STUDY REPORT

AI-NETRA for Tuberculosis Active Case Finding and Diagnostic Network Optimization

A Three-Part Mixed-Methods Pilot Evaluation from Pune District, India

---

Government Partner: **State TB Office, Maharashtra** | Implementation Partner: **FIND**

Development Partner: **Froncort.AI** with support from **Google.org.**

Ethics approval**: JCDC/BHR/26/005**, 13 March 2026

22 July 2026

v1.1 (10 November 2026)

Govt. of Maharashtra, Health Services
Jt. Director of Health Services (Leprosy & TB)
"AROGYA BHAVAN" Opp. Vishrantwadi Police Station,
Alandi Road, Yerwada, Pune-411006.

Jt.Director - (020) 26686955
Dy.Director - 26686951
Office - 26686952-54
Fax - 26686956

Section wise e-mail
TB section- stomh@rntcp.org
Lep section - jtlepnms@rediffmail.com
Est section - jdhsest99@gmail.com

Foreword

Tuberculosis remains among the most pressing public health challenges before Maharashtra and before the nation. Even as notifications rise and treatment outcomes improve, the disease continues to concentrate in the neighbourhoods that are hardest to reach and easiest to overlook. Our shared commitment, in step with the Sustainable Development Goal target of ending the epidemic by 2030, obliges us to find these cases earlier, to reach the households that routine campaigns pass by, and to use every rupee of programme effort where it will do the most good.

Active case finding has long been central to that effort. Its limitation has never been intent or diligence; it has been precision. Teams screen very large populations to find comparatively few cases, and the block or the ward, our usual unit of planning, is far larger than the pockets in which transmission actually clusters.

It is against this background that the State Tuberculosis Office welcomed the opportunity to evaluate AI-NETRA in Pune district. The tool resolves risk to small settlement clusters and turning it into a screening plan that a frontline worker can follow, household by household. Our interest was never in the novelty of the technology, but in whether it could make honest, measurable use of the resources we already have.

I am encouraged by what this evaluation has shown. Our district teams and our ASHAs took to the tool readily and told us, in their own words, where it helped and where it must improve; that candour is itself a mark of a healthy programme. The accompanying work on diagnostic network optimization, which routes samples to the nearest capable laboratory, points to gains in efficiency that can be realised with the machines already in place. Throughout, the study has been disciplined about the difference between what has been demonstrated and what has not.

I would equally underline the study's own counsel. This is the experience of a single district, and the findings, though promising, rest on a small number of confirmed cases. Before any wider adoption, a larger, concurrent study, in which AI-guided and routine case finding run side by side under matched conditions, will be necessary to establish that the results are comparable and replicable. That is the right and responsible next step, and it is one the State would be glad to support.

Maharashtra is glad to have contributed this field evidence, and we look forward to the next phase of study.

Our mission is clear, and it is urgent. Work of this kind, honest about its limits and disciplined in its evidence, brings us closer to a TB-Mukt Maharashtra and a TB-Mukt Bharat.

Dr. Radhakishan Pawar
Joint Director of Health Services,
(TB & Leprosy), Arogya Bhavan, Pune

Revision note (version 1.1, 10 November 2026).

This version corrects the report of 22 July 2026 in four respects.

(1) Urban referral precision (Sections 3.3.3 and 4, Table 3.11): the earlier figure of 4.3% versus 1.2% (3.6-fold) came from an interim data cut of 23 June 2026 pooled across all four units; on the final data under the primary definition it is 9 of 271 (3.3%) versus 7 of 523 (1.3%) in the urban units, 2.5-fold (95% CI 0.93 to 6.59, p=0.067), and 10 of 325 (3.1%) versus 9 of 746 (1.2%) pooled, 2.6-fold (p=0.043).

(2) Case ascertainment (Sections 2.5, 3.3.1, 3.3.2 and 4, Tables 3.6 and 3.8): of the 15 all-forms cases in the AI arm, 10 were NAAT-positive pulmonary TB and 5 were diagnosed at the referral facility on chest radiography or cytopathology; the earlier text described all 15 as bacteriologically confirmed and the AI arm’s ascertainment as sputum-only, and the primary endpoint definition is now stated for each arm.

(3) Acceptability and time-and-motion (Section 3.2.2): focus-group results are reported as group-level counts from the compiled discussion template, and plan-access time as a median of 54 seconds (IQR 42 to 69) among the 37 workers with all five steps timed.

(4) Figures 2 and 3 are redrawn without embedded figure numbers, and the field-data extract is dated 29 June 2026 throughout. No primary or secondary rate ratio, NNS or cost estimate has changed.

Table of Contents

Abbreviations and Key Terms 5
Abstract 6
1 Introduction 7
1.1 Objectives 7
2 Methods 7
2.1 Design and setting 7
2.2 The AI-NETRA prediction model 7
2.3 Part 1: retrospective validation 8
2.4 Part 2: acceptability and usability 8
2.5 Part 3: prospective field evaluation 8
2.6 Sub-study: AI-DNO diagnostic network optimization 9
2.7 Sub-study: cold-start prediction (PDFM) 9
2.8 Statistical methods and analysis plan 9
2.8.1 Estimand and design 9
2.8.2 Analysis populations and comparators 9
2.8.3 Outcomes and statistical methods 10
2.8.4 Statistical power 10
2.8.5 Reporting standards 10
2.9 Ethics 11
3 Results 12
3.1 Part 1: Retrospective validation 12
3.1.1 Campaign characteristics 12
3.1.2 Case capture and prediction error 12
3.1.3 Screening efficiency 13
3.1.4 Taluka-level prediction accuracy 13
3.1.5 Cost implications 14
3.2 Part 2: Acceptability and usability 14
3.2.1 Program-leadership stratum 14
3.2.2 Frontline-worker stratum 15
3.2.3 Post-campaign frontline survey 15
3.3 Part 3: Prospective field evaluation 16
3.3.1 Screening, case ascertainment, and primary endpoint 16
3.3.2 Total case detection (all forms) 18
3.3.3 Presumptive yield and referral precision 18
3.3.4 Screening efficiency and cost 19

3.4 Sub-study: AI-DNO diagnostic network optimization 19
3.5 Sub-study: cold-start prediction (PDFM) 21
3.6 Mixed-methods integration 21
4 Discussion 22
4.1 Toward a precision-targeted saturation model 23
5 Limitations 23
6 Conclusions 24
6.1 For the Central TB Division and national policy 24
6.2 For state TB programs 24
6.3 For funders and research partners 24
7 Way Forward 25
7.1 A concurrent, like-for-like trial 25
7.2 An environmental cold-start layer 25
7.3 An offline, field-first local language application 25
7.4 Program enablers 25
Competing interests 25
Funding 25
Contributions 26
Acknowledgments 26
8 References 27
8.1 Study documents 27
8.2 Cited literature 27
Appendix A. Reporting-standard checklists 28
A.1 STROBE (observational analyses) 28
A.2 TREND (quasi-experimental comparison) 28
A.3 TRIPOD+AI (prediction model) 28
A.4 COREQ (focus group discussions) 29

Abbreviations and Key Terms

Term	Definition
ACF	Active case finding; systematic screening to detect TB in people who have not presented to care.
AI-ACF / AI-DNO	The two AI-NETRA engines: geospatial active case finding, and diagnostic network optimization.
ASD	Average service distance; mean distance a sample travels from a collection point to its laboratory.
ASHA	Accredited Social Health Activist; community frontline health worker.
CBNAAT / NAAT	Cartridge-based and other nucleic-acid amplification tests for bacteriological TB confirmation.
COREQ	Consolidated criteria for reporting qualitative research.
FGD	Focus group discussion.
H3	Uber's hexagonal hierarchical geospatial index; the analytical grid used by AI-NETRA.
MAE	Mean absolute error; average absolute difference between predicted and observed case counts.
Nikshay	India's national TB notification and case-management information system.
NNS	Number needed to screen; persons screened per confirmed case (the inverse of yield).
NTEP	National TB Elimination Program.
PDFM	Population Dynamics Foundation Model (Google Research).
RR	Rate ratio; the ratio of two rates (here, AI-guided versus standard ACF).
SAP	Statistical analysis plan.
STROBE / TREND	Reporting standards for observational studies and non-randomized evaluations.
TRIPOD+AI	Reporting standard for clinical prediction models, including those using machine learning.
TB Unit (TU)	The NTEP operational unit for TB service delivery within a district.

Abstract

Background. Tuberculosis (TB) active case finding (ACF) under India's National TB Elimination Program is effective but inefficient when screening is untargeted, producing a high number needed to screen (NNS). TB is spatially clustered, which creates an opportunity for geographic targeting. AI-NETRA is a geospatial, agentic artificial-intelligence system that predicts neighborhood-level TB risk on an H3 hexagonal grid from Nikshay notification data and translates that prediction into field-ready microplans.

Methods. We evaluated AI-NETRA in three parts. Part 1 was a retrospective validation against three successive ACF campaigns in Pune district. Part 2 assessed acceptability and usability with program leadership (n=15), through focus group discussions with frontline workers (n=100 across 16 groups), a controlled time-and-motion study (n=57), and a post-campaign frontline survey (n=101). Part 3 was a prospective field evaluation comparing AI-guided ACF against standard ACF across four TB Units (rural Bhor and Shirur; urban Haveli PMC and Haveli PCMC). The pre-specified primary endpoint was bacteriologically confirmed, nucleic-acid amplification test (NAAT) positive pulmonary TB, analyzed as-screened. Secondary endpoints were all-forms case yield, presumptive yield, NNS, and cost per confirmed case. Two sub-studies evaluated diagnostic network optimization (AI-DNO) and cold-start risk prediction using the Google Population Dynamics Foundation Model (PDFM).

