bias on core outcomes.** Low birth weight relied primarily on maternal recall, as birth weight was missing from the Mother and Child Protection (MCP) card for roughly two-thirds of children, and recorded values showed heaping at exactly 2,500g. Pregnancy and delivery complications were also self-reported with no clinical or record-based verification. As these are central indicators for the DiD comparison, agencies must propose a verification or triangulation approach (e.g., cross-checking against MCP card, or improved probing protocols) rather than relying on recall alone. Agencies should assess the reliability of any facility- or frontline-worker-based records or recall used for this verification, rather than assuming ANM/ASHA recall is itself accurate.
- **Non-response and migration-related attrition.** At baseline, 6,339 women were approached to achieve 3,677 valid interviews (a shortfall of roughly 42%), driven largely by seasonal migration and festival-season travel. Since migrant households may differ systematically from settled ones, agencies must propose a non-response tracking and reporting protocol, including reporting on the socio-demographic profile of non-responders versus responders, so that any bias introduced by differential attrition can be assessed. Since the endline is planned for the same Sept–Oct window as baseline, non-response driven by seasonal migration should be broadly comparable across rounds, which aids comparability; agencies should still track and report on this per the protocol above.
- **Baseline–endline sample comparability.** As a repeated cross-sectional design, endline respondents will necessarily be different individuals from baseline respondents (drawn afresh from women with children 0–23 months at the time of endline). Agencies must report on the comparability of socio-demographic profiles of the baseline and endline samples, to help distinguish genuine program effect from population composition change over the intervening period.
- **Baseline does not pre-date the supply-side program.** Damoh had already received approximately two years of supply-side interventions before this baseline was fielded (see Section I). The baseline report itself documents that MNCHN service uptake in Damoh already exceeded Panna's on roughly 15 indicators at the time of the survey. This baseline is therefore a valid pre-intervention state only for the randomised PLA/MEG layer (RQ1a); for the quasi-experimental Damoh-vs-Panna comparison (RQ2), the endline DiD will capture the incremental effect of the program from 2024 to 2026, not its full effect since 2022. Agencies should frame RQ2 estimates accordingly, and should not describe the DiD as capturing the total effect of the Akshita program.
- **Baseline covariate imbalance between the randomised arms.** With 17 clusters achieved in each of the integrated and supply-only arms, the baseline survey found statistically significant differences between these two randomised arms on at least two household-level indicators (handwashing with soap; access to safe drinking water). Agencies must propose a covariate-adjusted estimation strategy (e.g., ANCOVA-style regression controlling for baseline cluster/household characteristics) for the

RQ1a (incremental) comparison, rather than a simple, unadjusted DiD, for any indicator where baseline imbalance is detected.

- **Baseline covariate imbalance between Damoh and the comparison district, Panna.** The baseline also found statistically significant differences between Damoh and Panna on socio-demographic and household indicators (respondent age, education, age at first birth, mobile phone ownership, wealth quintile, household structure, and water access. Since the Damoh-Panna comparison underlying RQ2 is quasi-experimental, agencies must apply a covariate-adjusted estimation strategy, controlling for these baseline differences.
- **Parallel trends assumption.** This design has only one baseline round, so the parallel-trends assumption underlying the DiD estimator cannot be formally tested against pre-treatment data. Agencies should treat this as a limitation, discuss its plausibility qualitatively (e.g., using contextual or secular trend information), and state it explicitly as a limitation in the endline report rather than asserting parallel trends are satisfied. This limitation applies specifically to the quasi-experimental Damoh-vs-Panna comparison (RQ2); it does not affect RQ1, which relies on cluster randomisation rather than a parallel-trends assumption. As supportive, descriptive evidence for RQ2 only, agencies may plot district-level NFHS-4 (2015-16) and NFHS-5 (2019-21) estimates for the overlapping MNCHN indicators to see whether Damoh and Panna appear to be moving together prior to TAF's 2022 supply-side rollout. Given NFHS district samples are modest in size (roughly 30-40 clusters per district), this should be treated as suggestive, not a formal statistical test of parallel trends, and it will not cover all TAF-specific indicators (e.g., the MNCHN agency module) that NFHS does not measure.
- The comparison district (Panna) was matched to Damoh using a district-level composite index of six indicators rather than cluster-level outcome matching.
- At baseline, both the integrated and supply-only arms in Damoh achieved 17 clusters each (34 total), one more per arm than the planned 15, after adding replacement clusters where originally sampled clusters had no available respondents. Panna's comparison arm achieved the planned 15 clusters with no such replacement. This leaves a minor deviation from the intended 15:15:15 symmetry (17:17:15); these are the same 49 clusters the endline will reuse (see Section III.A), and the deviation does not affect the district-level Panna matching described above.

