WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Digital Health Transformation in Primary Healthcare: A 12-Month Multicentre Evaluation of AI-Enabled Telemedicine Across Three Primary Health Centres in Karimnagar District, Telangana, India

***Atul Gopal Agrawal*[1,*] *Ankita Agrawal*[2]**

[1] Professor and Head, Department of Community Medicine, Surabhi Institute of Medical Sciences, Siddipet, Telangana, India (MBBS, MD, PGDAST, PGDMCH, PGDHHM, ACME, MBA, CCH, CAI, CSD)

[2] Consultant Physiotherapist and Researcher, Department of Physiotherapy, Shree Sathya Sai Sanjeevani Mother and Child Hospital, Yavatmal, Maharashtra, India (MPT, Orthopaedics)

***Corresponding author:** Dr. Atul Gopal Agrawal, Department of Community Medicine, Surabhi Institute of Medical Sciences, Siddipet, Telangana, India. E-mail: dratulu9@gmail.com

Study design: *Prospective multicentre implementation evaluation with a historical (before-and-after) comparator. Reporting guidance: STROBE; DECIDE-AI principles.*

KEY POINTS

Question. Can an AI-enabled telemedicine pathway linking three rural primary health centres to a district hospital sustain high-volume, specialist-supported care in routine practice?

Findings. Over 12 months, 48,650 encounters (34,120 individuals) were recorded; 83.0% used AI-assisted triage and 85.4% were resolved at the PHC. Among 312 STEMI cases, crude mortality was 4.8% versus 14.1% historically (unadjusted RR 0.34, 95% CI 0.19–0.60).

Meaning. The pathway was feasible at scale and associated with faster operations and lower patient expenditure; causal and safety claims require task-specific diagnostic validation and controlled, adjusted evaluation.

ABSTRACT :

Background: Rural primary care in India is constrained by limited access to diagnostics and specialist review. Artificial intelligence (AI)-assisted decision support combined with telemedicine may extend specialist capacity to peripheral facilities, but evidence from routine implementation remains limited.

Objective: To describe utilisation, care disposition, operational performance, ST-elevation myocardial infarction (STEMI) pathway indicators and patient expenditure during a 12-month AI-enabled telemedicine implementation across three primary health centres (PHCs) in Karimnagar District, Telangana, India.

Methods: Prospective multicentre implementation evaluation (1 January–31 December 2024) at PHC Choppadandi, PHC Gangadhara and PHC Manakondur (catchment 135,800), linked to a district tertiary-care hospital. Aggregate programme data covered 34,120 individuals and 48,650 encounters. Proportions are presented with Wilson 95% confidence intervals (CIs); crude STEMI mortality was compared with a historical cohort using risk ratio, Newcombe risk-difference CI and χ^2 test.

Results: There were 1.43 encounters per individual (358 per 1,000 catchment population). AI-assisted triage was recorded in 83.0% (95% CI 82.7–83.3), PHC-level resolution in 85.4% (85.0–85.7) and referral in 14.6% (14.3–15.0). Mean consultation turnaround fell from 48.2 to 11.4 minutes. Among 312 STEMI cases, 285 were thrombolysed (91.3%; 87.7–94.0); crude mortality was 4.8% (2.9–7.8) versus 14.1% (10.6–18.5) in 305 historical cases (risk ratio 0.34, 95% CI 0.19–0.60; risk difference −9.3 percentage points, −14.0 to −4.7; $p<0.001$), and door-to-needle time fell from 142 to 38 minutes. Mean out-of-pocket expenditure fell from ₹1,850 to ₹585 per episode (−68.4%). Pooled sensitivity (94.8%) and specificity (92.6%) were reported without task-specific contingency tables; maternal, neonatal and referral-completion outcomes could not be quantified.

Conclusions: The pathway achieved high utilisation, rapid consultations and a favourable crude STEMI signal in rural primary care. Because the design was uncontrolled and diagnostic validation incomplete, these findings are hypothesis-generating. Task-specific validation, complete follow-up and adjusted or interrupted time-series evaluation are required before conclusions on effectiveness, safety or scalability.

Keywords: *artificial intelligence; telemedicine; primary health care; rural health; diagnostic decision support; STEMI; digital health; implementation evaluation; India*

1. Introduction

Timely access to diagnostics and specialist opinion remains a persistent challenge in rural health systems. Distance, transport costs, uneven workforce distribution and fragmented referral pathways delay care and impose catastrophic expenditure on households [1–3]. Telemedicine can connect peripheral facilities with higher-level clinicians, while AI-enabled tools may support prioritisation of selected clinical inputs such as chest radiographs and electrocardiograms (ECGs) [4–6]. Their real-world value, however, depends on workflow integration, clinician oversight, patient trust and reliable escalation pathways rather than on algorithmic accuracy alone [7,8].