Results. In the retrospective validation, AI-NETRA captured 89% of notified cases from under one-third of the screened population, reduced NNS by 64%, and improved the rank correlation between predicted risk and observed cases from 0.55 to 0.78 (p=0.002). In the prospective evaluation, the AI arm confirmed 10 NAAT-positive pulmonary cases from 66,095 individuals screened versus 9 from 177,322 in the standard arm, a rate ratio (RR) of 2.98 (95% CI 1.09 to 8.29, exact p=0.019); NNS was 6,610 versus 19,702 (66% lower) and cost per bacteriological case fell from Rs. 27.6 lakh to Rs. 9.3 lakh. Total case detection using an all-forms definition did not differ (AI 15 versus standard 32; RR 1.26, 95% CI 0.63 to 2.39, p=0.51). Presumptive yield favored the AI arm in rural units (RR 2.22) but not urban units (RR 1.01). Acceptability was high (93% leadership willingness; 96% of frontline workers reported finding the same suspects while screening fewer people) alongside specific usability gaps in dense urban navigation and plan-access overhead. AI-DNO reached 77.1% concordance with an expert panel (about three times faster than manual planning). The PDFM cold-start sub-study was directional only.

Conclusions. AI-NETRA is a usable, well-received targeting layer that concentrates ACF into higher-positivity pockets and lowers NNS and cost per confirmed case. Per-person confirmed-case yield favored the AI arm, but total case detection did not differ in a non-concurrent comparison with unmatched case ascertainment. A concurrent, like-for-like cluster-randomized trial with harmonized case definitions is the appropriate next step.

Keywords. *tuberculosis; active case finding; geospatial artificial intelligence; H3; number needed to screen; diagnostic network optimization; India.*

1 Introduction

India carries the largest share of the global TB burden, and the National TB Elimination Program (NTEP) has made ACF a central instrument for finding people with TB who have not presented to care. [1,2] ACF is effective, but when screening is spread evenly across a population it is expensive per case found, because the number needed to screen to detect one confirmed case is high in low-prevalence areas. TB, however, is not evenly distributed. It clusters in identifiable neighborhoods, which means that where screening is directed matters as much as how much screening is done. [3]

AI-NETRA was built to make that targeting decision data-driven and reproducible. The platform ingests routine Nikshay notification data and other geospatial signals, learns the spatial structure of TB risk, and produces a risk surface on an H3 hexagonal grid. [5] Each hexagon is scored and ranked, and the highest-risk hexagons are assembled into field-ready microplans that tell a frontline worker where to go. The intent is not to replace program judgment but to concentrate a fixed screening effort where the yield is highest, lowering NNS without lowering the number of cases found.

1.1 Objectives

This report sets out to (i) validate AI-NETRA's predictive accuracy retrospectively against real ACF outcomes; (ii) establish whether the tool is acceptable and usable to the people who must operate it; and (iii) evaluate, prospectively and in the field, whether AI-guided ACF improves case yield and screening efficiency relative to standard practice. Two sub-studies extend the evaluation to diagnostic network optimization and to the cold-start problem of predicting risk where notification history is sparse.

2 Methods

2.1 Design and setting

The evaluation was conducted in Pune district, Maharashtra, across four TB Units selected to span the rural and urban ACF context: Bhor and Shirur (rural) and Haveli PMC and Haveli PCMC (urban). The overall design is a three-part mixed-methods pilot: a retrospective prediction-validation component, a mixed-methods acceptability component, and a prospective quasi-experimental field comparison of AI-guided versus standard ACF. Reporting follows STROBE for the observational analyses, TREND for the quasi-experimental comparison, TRIPOD+AI for the prediction model, and COREQ for the qualitative work; item-level checklists are provided in Appendix A. [9,10,11,12]

2.2 The AI-NETRA prediction model

AI-NETRA represents geographical areas as H3 hexagons and predicts a TB-risk score per hexagon from routine Nikshay notification data and associated geospatial features. Hexagons are ranked as per predicted TB risk, and the highest-risk hexagons are aggregated into microplans. The prediction model, its inputs and outputs, and its intended use are reported in line with TRIPOD+AI. [11] The retrospective analyses use the model as it was refined over successive campaigns; the prospective analyses use a single locked model version applied ahead of the field sweep.

2.3 Part 1: retrospective validation

Aggregate taluka-level ACF outcomes were obtained from the District TB Office for three campaign periods (October 2023, December 2024, and March 2025) as reported on the standard TB-4 form. A fourth campaign (September 2022) was excluded from the primary analysis because its reported screening volumes exceeded the census population, and it is noted only as supplementary detail. For each period, per-taluka AI-NETRA predictions specified the number and priority of predicted hotspot hexagons, the estimated population within hotspot zones, and the predicted case counts. All predictions used only notification data available before the respective campaign, a rolling temporal validation that preserves strict temporal separation between the data the model learned from and the outcomes it was tested against.

The prediction model (version 1) was trained on the Nikshay notification register for Pune Rural, covering actively and passively detected cases from 2022 to 2025 and restricted to bacteriologically confirmed cases, so that the model learns the spatial distribution of confirmed TB in order to predict where future cases are likely to arise. Geocoding assigned about 92% of training records to H3 cells; roughly 16% resolved to building-level precision and 72% to area-level precision. All geocoded records were retained regardless of precision, because restricting to high-precision addresses would have reduced the training set to a fraction too small for spatial learning. The target population and the NNS denominator were computed at H3 resolution 9 (about 325x325 meter) for operational feasibility, while hotspot locations for navigation were provided at resolution 10 (about 123x123 meter). Both figures are expressed as the side of a square of equal area to the hexagonal cell, for ease of interpretation.

The primary retrospective outcome was case capture, the percentage of actual ACF-detected cases falling within the model's prioritized hexagons. Secondary outcomes were the reduction in population screened, the reduction in NNS, and taluka-level discriminative ability. Analyses used the predicted-to-actual ratio, the absolute prediction error, the Spearman rank correlation, the Pearson correlation with and without the dominant Haveli outlier, and the mean absolute error per taluka. Cost implications used standardized program assumptions (Rs. 700 per test and a 20% testing rate). Analyses were conducted in Python.

2.4 Part 2: acceptability and usability

Acceptability was assessed at two levels. Program leadership (n=15) completed a structured willingness-to-adopt survey. Frontline acceptability used three instruments: focus group discussions with 100 frontline workers across 16 groups (17 April to 11 May 2026, analyzed under COREQ); a controlled time-and-motion study of 57 workers measuring the operational overhead of accessing and following a microplan; and a post-campaign survey of 101 ASHAs administered after they had actually used AI-NETRA microplans in the field, which is treated here as the primary in-practice acceptability signal.

2.5 Part 3: prospective field evaluation

The AI arm was a two-week AI-guided ACF sweep across the four TB Units, in which teams screened the individuals in AI-NETRA microplans. The comparator (standard arm) was the corresponding routine ACF campaign conducted in the same four units over December 2024 to January 2025. The pre-specified primary endpoint was bacteriologically confirmed, NAAT-positive pulmonary TB, matching the 2025 standard case definition, analyzed on the as-screened population. Secondary endpoints were the all-forms case yield (any bacteriologically or clinically confirmed TB, used as the total-case-detection measure), presumptive yield, NNS, and cost per confirmed case. Denominators were the number of individuals screened in each arm.

Within each unit, all high-priority (red) hotspots were covered comprehensively, medium-priority (orange) hotspots were sampled by about one-third, ranked, geographically stratified selection, and non-hotspot (green) areas served as an opportunistic specificity check. Field teams were trained before deployment in each unit. The

mobile X-ray van was not deployed with the AI sweep, where chest radiography was available only on referral to a health facility, whereas the standard campaign screened with the van at the point of contact. The primary endpoint was therefore held to bacteriologically confirmed pulmonary TB in both arms: NAAT-positive pulmonary TB in the AI arm (the NTEP standard definition, and every bacteriologically confirmed pulmonary case in that arm was NAAT-positive), and the TB-4 reporting category of sputum smear- or NAAT-positive pulmonary TB in the standard arm, which also admits smear-only diagnoses and is, if anything, the broader of the two. The all-forms definition serves as a conservative sensitivity analysis that favors the standard arm.

2.6 Sub-study: AI-DNO diagnostic network optimization

AI-DNO was evaluated over a diagnostic network of 49 devices, 43 laboratories, 197 sample-transport routes, and 53,705 tests. Model-recommended network changes were compared against an expert panel across multiple optimization domains in a multi-rater evaluation (Union World Conference on Lung Health, Abstract 3160), and routing efficiency was measured as the change in mean sample-to-laboratory distance. [S7]

Network state data (laboratories, device inventories, referral routes, and annual test volumes) were ingested from program-entered district data. Device capacity was derived from program parameters, and utilization was computed as routed demand divided by capacity. Transport distance used a circuitry-adjusted great-circle approximation (Haversine distance multiplied by 1.4), with transit time derived from local vehicle speed. Routing efficiency was summarized by the average service distance (ASD), reported both unweighted (per referring facility) and volume-weighted. Optimized routing was generated by a mixed-integer linear programming solver that minimizes aggregate transport distance subject to device-capacity, maximum-distance, and geographic-feasibility constraints. The reported scenario is a pure referral redesign, with no device relocation or procurement.