II. Objectives of the Study

Overall Objective: To complete the impact evaluation of the Akshita program's integrated model in Damoh through an endline assessment, expected to be fielded approximately two years after baseline (around September–October 2026).

Specific objectives:

- Assess the status of women in the study areas at endline, on the same socio-demographic, knowledge, practice, service-access, and health outcome domains measured at baseline, to enable a valid before-after and between-arm comparison.
- Estimate the incremental and overall effectiveness of the integrated model using DiD analysis.
- Assess women's agency specifically in MNCHN care-seeking decisions (new domain; see Section III.B), building on the general empowerment measures established at baseline.
- Through a qualitative component tied to the wider impact evaluation: (a) document how the PLA and MEG models were implemented, adopted, and experienced on the ground in the integrated arm (noting that PLA and MEG operate in different blocks; see Table A2), (b) use this to help interpret the quantitative DiD findings, and (c) document TAF's design models and implementation plans, how these were adapted, and what lessons these offer, through key informant interviews with TAF leadership and the district team, (see Section III.C).

III. Evaluation Design and Methodology

The endline must preserve the baseline's evaluation design to permit valid comparison. The agency is not expected to redesign the evaluation, but to execute the endline round consistently with the parameters below and flag any deviations required due to changed ground conditions (e.g., program roll-out changes, cluster non-availability).

III.A Study Arms, Clusters, and Sampling

The same three study arms, five Damoh blocks, and cluster allocations used at baseline apply at endline:

Table A2. Cluster Allocation by Block

Block	Integrated (Supply + PLA or MEG, by block) clusters	Supply-only clusters
Batiyagarh	18	6
Hatta	16	4
Jabera	21	9
Patera	16	5
Tendukheda	19	6
Total	90	30

The comparison arm remains Panna district (matched at baseline via a six-indicator composite index; see Section I.D).

Within the integrated arm, the two demand-side interventions were rolled out on different timelines by block, using the same cluster allocation as Table A2: PLA was rolled out in 2024 in Jabera and Batiyagarh; MEG was rolled out in 2025 in Patera, Hatta, and Tendukheda. By endline, PLA-block clusters will have roughly two years of exposure and MEG-block clusters roughly one year -- a difference agencies should account for when interpreting RQ1b (Section 2) and the qualitative findings (Section III.C).

Sampling approach (fresh cross-sectional sample, not panel): The endline sample must be a fresh, independent, systematic random sample of women with children 0–23 months, drawn from the same clusters and arms as baseline, consistent with how the baseline itself was drawn and analysed. Proposals suggesting a panel design (re-contacting the same baseline respondents) will not be considered appropriate for this study, for the following reasons, and agencies should not submit panel-based proposals:

- Baseline eligibility was tied to a life-stage window (children aged 0–23 months at time of survey), not a fixed respondent trait; by endline, these same children would be 24–47 months old and outside the eligible window for the indicators being measured (e.g., recent pregnancy/delivery practices, exclusive breastfeeding).
- Randomisation was conducted at the cluster (HWC) level, and the DiD estimator $[I(t)-I(0)] - [C(t)-C(0)]$ is defined using independent arm-level estimates at each round; it does not require paired individual observations.
- Given baseline's already-high attrition from seasonal migration, re-locating the same individuals two years on would likely see worse and non-random attrition, biasing any panel toward more settled households.
- The baseline power calculation was conducted across nine MNCHN service-utilisation and care indicators. The sample size was set to maximum N required across these indicators to ensure the designed was adequately powered. TAF's ToC (Annex 1, Section I.A) maps each of the nine indicators to whichever comparison it is best suited to inform. Agencies should draw on this theory of change, together with their own assessment, to propose which indicator(s) should serve as the

primary outcome(s) for sampling and power purposes, and to propose their sampling approach accordingly, justifying their choice in the technical proposal. Regardless of which indicator(s) are proposed as primary, agencies must apply an appropriate correction for multiple hypothesis testing across all nine indicators at the analysis stage, and should specify their proposed correction method (e.g., false discovery rate control, family-wise error rate control) in the technical proposal. The following nine indicators were considered for sampling at baseline –

- Ante-natal care (ANC) visit in the first trimester
- 4+ ANC visits
- 180+ IFA tablet consumption
- Full ANC
- PNC within 2 days of delivery
- Timely initiation of breastfeeding
- Exclusive breastfeeding
- Dietary diversity of children aged 6-23 months
- Full immunization

DiD analysis: the agency will estimate the incremental effect of the integrated model over the supply-only model, and the overall effect of the integrated model relative to no intervention. This is not a prescribed model: agencies should specify their own regression specification (e.g., outcome ~ treatment + time + treatment×time interaction, plus covariates where indicated in Section I.D), rather than reporting only the aggregate difference of arm-level means. Agencies may propose Stata (version 16.0 or higher), R, or other comparable statistical software for this analysis.