India has built substantial telemedicine infrastructure. The eSanjeevani platform provides provider-to-provider and patient-to-provider teleconsultation, and national Telemedicine Practice Guidelines define professional responsibility, consent, documentation and privacy [9,10]. Governance of health AI has simultaneously matured: WHO and the Indian Council of Medical Research emphasise autonomy, safety, transparency, accountability, inclusiveness, human oversight and responsible data use [11–14]; the Digital Personal Data Protection Act, 2023 sets the statutory data framework [15]; and the Ministry of Health and Family Welfare launched the Strategy for Artificial Intelligence in Healthcare for India (SAHI) and the Benchmarking Open Data Platform (BODH) in February 2026 [16]. The WHO Global Strategy on Digital Health has been extended to 2027 [17].

Despite this policy momentum, empirical reports from routine primary-care deployments remain scarce, and many AI studies are limited by weak design and over-optimistic claims [8]. Pooled accuracy figures can conceal differences across clinical tasks, and before-and-after comparisons are vulnerable to temporal confounding. Transparent reporting is therefore essential: STROBE supports observational reporting [18]; STARD-AI addresses reporting of diagnostic accuracy studies of AI [19]; DECIDE-AI guides early-stage clinical evaluation of AI decision support [20]; CLAIM covers imaging AI [21]; and CHEERS 2022 guides economic evaluation [22].

We therefore evaluated a 12-month AI-enabled telemedicine implementation across three PHCs in Karimnagar District, Telangana. Our objectives were to (i) quantify utilisation and care disposition; (ii) describe operational and STEMI-pathway indicators against pre-implementation values; (iii) describe patient expenditure; and (iv) identify precisely which evidence is still needed for a reproducible, policy-grade evaluation.

RESEARCH IN CONTEXT: WHAT IS ALREADY KNOWN	**WHAT THIS STUDY ADDS**
• AI tools and telemedicine are increasingly promoted for rural primary care, but most evidence concerns algorithm accuracy rather than routine service delivery. • Reporting quality of AI studies is often weak, and pooled accuracy figures can mislead. • India has national telemedicine and AI-governance frameworks, but few published PHC-level implementation evaluations.	• A 12-month, three-PHC evaluation (48,650 encounters) with reconciled aggregate indicators and transparent confidence intervals. • Crude STEMI mortality of 4.8% versus 14.1% historically (RR 0.34), with sensitivity analysis showing the estimate is large but not causal. • An explicit map of evidence gaps (task-specific accuracy, maternal and neonatal outcomes, referral completion, costs).

2. Methods

2.1 Study design and reporting standards

We conducted a prospective multicentre implementation evaluation from 1 January to 31 December 2024, using pre-implementation (historical) values as a descriptive comparator for selected operational, clinical and economic indicators. The report follows STROBE [18] and, for the AI-supported pathway, DECIDE-AI principles [20]. Because only aggregate programme data were analysed, STARD-AI and CLAIM items on diagnostic accuracy could be addressed only partially (Section 4.3).