2.7 Sub-study: cold-start prediction (PDFM)

To test prediction in zones without notification history, we evaluated Google's Population Dynamics Foundation Model (PDFM) embeddings under walk-forward validation in cold-start hexagons, alone and in a hybrid with weather and air-quality signals, benchmarked against a naive baseline. [S6]

2.8 Statistical methods and analysis plan

2.8.1 Estimand and design

The prospective evaluation is a quasi-experimental comparative-effectiveness study using a controlled before-after design, contrasting two ACF planning strategies (AI-guided micro-targeting versus standard blanket planning) in the same geographies. It is not a case-control study; it contrasts strategies, not exposures. The primary estimand is the relative change in bacteriological TB yield and in the number needed to screen achieved by AI-guided ACF relative to standard ACF, across all four units. The estimand is expressed as a rate (cases per persons screened), with the number of persons screened entering as an offset, so that units of differing size are compared on a common footing.

2.8.2 Analysis populations and comparators

The primary population is as-screened: all individuals actually screened under each approach, reflecting real-world delivery. A secondary as-planned population (individuals in the zones the model designated for screening) assesses plan fidelity and is deferred to data lock. The primary comparator is the standard blanket ACF campaign run in the same units, with the December 2024 district campaign as the matched reference; rural-only and urban-only contrasts are reported as subgroup detail. An optional concurrent difference-in-differences contrast against the contemporaneous campaign is available where a valid concurrent control exists.

2.8.3 Outcomes and statistical methods

Rate ratios comparing the arms were estimated with exact Poisson methods, using the exact conditional (binomial) form of the Poisson rate-ratio test given the small case counts, with exact 95% confidence intervals. NNS was computed as individuals screened per confirmed case, with intervals obtained by inverting the exact Poisson interval. Presumptive yield used Fisher exact and log-binomial methods. Cost per confirmed case applied program ACF costing to each arm's screening effort and case yield. Outcome-to-method mapping is summarized in Table 2.1.

Table 2.1. Outcome-to-method mapping

Outcome or question	Primary method	Supporting method
TB positivity (yield) rate ratio	Exact Poisson (conditional binomial) test	Poisson regression with screened offset
Number needed to screen	Inversion of the exact Poisson interval	Delta-method interval as cross-check
Presumptive yield	Fisher exact test	Log-binomial regression
Spatial concentration	Share of cases in top hexagons; precision-at-k	Exact binomial; bootstrap intervals
Spatial calibration	Poisson calibration regression	Spearman rank correlation
Concurrent comparison (DiD)	Poisson or negative-binomial, arm by period	Overdispersion check; negative-binomial fallback

Negative-binomial models are the fallback wherever Poisson overdispersion is detected. Source: statistical analysis plan.

2.8.4 Statistical power

The prospective phase is powered to estimate effect sizes with confidence intervals, not to confirm small differences through hypothesis testing, a decision driven by the number of confirmed cases achievable within the field window. Detecting a rate ratio of 1.25 at 80% power would require on the order of 640 confirmed cases; a ratio of 1.5 about 200; a ratio of 2.0 about 75; and a ratio of 3.0 about 35 (Table 2.2). At small counts, large effects remain detectable while modest ones do not, and every prospective comparison should be read in that light. For interval estimation, an exact 95% Poisson interval around an observed count of 10 spans roughly 4.8 to 18.4, illustrating the precision available at this scale. The retrospective dataset, with about 380 confirmed cases, carries the formal inferential weight, and Nikshay hexagon attribution continues to accrue prospective events over time.

Table 2.2. Approximate confirmed cases required for 80% power, by effect size

True rate ratio (targeted versus standard)	Approximate confirmed cases for 80% power
1.25	about 640
1.50	about 200
2.00	about 75
3.00	about 35

2.8.5 Reporting standards

Reporting follows the standard appropriate to each component: TREND for the primary non-randomized evaluation, STROBE for observational comparisons, TRIPOD+AI for the prediction model, and COREQ for the focus group work, with mixed-methods integration reported per GRAMMS. [9,10,11,12,14] Item-level checklists for the core standards are provided in Appendix A.

2.9 Ethics

The prospective phase was approved under JCDC/BHR/26/005 (13 March 2026). Routine program data were used under the applicable NTEP data-governance arrangements.

3 Results

3.1 Part 1: Retrospective validation

In brief. The model was checked against three real Pune ACF campaigns. In the most recent, its plan would have found about 89 of every 100 cases the campaign actually detected while sending workers to under a third of the population. It also improved each time it was run: the district case-count error fell from 47 to 12, and the ranking of which sub-districts carried the most TB went from a weak match to a strong one.

3.1.1 Campaign characteristics

Three ACF campaigns with temporally matched predictions were evaluated. A September 2022 campaign was excluded from the primary analysis because its screening denominators were implausible (reported volumes exceeded the census population). The three retained campaigns are characterized in Table 3.1.

Table 3.1. ACF campaign characteristics, Pune district

ACF campaign	Duration	Population screened	TB cases	Positivity	NNS
October 2023	11 days	461,467	163	0.035%	2,831
December 2024	14 days	523,911	108	0.021%	4,851
March 2025	13 days	535,978	109	0.020%	4,917

Positivity and NNS reflect blanket (untargeted) ACF.

3.1.2 Case capture and prediction error

Averaged across the three campaigns, the model predicted about 93% of actual cases. In the most recent campaign (March 2025) it predicted 89% of detected cases (97 of 109), and the absolute prediction error improved consistently from 47 to 33 to 12 cases (Table 3.2; Figure 1). The December 2024 figure of 131% reflects the model predicting about 31% more cases than the campaign detected, consistent with campaign under-detection rather than model over-prediction.

Table 3.2. Predicted versus actual case capture, by campaign

ACF campaign	Actual	Predicted	Capture	Absolute error	Trend
October 2023	163	116	71%	47	Baseline
December 2024	108	141	131%	33	Improving
March 2025	109	97	89%	12	Best

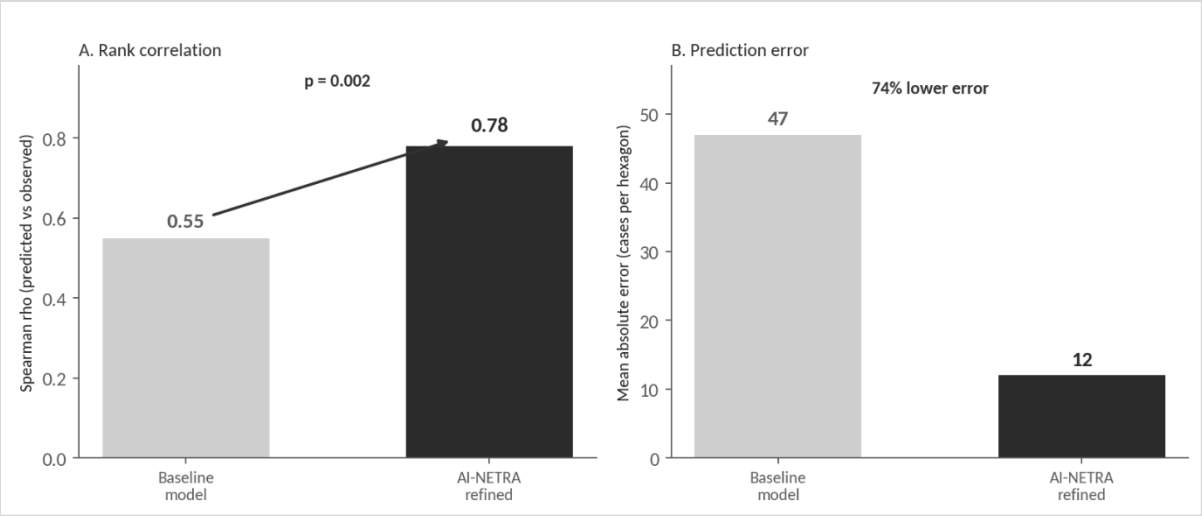


Figure 1. Retrospective predictive accuracy across three ACF campaigns.
Source: AI-NETRA Pune Pilot Study Report [S1].

3.1.3 Screening efficiency

In March 2025 the model concentrated screening into under one-third of the original population (about 169,973 versus 535,978), a 68% reduction in population screened and a 64% reduction in NNS. As iterations progressed, the targeted population tightened from about 296,000 to 170,000 (a 43% tightening) while case capture rose from 71% to 89%, evidence of greater precision without loss of detection (Table 3.3).

Table 3.3. Screening efficiency across campaigns

ACF campaign	Baseline screened	AI target	Population reduction	Baseline NNS	AI NNS	NNS reduction
October 2023	461,467	296,121	36%	2,831	2,553	10%
December 2024	523,911	211,626	60%	4,851	1,501	69%
March 2025	535,978	169,973	68%	4,917	1,752	64%

The AI target population (the NNS denominator) is computed at H3 resolution 9 (about 325x325 m) for operational feasibility; navigation hotspots are provided at resolution 10 (about 123 x 123 m). Source: AI-NETRA Pune Pilot Study Report [S1].