As a secondary, exploratory analysis for RQ1b, the agency should also report the same DiD estimate disaggregated by PLA-block clusters (Jabera, Batiyagarh) versus MEG-block clusters (Patera, Hatta, Tendukheda), clearly labelled as descriptive rather than causal, given that intervention type follows block rather than a randomised assignment and the two interventions differ in exposure duration by endline (see Section I.B).

The agency will also:

- Adapt the baseline CAPI tool for the endline round, retaining all indicator definitions and question wording needed for valid comparison, while adding the new women's agency module (Section III.B) and any TAF-approved program updates.
- Include a brief module (embedded in the CAPI tool) to verify treatment exposure and perceived usefulness/fidelity among integrated-arm respondents (e.g., self-reported participation in PLA or MEG sessions, as applicable to the respondent's block, plus perceived usefulness), to support interpretation of RQ1a/RQ1b and RQ3.
- Update the sampling frame using the ANMOL database (or equivalent government data source current at the time of fieldwork), following the same systematic random sampling procedure used at baseline.
- Field a fresh cross-sectional sample of women with children 0–23 months across all 49 (or currently available) clusters, powered consistently with the baseline's assumptions (80% power, 95% CI, ICC of 0.04, 40 respondents per cluster) unless a revised power calculation is proposed and justified. Consistent with Section I.A and I.D, agencies must calculate and justify their own required sample size based on their proposed primary outcome(s), rather than assuming a fixed target; the baseline's own target of approximately 3,600 was calculated differently (maximum N required across nine indicators) and should not be treated as a floor or ceiling. Sampling will draw from the same 49 clusters. A revised power calculation will only be accepted if justified by a material change in cluster availability, the observed ICC, or the eligible population's characteristics, and must be signed off by TAF's M&E team before fieldwork begins.

- Propose and implement a verification/triangulation approach for recall-sensitive indicators (low birth weight, pregnancy/delivery complications), and a non-response and migration-tracking protocol, reporting on responder/non-responder profiles; see Section I.D.
- Conduct descriptive, bivariate (t-test/z-test), and multivariate linear and logistic regression analysis, including DiD estimation, consistent with the baseline's analytical approach (Table A1), using cluster-robust standard errors or mixed-effects models to account for the clustered design; and report explicitly on baseline–endline sample comparability; see Section I.D.
- Where baseline balance tests show significant differences between the integrated and supply-only arms, apply the covariate-adjusted estimation approach set out in Section I.D rather than an unadjusted DiD.

III.B New Measurement Domain: Women's Agency in MNCHN Care-Seeking

The baseline measured general women's empowerment (self-confidence, self-esteem, mobility, and decision-making across agriculture, household expenditure, own health expenditure, contraception, and visiting family/relatives). It did not measure agency specific to MNCHN care-seeking, i.e., whether a woman can decide and act, with or without needing permission or a companion, to seek ANC care, choose a facility for delivery, take her child for immunisation, or act on a danger sign.

The agency will design and field a short additional module (approximately 3–5 items) on MNCHN-specific care-seeking autonomy, to be analysed both independently and in relation to the general empowerment score established at baseline, so that TAF can understand whether general empowerment translates into health-specific agency.

III.C Qualitative Component: MEG/PLA Perspectives, Tied to the Wider Impact Evaluation

Unlike the recall-sensitive quantitative indicators, the baseline included no qualitative component at all. The endline will introduce a qualitative "dipstick" designed not as a standalone study, but as an integral part of the same endline exercise, explicitly intended to help explain the quantitative DiD findings on the integrated-vs-supply-only comparison.

Requirements for the agency:

- **Purpose:** the qualitative component serves three connected purposes. First, it must document how the PLA and MEG components (in their respective blocks) of the integrated model were adopted, implemented, and experienced on the ground, using a lens broadly consistent with the RE-AIM framework (reach, effectiveness, adoption, implementation, maintenance); this documentation of the models and implementation experience stands on its own value for programme learning. Second, it must help explain why the integrated arm did or did not outperform the supply-only arm on the quantitative DiD findings. Third, it must document TAF's design models and implementation plans for the integrated model, how these were adapted along the way (e.g., the phased PLA-then-MEG sequencing), and what lessons these offer, through key informant interviews with TAF leadership and the district team; this extends the same implementation-research lens (e.g., the process domain of planning, engaging, executing, and reflecting/adapting) from field-level respondents to the programme's own design and rollout decisions.
- **Coverage:** respondents drawn only from integrated-arm blocks, since MEG/PLA does not operate in the supply-only or comparison arms. This applies to the field-level respondent groups (first and second purposes above); the design-intent KIIs with TAF leadership and the district team (third purpose) are not tied to specific blocks.
- **Suggested respondent groups:** MEG members/beneficiaries, PLA facilitators, and frontline workers (ASHA/AWW/ANM) who interface with MEG/PLA, plus TAF leadership and the district team (for the design-and-implementation strand); agencies may propose additions (e.g., supervisors, CHOs) with justification.