3.1.4 Taluka-level prediction accuracy

The model's ability to rank which talukas would carry higher case burdens improved markedly. The Spearman rank correlation rose from 0.547 (borderline, p=0.053) in October 2023 to 0.779 (p=0.002) in December 2024. Excluding the dominant Haveli outlier, the Pearson correlation rose from a non-significant 0.271 to 0.738 (p=0.006), confirming discrimination across the full range of burden independent of the outlier. The improvement trajectory is shown in Table 3.4.

Table 3.4. Prediction accuracy, improvement trajectory

Metric	October 2023	December 2024	Trend
Aggregate case capture	71%	131% (Mar 2025: 89%)	Toward 100%
Absolute error (cases)	47	33 (Mar 2025: 12)	Improving
Spearman rho	0.547	0.779 (p<0.01)	Improving
Pearson r (all talukas)	0.755	0.866	Improving
Pearson r (excluding Haveli)	0.271	0.738 (p<0.01)	Improving
MAE (cases per taluka)	5.3	4.2	Improving

March 2025 per-taluka actuals were available in aggregate only. Source: AI-NETRA Pune Pilot Study Report [S1].

3.1.5 Cost implications

Under the standardized costing assumptions, concentrating screening into the predicted hotspots reduces the cost of a full district ACF cycle from about Rs. 7.3 crore to about Rs. 3.3 crore [S1]. These are modeled projections contingent on achieving the predicted detection in the field.

> ***Interpretation.*** *The retrospective signal is consistent and strong: the model reliably concentrates cases into a minority of the geography, and its accuracy improves as it is retrained. The December 2024 capture above 100% should be read as directional evidence of concentration and of likely campaign under-detection, not as a precise capture fraction. These are simulated projections against historical campaigns; the prospective phase in Section 3.3 is what tests them in the field.*

3.2 Part 2: Acceptability and usability

In brief. Acceptability was tested in two strata. District and state TB leaders reviewed the platform running on their own data and voted on each feature; 14 of 15 said they would use it, and the features they most wanted are ones AI-NETRA already provides. Frontline ASHAs and TB staff then used the tool in the field: most found the maps useful and easy to read and would recommend it, with two clear messages, that dense-urban navigation often failed and that a people-to-screen quota display was poorly received.

3.2.1 Program-leadership stratum

Structured focus group discussions with an embedded feature-prioritization survey were conducted with TB program stakeholders (District TB Officers, supervisory staff, and state program leadership; n=14 to 16 per item), who reviewed AI-NETRA outputs generated from real district data. Overall acceptability was 93% (14 of 15 willing to use the tool for ACF), with usefulness ratings clustered at the top of the scale.

Across six functional domains and 22 functionalities, 71 feature options were assessed. Fourteen options (20%) received at least 80% endorsement and 29 (41%) received at least 67%. The two highest-endorsed case-finding features, landmark-based hotspot identification and key-population mapping (both 81%), are core implemented capabilities of AI-NETRA. Within the case-finding domain, high-resolution (100x100 m) hotspot prediction was the most endorsed targeting option, offline map capability was strongly endorsed (75%, reflecting poor rural connectivity), and next-month forecasts were preferred (71%, aligning with the campaign cycle). Continuation of paper-based methods (19%) and web-only portals (13%) were the least endorsed, indicating a clear preference for mobile-first tools. The most-endorsed options are shown in Table 3.5.

Leadership identified geospatial data quality as the primary implementation constraint: free-text, inconsistent address quality in the notification register; reliance on permanent rather than current addresses; and the absence of geo-coordinates in standard Nikshay exports despite their existence within the system. Additional constraints were the absence of real-time data infrastructure, limited field-level digitization, and shapefile availability restricted to Pune Rural.

Table 3.5. Feature options receiving at least 80% leadership endorsement

Domain	Feature option	Endorsement
Interface	Voice-based commands for hands-free querying	100%
Monitoring	Comprehensive role-based dashboard	93%
Interface	Responsive web portal (desktop and mobile)	93%
Planning	Resource-constraint simulation	87%
Monitoring	Real-time field-team location map	87%
Case finding	Daily target-area summary via SMS or WhatsApp	86%
Case finding	High-priority geographic areas with landmarks	81%
Case finding	Key-population and vulnerable-group mapping	81%
Network	Real-time GPS sample tracking	81%
Network	Sample backlog alerts	81%
Network	Supervisor override of transport routes	80%
Planning	Cross-program data integration	80%
Reporting	Custom report builder and in-app feedback	80%

Source: AI-NETRA Pune Pilot Study Report [S1].

3.2.2 Frontline-worker stratum

Sixteen focus group discussions were held across the four TB Units between 17 April and 11 May 2026, with 100 participants (86 ASHAs alongside supervisory staff, multipurpose workers, and TB health visitors); 32 had more than 10 years of field experience. All owned smartphones, 87% used Google Maps, 96% had a PDF viewer installed, and all preferred Marathi-language outputs. Acceptability was high. The group was the unit of analysis: 12 of the 16 groups said they would use the plan in their next campaign (two further groups were split about 70:30 in favor, one was undecided and one, in a dense urban PCMC area, was opposed); 15 of 16 would recommend it to a colleague; 12 of 16 scored its usefulness at the maximum of 5, with a group median of 5 for both usefulness and ease of use; and 15 of 16 rated it better than the paper-based method.

Two concrete usability problems surfaced. The People-to-Screen target figure shown in the plan was acceptable to only 19% of workers, who read it as an unrealistic quota on top of an already heavy workload. Turn-by-turn navigation failed for 5 of 7 workers in one dense urban unit. The time-and-motion observations of 57 workers found a median plan-access overhead of 54 seconds per hotspot (IQR 42 to 69) among the 37 workers for whom all five manual steps, spread across three separate apps, were timed, and 49 seconds (IQR 40 to 69) across all 57 when summed over the steps recorded. A timed comparison against the paper workflow was not made.

3.2.3 Post-campaign frontline survey

The post-campaign survey of 101 ASHAs, taken after real field use, is the most informative acceptability result. In that survey, 83% agreed the hotspots were genuinely higher-risk, 92% reported reduced workload, 86% reported time saved, 94% found navigation easy, 96% reported finding the same suspects while screening fewer people, and 98% felt adequately trained. Notably, only 41% perceived more presumptive cases in hotspots than in general areas.

> ***Interpretation.*** *The frontline verdict is favorable on efficiency and usability, and the 96% who found the same suspects while screening fewer people is the acceptability counterpart of the NNS reduction. The dissenting 41% figure is an honest counterweight: workers did not consistently perceive a higher presumptive rate in hotspots, which is congruent with the all-forms total-case result in Section 3.3.2 and should temper any claim that targeting is self-evident to the people doing it.*

3.3 Part 3: Prospective field evaluation

In brief. Across the four units, AI-guided teams screened 66,095 people, referred 326 for testing, and diagnosed 15 cases across all forms, of which 10 met the primary NAAT-positive pulmonary definition. On a per-person basis this is close to three times the standard campaign's confirmed-case rate. In absolute, total-case terms the standard campaign, which had a mobile X-ray van the AI sweep lacked, recorded more cases. Both results are reported side by side.

3.3.1 Screening, case ascertainment, and primary endpoint

Field teams screened the populations summarized in Table 3.6, identifying 326 presumptive cases across the four units, of whom 325 were tested and 15 were diagnosed across all forms. Of these, 10 met the primary definition of NAAT-positive pulmonary TB (Shirur 1, Bhor 0, Haveli PMC 7, Haveli PCMC 2). The urban (PMC and PCMC) denominators are drawn from a population-covered field, a different basis from the rural individual-screening denominators, and are reconciled in the pre-specified denominator sensitivity analysis. The full cascade is shown in Figure 2.

Table 3.6. Prospective field results by TB Unit (as of 29 June 2026)

TB Unit	Setting	Screened	Households	Presumptive	Tested	Diagnosed (all forms)	Primary (NAAT+ pulmonary)
Shirur	Rural	3,710	855	31	31	2	1
Bhor	Rural	3,098	815	23	23	0	0
Haveli PMC	Urban	27,047	7,004	204	204	9	7
Haveli PCMC	Urban	32,240	9,143	68	67	4	2
Total		66,095	17,817	326	325	15	10

All-forms cases are TB of any site diagnosed at the referral facility: of the 15, 10 were NAAT-positive pulmonary TB (7 Truenat, 3 CBNAAT), 1 pulmonary case was diagnosed on chest radiography, 3 were pleural TB diagnosed radiologically, and 1 was abdominal TB diagnosed on cytopathology. The primary column counts the 10 NAAT-positive pulmonary cases. Urban denominators are population-covered, a different basis from the rural individual-screening denominators, reconciled in the pre-specified denominator sensitivity analysis. Source: ACF field data extract, 29 June 2026 [S5].

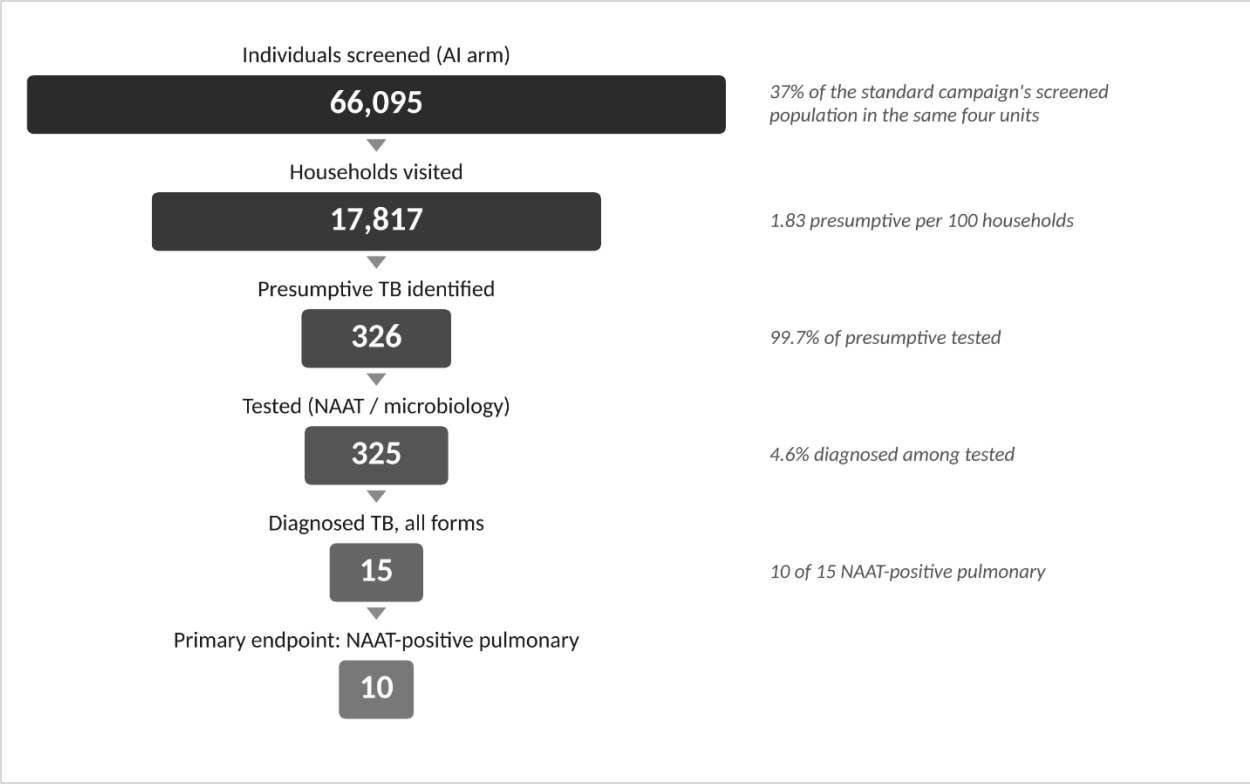


Figure 2. Prospective screening cascade, AI-guided arm.
Source: ACF field data extract, 29 June 2026 [S5].

Under the primary bacteriological pulmonary definition, the AI arm confirmed 10 cases from 66,095 persons screened, a pooled yield of 1.51 per 10,000 (exact Poisson 95% CI 0.73 to 2.78) and a pooled NNS of 6,610 (95% CI 3,594 to 13,783). Yields and NNS by unit are shown in Table 3.7; the per-unit intervals are wide, as expected at small counts, so the pooled estimate carries the interpretive weight.

Table 3.7. Primary outcome: bacteriological yield and NNS by TB Unit (as-screened)

TB Unit	Cases	Screened	Yield per 10,000 (95% CI)	NNS
Shirur	1	3,710	2.70 (0.07 to 15.0)	3,710
Bhor	0	3,098	0 (0 to 11.9)	> 840
Haveli PMC	7	27,047	2.59 (1.04 to 5.33)	3,864
Haveli PCMC	2	32,240	0.62 (0.08 to 2.24)	16,120
Pooled (4 units)	10	66,095	1.51 (0.73 to 2.78)	6,610

Exact (Garwood) Poisson intervals. Bhor recorded zero cases, so its NNS is a lower bound. Source: ACF field data extract, 29 June 2026 [S5]; analysis per Section 2.8.

Compared against the standard blanket ACF campaign in the same units, AI-guided screening confirmed bacteriological pulmonary TB at 2.98 times the standard rate (10 of 66,095 versus 9 of 177,322; exact 95% CI 1.09 to 8.29; exact Poisson p=0.019). The number needed to screen fell from 19,702 under the standard campaign to 6,610 under AI targeting, a 66% reduction, achieved while screening roughly one-third of the persons. The unit-level and pooled comparison is shown in Table 3.8: the advantage is carried by the urban units combined, is large but imprecise in Shirur (one case), and is null in Bhor (zero cases).

Table 3.8. Bacteriological positivity: AI-guided versus standard ACF, by TB Unit (as-screened)

TB Unit	AI cases / screened	Standard cases / screened	Rate ratio (95% CI)	p
Bhor	0 / 3,098	1 / 19,842	0	1.00
Shirur	1 / 3,710	1 / 42,664	11.5 (imprecise)	0.15
Haveli urban (PMC, PCMC)	9 / 59,287	7 / 114,816	2.49 (0.83 to 7.87)	0.069
Pooled (4 units)	10 / 66,095	9 / 177,322	2.98 (1.09 to 8.29)	0.019

Exact Poisson (conditional binomial) rate-ratio test; bacteriologically confirmed pulmonary TB in both arms (NAAT-positive in the AI arm; sputum smear- or NAAT-positive in the TB-4 category of the standard arm). The Haveli urban row pools PMC and PCMC against the campaign's combined Haveli urban units. Per-unit intervals are wide at small counts; the pooled estimate is the primary result. Source: standard-campaign reporting and ACF field data extract, 29 June 2026 [S5].

3.3.2 Total case detection (all forms)

Using a like-for-like, sector-agnostic all-forms definition, the AI arm diagnosed 15 cases and the standard arm 32, a rate ratio of 1.26 (95% CI 0.63 to 2.39, p=0.51), which is not statistically significant. By taluka, all-forms cases were Shirur 2 versus 4, Bhor 0 versus 2, and Haveli 13 versus 26.

> ***Interpretation.*** *This is the total-case-detection result, and it is a null. In absolute terms the standard arm found more than twice as many cases. The dominant reason is a difference in case ascertainment. The standard campaign ran with the mobile X-ray van, which screens by imaging at the point of contact and detects clinically diagnosed pulmonary and extrapulmonary disease that a symptom screen with sputum NAAT does not: 23 of its 32 cases were X-ray-based or extrapulmonary diagnoses. The van was not deployed with the AI sweep, where chest radiography was available only on referral to a health facility, and 5 of the AI arm's 15 cases were diagnosed that way (1 pulmonary and 3 pleural on radiography, 1 abdominal on cytopathology). The per-person positivity advantage in Section 3.3.1 and the null in total detection here are both true and are reconciled by the difference in case ascertainment between the two arms, not by a difference in the underlying targeting.*

3.3.3 Presumptive yield and referral precision

Presumptive yield, expressed as the presumptive rate in the AI arm relative to the standard arm, was pooled RR 1.17 (95% CI 1.03 to 1.34, p=0.018). The effect was concentrated in rural units (RR 2.22, 95% CI 1.62 to 3.00, p<0.001) and absent in urban units (RR 1.01, 95% CI 0.87 to 1.17, p=0.94). Where the urban signal did appear was in referral precision, the proportion of tested presumptives in whom bacteriologically confirmed pulmonary TB was found. In the Haveli urban units this was 9 of 271 (3.3%) in the AI arm against 7 of 523 (1.3%) in the standard arm, a 2.5-fold difference (95% CI 0.93 to 6.59, p=0.067). Pooled across the four units it was 10 of 325 (3.1%) against 9 of 746 (1.2%), 2.6-fold (95% CI 1.05 to 6.22, p=0.043). The full set of rate ratios is shown in Figure 3.