- **Timing:** qualitative fieldwork should run concurrently with (or immediately adjacent to) the quantitative fieldwork, not as a separate phase, so that emerging quantitative patterns can inform qualitative probing where feasible.
- **Integration, not parallel reporting:** the agency's technical proposal must describe a specific approach to triangulating qualitative and quantitative findings (e.g., a joint interpretation workshop, a structured mixed-methods matrix, or embedding qualitative narrative directly against relevant quantitative findings). A single integrated endline report is expected, with a dedicated triangulation/mixed-methods section, not a standalone qualitative annex disconnected from the main findings.
- **Theory of change validation:** drawing on both the quantitative DiD findings and the qualitative evidence above, the agency must assess which elements of TAF's theory of change (Section I.A) are supported, contradicted, or require revision, and present this assessment, together with a more detailed theory of change developed through the study, as part of the triangulation/mixed-methods section of the endline report.

IV. Measurement Domains and Indicators

The endline retains all baseline measurement domains (table below) to preserve comparability, with the addition of the MNCHN agency module (Section III.B) and the qualitative component (Section III.C). Full baseline indicator definitions are contained in the baseline survey tool, which will be shared with the selected study partner.

Table A3. Measurement Domains and Indicators

Domain	Illustrative indicators
Maternal and newborn health	Low birth weight, preterm delivery, pregnancy/delivery complications (identification, management, referral) (recall-sensitive; see Section I.D)
Maternal health care practices	Full ANC coverage, IFA/calcium consumption, TT, ANC tests, institutional/c-section delivery, postnatal check-ups, dietary diversity
Newborn and child health care practices	Skin-to-skin care, clean cord care, timely breastfeeding initiation, exclusive breastfeeding, complementary feeding, immunisation
Access to health and nutrition services	AWC services, ASHA/AWW counselling and home visits, VHSND attendance, JSY/JSSK entitlements, SAM/MAM identification and referral
Women's awareness of MNCHN practices	Danger sign recognition, awareness of entitlements and FLW/community platform services
Women's individual/household characteristics	Socio-demographics, food security, migration status
Women's empowerment (general)	Self-confidence, self-esteem, mobility, decision-making across household domains
Women's agency in MNCHN care-seeking (NEW)	Autonomy in ANC-seeking, facility choice for delivery, immunisation follow-through, acting on danger signs
Qualitative: PLA and MEG implementation and design (NEW)	Adoption, implementation fidelity, perceived effectiveness, and maintenance of PLA (Jabera, Batiyagarh) and MEG (Patera, Hatta, Tendukheda), from beneficiary, facilitator, and FLW perspectives; plus design models, implementation plans, and adaptations, from TAF leadership and district-team perspectives

V. Data Management, Quality Assurance

- Data collection via CAPI (or agency-proposed equivalent), with daily data sync and centralised quality checks, consistent with baseline practice.

- Spot checks and back-checks by an independent, trained team member (minimum 10% of interviews, or as otherwise justified by the agency), with periodic debrief sessions during fieldwork.
- High-frequency checks (HFCs) on incoming CAPI data, run daily or near-daily, to flag outliers, skipped items, unusually short/long interview durations, and enumerator-level anomalies.
- GPS verification of interview locations against expected cluster/village boundaries.
- Enumerators must complete a structured training (classroom plus supervised field practice) and a pilot test of the tool prior to full rollout; the agency's team lead must certify enumerators as meeting quality standards before data collection begins.
- Written informed consent in Hindi (or locally appropriate language), voluntary participation, and de-identified data storage.
- Study protocol and tools to be reviewed and approved by the IRB prior to fieldwork.

VI. Deliverables

Table A4. Study Deliverables

Deliverable	Description
Inception Report	Finalised quantitative and qualitative tools, updated sampling frame, and detailed protocols for recall-bias mitigation and non-response tracking (Section I.D)
IRB Approval Documentation	Ethics approval prior to fieldwork
Datasets and Transcripts	Cleaned, documented quantitative dataset and qualitative transcripts/notes
Draft Endline Report	Quantitative findings, DiD analysis, qualitative findings (including documentation of the PLA and MEG models and implementation experience), and the mixed-methods triangulation section
Final Endline Report	Incorporating TAF feedback
Presentation	Presentation of key findings to TAF