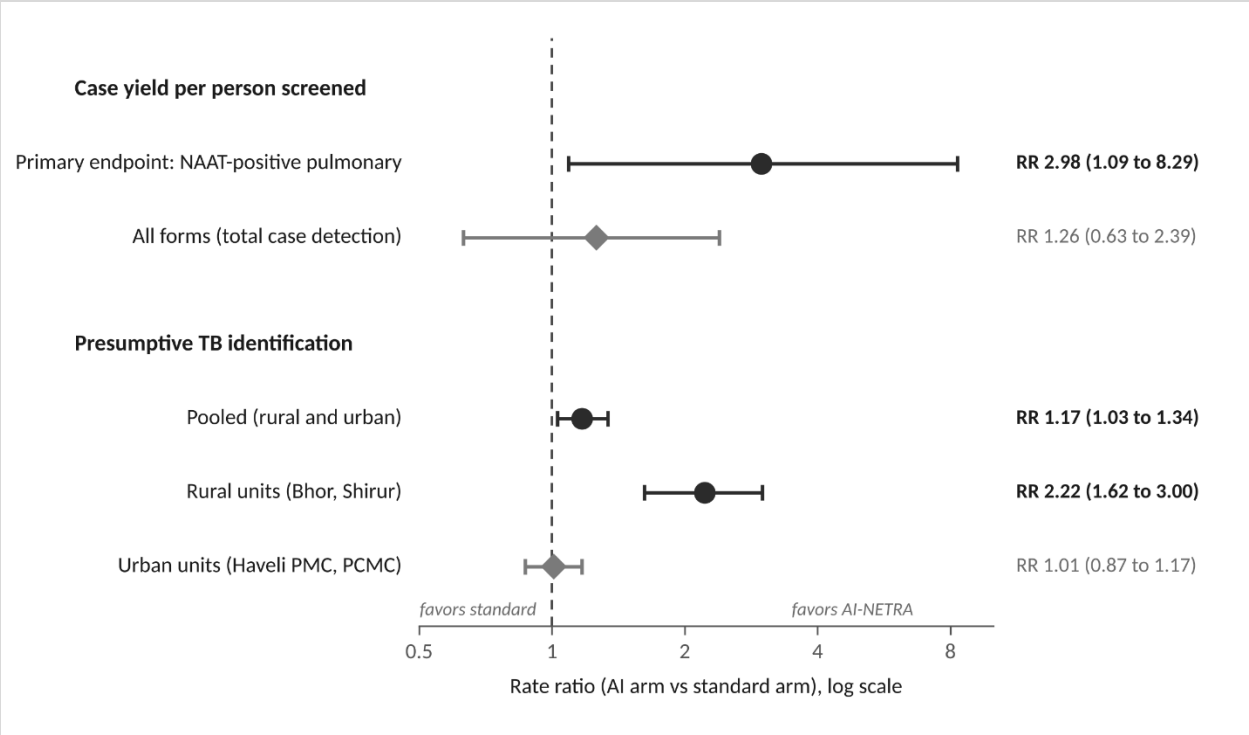


Figure 3. Prospective rate ratios, AI-guided versus standard ACF. Rate ratios above 1 favor AI-guided ACF.
Source: ACF field data extract, 29 June 2026 [S5].

> ***Interpretation.*** *The presumptive picture is two-sided and honest. In rural units, AI-NETRA raised the presumptive rate substantially. In urban units it did not raise the presumptive rate at all, but it improved the quality of referrals so that a given presumptive was about two and a half times as likely to be a true case. The rural yield effect and the urban precision effect are different mechanisms, and both are reported rather than blended into a single headline.*

3.3.4 Screening efficiency and cost

The AI arm needed to screen 6,610 people to confirm one pulmonary case, against 19,702 in the standard arm, a 66% reduction in NNS. Cost per bacteriologically confirmed case fell correspondingly, from Rs. 27.6 lakh in the standard arm to Rs. 9.3 lakh in the AI arm (66% lower). Applying the retrospective concentration to a full district ACF cycle implies a reduction from about Rs. 7.3 crore to about Rs. 3.3 crore [S1]. The AI sweep completed in two weeks, an operational advantage for an on-demand campaign rather than a limitation.

3.4 Sub-study: AI-DNO diagnostic network optimization

In brief. We mapped how TB samples move from collection points to testing laboratories across Pune. The network was uneven: one laboratory was overloaded while eight sat nearly idle, and some samples traveled long distances. By reassigning which laboratory each collection point sends samples to, with no new machines, the average travel distance drops by about a quarter, the overloaded laboratory is relieved, and the district saves about Rs. 22 lakh a year.

The Pune diagnostic network comprises 49 devices (38 TrueNat Duo, 10 GeneXpert four-module, and 1 GeneXpert sixteen-module) across 43 laboratories, processing 53,705 TB tests through 197 sample-transport routes. Overall laboratory utilization was 45.2%, masking wide variation (coefficient of variation 53%): Yamuna Nagar was the sole over-capacity laboratory at 126.4%, while eight laboratories (19%) operated below 25% utilization. Spare capacity at under-utilized laboratories far exceeded the over-capacity backlog, indicating that load balancing through re-routing was feasible without new procurement. The utilization distribution is shown in Table 3.9.

Table 3.9. Laboratory utilization distribution (baseline)

Utilization band	Laboratories	Percentage
Over 100% (over capacity)	1	2%
80 to 100%	4	9%
50 to 80%	18	42%
25 to 50%	12	28%
Under 25% (under-utilized)	8	19%

Coefficient of variation in utilization across laboratories was 53%. Source: AI-NETRA Pune Pilot Study Report [S1].

Optimized referral redesign reassigned 72 of 154 non-onsite routes to geographically closer laboratories, leaving device stock unchanged. The unweighted average service distance fell from 19.5 to 14.7 km (24.9% lower) and the volume-weighted distance from 11.9 to 10.1 km (15.2% lower), mean transit time fell by 7.3 minutes per trip, and routes longer than 30 km fell from 15 to 11. Yamuna Nagar was relieved from 126.4% to 96.2%, leaving no laboratory over capacity. Total annual system cost fell from Rs. 7.89 crore to Rs. 7.67 crore (a 2.8% reduction, about Rs. 22.14 lakh saved), and cost per test from Rs. 1,470 to Rs. 1,428 (Table 3.10).

Table 3.10. AI-DNO referral redesign summary

Metric	Baseline	Optimized	Change
Unweighted ASD (km)	19.5	14.7	24.9% lower
Volume-weighted ASD (km)	11.9	10.1	15.2% lower
Mean transit time per trip (min)	29.3	22.0	7.3 min lower
Routes over 30 km	15	11	27% lower
Yamuna Nagar utilization	126.4%	96.2%	relieved
Annual transport cost (Rs. lakh)	52.0	35.9	30.9% lower
Total annual system cost (Rs. crore)	7.89	7.67	2.8% lower
Cost per test (Rs.)	1,470	1,428	2.9% lower

All gains are achievable through route reassignment alone, with no device procurement or relocation. Source: AI-NETRA Pune Pilot Study Report [S1].

Against an expert panel, model recommendations reached 77.1% concordance (54 of 70 rater-criterion pairs, 95% CI 66.0 to 85.4), produced about three times faster than manual planning. Concordance and efficiency are shown in Figure 4.

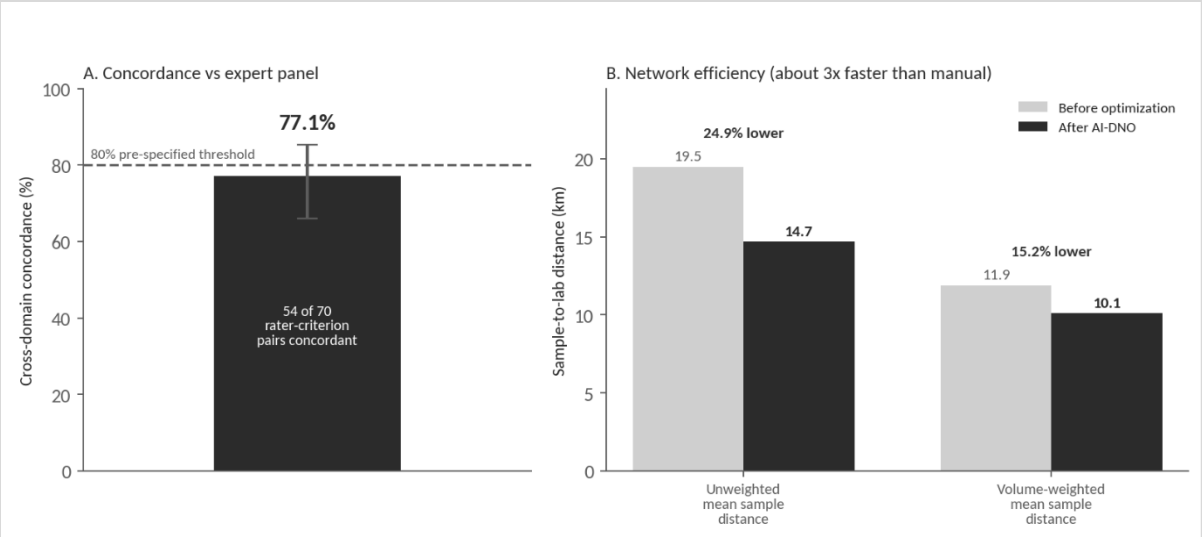


Figure 4. AI-DNO efficiency gains and expert-panel concordance. Source: AI-DNO multi-rater evaluation, Union World Conference on Lung Health Abstract 3160 [S7].

> ***Interpretation.*** *AI-DNO is fast and broadly sensible, but the concordance of 77.1% sits below its own pre-specified 80% acceptance threshold, and the evaluation used two raters whose independence was not formally established. The efficiency gains are real and worth capturing; the concordance result should be read as promising but not yet meeting the bar the team set for it, and is reported as such.*

3.5 Sub-study: cold-start prediction (PDFM)

In cold-start hexagons, PDFM embeddings alone reached an F1 of 0.538 (sensitivity 24.6%, specificity 79.0%), below the naive baseline of F1 0.582 (Wilcoxon p=0.004; Cohen's d of 1.35 in favor of the baseline). Adding weather and air-quality signals in a hybrid improved precision-at-10% from 0.675 to 0.783 (a 16% relative gain) and the Spearman correlation from 0.509 to 0.541. Three independent methods (subspace ablation, with an effect size of 0.335; neural attention; and self-organizing-map topology) converged on weather and air quality as the operative TB-relevant signal.

> ***Interpretation.*** *The cold-start result is directional, not deployable. PDFM alone underperforms a simple baseline, but the environmental signals it helps surface are consistently identified as informative by three separate methods. The finding is limited to a single city and a single year and should be treated as a pointer toward an environmental cold-start layer, not as a validated prediction capability.*

3.6 Mixed-methods integration

Following GRAMMS guidance, the quantitative and qualitative strands are integrated in a joint display (Table 3.11) so that agreement and tension between the numbers and the field experience are visible in one place. [14]

Table 3.11. Mixed-methods joint display

Dimension	Quantitative finding	Field and qualitative finding	Integration
Targeting concentrates cases	Retrospective 89% capture from under a third of the population; prospective RR 2.98	96% of ASHAs found the same suspects while screening fewer people	Convergent
Screening efficiency	NNS fell from 19,702 to 6,610 (66% lower)	92% reported reduced workload and 86% reported time saved	Convergent
Total case detection	All-forms yield did not differ (AI 15 versus standard 32; RR 1.26)	Only 41% perceived more presumptive cases in hotspots	Convergent; both temper the yield story
Micro-targeting resolution	Urban referral precision 2.5 times higher for AI-flagged presumptives	Leadership most wanted 100x100 m landmark hotspot maps, already provided	Convergent
Last-mile usability	Time-motion access overhead 54 s per hotspot; urban navigation failed 5 of 7	People-to-screen quota display acceptable to only 19%	Convergent; both flag field UX

Integration follows GRAMMS guidance for mixed-methods reporting. Sources: Sections 3.1 to 3.4 and the acceptability strand [S1, S5].

4 Discussion

Three conclusions follow from the evidence. First, AI-NETRA's operational value is established. It concentrates screening into higher-positivity pockets, it reduces NNS by roughly two-thirds both retrospectively and prospectively, and the people who use it in the field find it usable and report finding the same suspects while screening fewer people. Second, superiority in total case detection is not established: the like-for-like all-forms comparison is a null (RR 1.26, p=0.51). Third, the per-person confirmed-case advantage (RR 2.98) is real but fragile and confounded, and should be read as hypothesis-generating.

The apparent tension between a strong per-person yield and a null in total detection is explained by the study conditions rather than by the targeting itself. The standard campaign ran with a mobile X-ray van and captured X-ray-based and clinically diagnosed pulmonary and extrapulmonary cases (23 of its 32); the AI sweep ran without the van, with radiography available only on facility referral, and its ascertainment rested mainly on sputum NAAT (10 of its 15 cases). The comparator was also drawn from a different time period, and the urban denominators were constructed differently in the two arms. Each of these pushes the two yield measures in the directions observed, and none of them is a property of the algorithm.

One observation on where AI-NETRA sits within the wider toolkit is worth noting. Other geospatial approaches, such as Wadhwani AI's Vulnerability Mapping, predict TB vulnerability at the level of administrative units and can indicate which blocks or wards to prioritize. AI-NETRA works at a finer grain, resolving risk to H3 hexagons within a prioritized area and turning it into a microplan a worker can follow. The two operate at different scales and are complementary rather than substitutable: a program could reasonably use an administrative-unit vulnerability layer to allocate effort across a district and AI-NETRA to spend that effort efficiently on the ground. [6,7,8]

A second observation concerns where the prospective advantage arose. The primary bacteriological result was carried by the urban units: nine of the ten NAAT-positive pulmonary cases came from Haveli PMC and PCMC, seven and two respectively. This matters because administrative vulnerability layers, including Wadhwani AI’s Vulnerability Mapping, resolve to the block or the ward and do not extend below it, so inside a dense urban ward there is no finer targeting layer and a whole-ward campaign is not operationally feasible. AI-NETRA fills that intra-urban gap. Its urban value in this pilot took the form of referral precision rather than presumptive volume:

the urban presumptive rate did not rise (RR 1.01), but a tested presumptive in the AI arm was about 2.5 times as likely to be confirmed (3.3% versus 1.3%, p=0.067; 2.6 times when the four units are pooled, p=0.043). Both facts are reported together, and the precision gain, which is modest in size and imprecise at these counts, does not obscure the volume null. The urban usability constraints found in the acceptability work, namely turn-by-turn navigation and the People-to-Screen display, bear on urban deployment specifically and are addressed by the field application described in Section 7.3.

An equity consideration runs through the prediction results. AI-NETRA learns from notification history, and its advantage is weakest precisely in the cold-start zones where notifications are sparse, which are often the least-served areas. This is the same gap the PDFM sub-study probes, and it is the reason an environmental cold-start layer is part of the way forward rather than an optional extra: without it, a notification-trained model risks under-serving the areas that most need ACF.

4.1 Toward a precision-targeted saturation model

These results also indicate how AI-NETRA should be deployed, not only how it performs. The central tension in this study, a strong per-person yield alongside a null in total case detection, is the familiar tension between precision targeting and blanket screening, and it is resolved not by choosing one but by sequencing them. A precision-targeted saturation approach, set out separately as a prioritize, saturate, diagnose, and iterate model, positions AI-NETRA as the prioritization layer that decides where community-wide screening is warranted; saturates each prioritized zone to the coverage the evidence indicates is needed to move prevalence; diagnoses efficiently with computer-aided chest X-ray and molecular testing; and re-maps each cycle so the targeting stays current. [S8,13]

Read through that model, the all-forms null is not a verdict against targeting but the expected result of prioritizing without then saturating and without an imaging arm. AI-NETRA identified the higher-yield pockets, but the pilot did not go on to saturate and image them, so the standard campaign, which ran with the mobile X-ray van, recorded more cases in absolute terms. The cold-start finding in Section 3.5 is the same model's answer to notification blindness: by ranking risk on environmental and population-dynamics signals rather than notification history, high-risk but under-notified zones surface for deployment with verification instead of being passed over. The practical implication is that AI-NETRA is most valuable not as a stand-alone efficiency tool but as the first layer of a sequence that ends in saturated, well-diagnosed coverage of the areas it flags.

5 Limitations

- **Non-concurrent comparator.** The standard arm was a campaign from a different time period, so secular trends and seasonal effects are not controlled.
- **Different case ascertainment between arms.** The mobile X-ray van was available in the standard arm but not the AI arm, so the arms did not confirm cases by the same means. This is the principal driver of the divergence between the per-person yield and the total-case result.
- **Different urban denominators.** Urban screening denominators (population-covered for Haveli PMC and PCMC) were constructed differently from the rural household-enumeration denominators, which affects rate comparisons.
- **Implementation wrapper.** The field results reflect AI-NETRA plus worker training plus the microplan workflow as a bundle. They should not be read as the effect of the prediction algorithm in isolation.
- **Fragility of the primary endpoint.** The primary result rests on small case counts. In a leave-one-out sensitivity analysis, dropping any single confirmed case reduces the rate ratio from 2.98 to 2.68 and

the p-value from 0.019 to 0.058, crossing conventional significance. The confidence interval is correspondingly wide (1.09 to 8.29). The primary finding is therefore statistically significant as observed but not robust to the loss of a single event, and is best treated as hypothesis-generating.

- **Equity and cold-start dependence.** Predictive advantage depends on notification history and is weakest in sparsely notified areas, as discussed in Section 4.
- **AI-DNO evaluation.** Concordance (77.1%) fell below the pre-specified 80% threshold, and rater independence was not formally established.
- **Cold-start prediction.** The PDFM sub-study is limited to a single city and year and is directional only.
- **Spatial attribution.** Hexagon-level attribution of cases relied on manual geolocation, which limits the precision of within-area spatial analyses. A geocoded field-capture step would address this.

6 Conclusions

AI-NETRA is a usable and well-received targeting layer that reliably concentrates ACF into higher-positivity neighborhoods, cutting the number needed to screen and the cost per confirmed case by roughly two-thirds. Its per-person confirmed-case yield favored the AI arm in this evaluation, while total case detection did not differ from standard practice in a non-concurrent comparison with unmatched case ascertainment. The operational case for deploying AI-NETRA as an efficiency layer is strong; the case that it finds more cases in absolute terms is not yet made. Establishing the latter requires a concurrent, like-for-like trial in which both arms use the same case-finding tools and definitions. The efficiency and precision case is strongest in dense urban settings, where no administrative targeting layer exists below the ward and blanket ward-wide screening is infeasible; urban wards are therefore a logical early focus for AI-guided ACF planning as an efficiency layer, distinct from the still-open question of absolute case detection.

6.1 For the Central TB Division and national policy

The retrospective and efficiency evidence supports piloting AI-guided ACF planning as an optional efficiency layer within the National TB Elimination Program, particularly in districts where the number needed to screen under blanket campaigns is high and yield is low. The prerequisite is data access: routine export of the geo-coordinates that already exist within Nikshay, so that spatial targeting and hexagon-level case attribution can be computed without bespoke geocoding.

6.2 For state TB programs

For state programs, the immediate value is operational: tighter microplans, lower per-case screening cost, and a diagnostic-network redesign (AI-DNO) that relieves overloaded laboratories and shortens sample-transport distances using existing machines. These gains are realizable now and do not depend on resolving the absolute-case-detection question.

6.3 For funders and research partners

For funders and research partners, the decisive next investment is a concurrent, like-for-like trial with harmonized case definitions and, ideally, cluster randomization, together with authorization for Nikshay hexagon attribution. That single study would convert a strong efficiency signal into a defensible claim about case detection at scale.

7 Way Forward

7.1 A concurrent, like-for-like trial

The decisive next study is a concurrent cluster-randomized (or stepped-wedge) trial that removes the confounders identified here. Both arms should run at the same time, use harmonized case definitions, deploy the mobile X-ray van equally, and receive equivalent training and supervision, so that the only difference between them is the planning method and the trial isolates the model rather than the effort around it. The co-primary outcomes should be total cases found and cases missed in deprioritized areas, so that the trial measures not only what targeting gains but what it might cost in coverage. The trial should be sited in a low-notification district, where the efficiency case is strongest and the equity risk is most testable. It should also include a dense urban arm, where the precision advantage is clearest and both the intra-urban targeting gap and the cold-start equity risk can be examined directly.

7.2 An environmental cold-start layer

To serve areas without notification history, AI-NETRA should be extended with an environmental layer built on weather and air-quality signals, which the PDFM sub-study identifies as informative. This layer is what allows targeting to reach the least-served zones rather than reinforcing existing notification patterns.

7.3 An offline, field-first local language application

The usability findings point to a specific tool. Frontline ACF needs an application that opens directly to the day's plan, navigates by voice in local languages, works fully offline, lets a worker add pockets discovered in the field, writes results back to Nikshay, and captures a geocoded location automatically through a single done action. That geocoding step also resolves the spatial-attribution limitation in Section 5. The People-to-Screen quota display, which most workers found unacceptable, should be removed. The application should be in local language.

7.4 Program enablers

None of the above succeeds without the program conditions that make ACF work: paying ASHA incentives, on time; reimbursing the data costs frontline workers incur using the tool; and restoring reliable access to mobile X-ray so that case ascertainment is complete and comparable across arms.

Competing interests

Dr. Ujjwal Rao is Founder and Chief Executive Officer of Froncort.AI, which developed AI-NETRA and led the analysis and preparation of this report. FIND served as implementation partner for the study, and the State TB Office, Maharashtra, hosted the evaluation. No other competing interests are declared. The role of the funder is described under Funding.

Funding

This work was supported by Google.org under its Health AI grant. The funder had no role in the design or conduct of the study, in the collection, analysis, or interpretation of the data, or in the decision to prepare this report.

Contributions

Dr. Archana Beri served as the Principal Investigator for the AI-NETRA study and provided overall scientific leadership, study conceptualization, technical oversight, interpretation of findings, and guidance on report development. Dr Beri led the study in her former role as Principal Scientist at FIND and currently serves as Head of Programs at The Union.

Dr. Ujjwal Rao, CEO of Froncort.AI, served as Co-Principal Investigator and led the technical development, adaptation, and implementation of the AI-enabled tool used in the study.

The Maharashtra State TB Cell provided overall programme guidance and strategic direction throughout the study. The team contributed valuable insights into the state TB programme, shared relevant programme information and data, supported alignment with NTEP priorities and processes, and facilitated coordination with state, district, and field-level stakeholders.

FIND supported the field implementation of the study, including programme coordination, stakeholder engagement, and operational oversight.

The Froncort.AI team supported the development, deployment, technical support, and field implementation of the AI-NETRA tool.

Acknowledgments

The study team acknowledges the valuable support of Dr R K Pawar, STO, Maharashtra, state and district programme consultants, NTEP officials, participating health facilities, and other stakeholders who contributed to site coordination, field activities, data collection, programme alignment, and implementation.

The study team gratefully acknowledges the support and guidance of Dr Sarabjit Singh Chadha - Country Director India, and Regional Director Asia Programs, Dr. Himanshu Vashistha - Technical Specialist, Praphulla Lakare - Data Analyst, Tajamul Showket - Data Analyst, and Mrinalini Das, M & E Specialist, from FIND who supported development, field coordination and implementation, including liaison with field teams and oversight of study activities.

Special thanks are extended to the Froncort.AI team for their support in tool development, deployment, and implementation throughout the study.

We sincerely acknowledge Mahesh Kolle, the ASHA workers, TB Health Visitors, field staff, and all other frontline personnel whose commitment and participation were central to the successful implementation of the study.

Finally, we thank all study participants and community members for their time, cooperation, and contribution to the AI-NETRA study.

8 References

8.1 Study documents

[S1] FIND, Froncort.AI. AI-NETRA Pune Pilot Study Report (Parts 1 and 2: retrospective simulation and acceptability). 4 January 2026. Internal study report.

[S2] FIND, Froncort.AI. Frontline FGD Analysis Draft Report, Version 3.0 (COREQ-structured). 4 June 2026. Internal qualitative analysis.

[S3] FIND, Froncort.AI. FGD and Time-Motion Data Analysis Tool (compiled focus-group and observation data). 18 May 2026. Internal dataset.

[S4] FIND, Froncort.AI. Post-campaign frontline (ASHA) acceptability survey, n=101. 2026. Internal dataset.

[S5] FIND, Froncort.AI. AI-NETRA prospective ACF field data extract. 29 June 2026. Final dataset.

[S6] Froncort.AI and Google Research. Population Dynamics Foundation Model for TB risk prediction: Pune cold-start zone evaluation (v4). 14 April 2026. Internal evaluation report.

[S7] FIND, Froncort.AI. AI-DNO diagnostic network optimization: multi-rater expert evaluation. Abstract 3160, Union World Conference on Lung Health, 2025. Conference abstract.

[S8] Froncort.AI. Precision-Targeted Saturation for TB Active Case Finding: a prioritize, saturate, diagnose, and iterate concept note. June 2026. Internal strategy note.

8.2 Cited literature

[1] World Health Organization. Global Tuberculosis Report 2024. Geneva: WHO; 2024.

[2] Central TB Division, Ministry of Health and Family Welfare, Government of India. India TB Report 2024. New Delhi: MoHFW; 2024.

[3] World Health Organization. WHO consolidated guidelines on tuberculosis. Module 2: Systematic screening for tuberculosis disease. Geneva: WHO; 2021.

[4] Burugina Nagaraja S, Thekkur P, Satyanarayana S, et al. Active case finding for tuberculosis in India: a synthesis of activities and outcomes reported by the National Tuberculosis Elimination Program. Trop Med Infect Dis. 2021.

[5] Uber Technologies. H3: a hexagonal hierarchical geospatial indexing system. Open-source documentation. Available at: https://h3geo.org

[6] Wadhwani AI. Vulnerability Mapping (VMTB) programme page. Available at: https://www.wadhwaniai.org/impact/healthcare-solutions/vulnerability-mapping/ (accessed 2026).

[7] Wadhwani AI. Healthcare programme overview. Available at: https://www.wadhwaniai.org/impact/healthcare/ (accessed 2026).

[8] Press Information Bureau, Government of India. TB Division of the Health Ministry signs MoU to explore Artificial Intelligence (AI) based solutions in combating TB. August 2019. Available at: https://www.pib.gov.in/PressReleasePage.aspx?PRID=1583584

[9] von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement. Ann Intern Med. 2007;147(8):573-577.

[10] Des Jarlais DC, Lyles C, Crepaz N; TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health. 2004;94(3):361-366.

[11] Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine-learning methods. BMJ. 2024;385:e078378.

[12] Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;19(6):349-357.

[13] Marks GB, Nguyen NV, Nguyen PTB, et al. Community-wide screening for tuberculosis in a high-prevalence setting (ACT3 trial). N Engl J Med. 2019;381(14):1347-1357.

[14] O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research (GRAMMS). J Health Serv Res Policy. 2008;13(2):92-98.

Appendix A. Reporting-standard checklists

This report follows four reporting standards. The tables below map the main item areas of each standard to the sections in which they are addressed. Item wording is paraphrased; the full checklists are cited in Section 8.2. Full field instruments (the FGD guide, the frontline survey instrument) and a protocol and methods summary are available on request and can be appended to this report when provided.

A.1 STROBE (observational analyses)

Item area	Addressed in
Structured title and abstract	Title page; Abstract
Background and rationale; objectives	Sections 1 and 1.2
Study design, setting, and participants	Sections 2.1, 2.5, 3.3.1
Variables and outcome (case) definitions	Sections 2.5, 3.3.1
Statistical methods; sensitivity analyses	Section 2.8; Section 5 (fragility)
Descriptive and outcome results	Section 3; Tables 3.1 to 3.11; Figures 1 to 4
Limitations and generalizability	Sections 5 and 6
Funding	Funding section

A.2 TREND (quasi-experimental comparison)

Item area	Addressed in
Intervention and comparison condition	Section 2.5
Theory or mechanism of the intervention	Sections 1 and 2.2
Assignment units and comparison groups	Section 2.5
Outcome measures and case ascertainment	Sections 2.5, 3.3
Sample sizes and denominators	Section 3.3.1; Figure 2
Baseline comparability and confounding	Sections 2.8 and 5
Effect estimates with precision	Section 3.3; Figure 3
Subgroup analyses (rural and urban)	Section 3.3.3
Limitations of nonrandomized design	Section 5

A.3 TRIPOD+AI (prediction model)

Item area	Addressed in
Source of data and setting	Sections 2.1 and 2.3
Predictors and inputs	Section 2.2
Outcome predicted	Sections 2.2 and 2.3
Model development and refinement	Sections 2.2 and 2.3
Performance measures (rank correlation, MAE, capture)	Section 3.1; Figure 1
Validation (retrospective and cold-start)	Sections 3.1 and 3.5
Fairness and equity considerations	Sections 3.5 and 4
Intended use and limitations	Sections 2.2 and 5

A.4 COREQ (focus group discussions)

Item area	Addressed in
Research team and reflexivity	Section 2.4; [S2]
Sampling and participants (16 groups, n=100)	Section 2.4
Setting and data collection (17 Apr to 11 May 2026)	Section 2.4
FGD guide	Available on request (see appendix note)
Analysis approach	Section 2.4; [S2]
Reporting of findings	Section 3.2
Group composition and saturation	[S2]