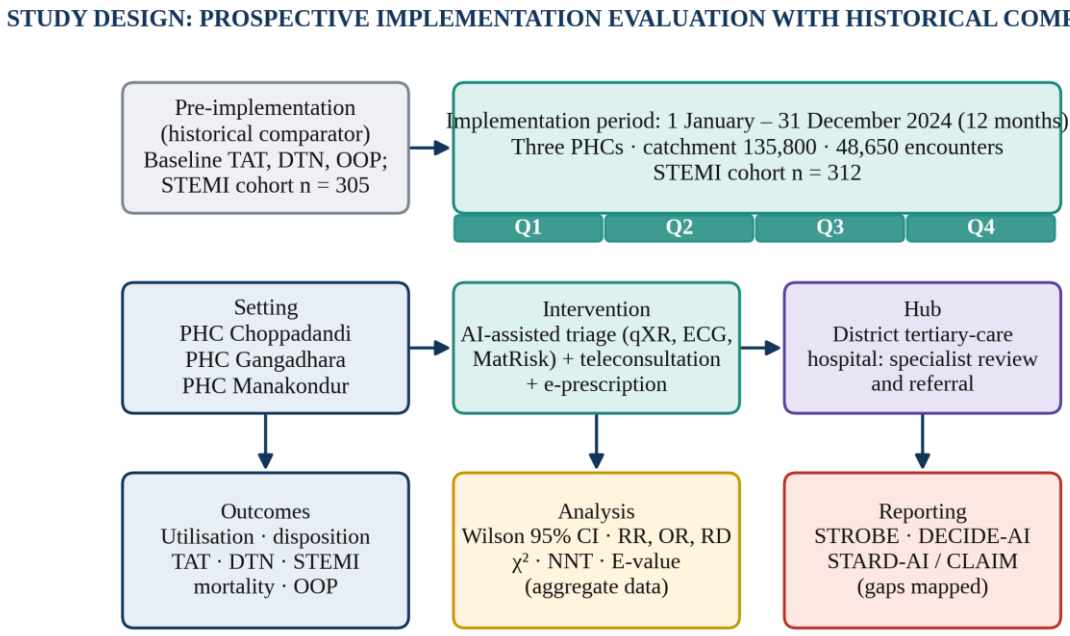


Figure 1. Study design schematic. Prospective implementation evaluation (1 January–31 December 2024) across three PHCs linked to a district hub, with a historical comparator for operational, STEMI and expenditure indicators. TAT, consultation turnaround time; DTN, door-to-needle time; OOP, out-of-pocket expenditure; NNT, number needed to treat.

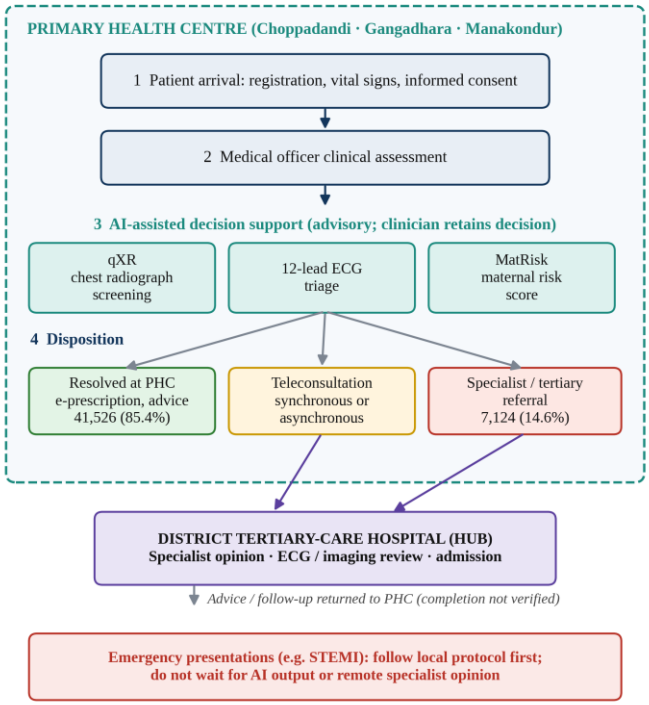
2.2 Setting

The study was conducted at PHC Choppadandi, PHC Gangadhara and PHC Manakondur in Karimnagar District, Telangana, with a combined catchment population of 135,800. The Karimnagar District tertiary-care hospital served as the specialist teleconsultation and referral hub.

2.3 Intervention architecture

The service comprised AI-assisted chest radiograph screening (qXR module), 12-lead ECG triage, maternal risk scoring (MatRisk module), synchronous or asynchronous specialist teleconsultation and electronic prescription transmission (Figure 2). AI outputs were designed to support, not replace, clinical assessment by the medical officer and specialist. Emergency presentations followed established local protocols and were not delayed pending AI output or remote opinion.

Figure 2. Care pathway and intervention architecture of the AI-enabled telemedicine service. Percentages are shares of all 48,650 encounters. The diagram is conceptual and does not represent verified referral completion.



2.4 Population and units of analysis

Aggregate programme data comprised 34,120 unique individuals and 48,650 encounters. Encounter-level indicators use 48,650 as denominator unless stated. The mean number of encounters per individual was 48,650/34,120 = 1.43; this mean does not describe the distribution of repeat visits. Person- versus encounter-level denominators for demographic summaries are specified in Table 2.

2.5 Outcomes and operational definitions

Service outcomes were encounter volume, AI-assisted triage, PHC-level resolution and specialist/tertiary referral. Operational outcomes were consultation turnaround time and STEMI door-to-needle time. Clinical outcomes were thrombolysis and in-hospital crude mortality among 312 STEMI cases during implementation, compared with 305 historical cases. Economic outcome was mean out-of-pocket expenditure per episode. Maternal mortality, neonatal mortality and referral completion were prespecified as relevant outcomes, but verified numerators, denominators, case definitions and follow-up windows were not available; no numerical estimates are presented for them.

Table 1. Operational definitions of study indicators

Indicator	Operational definition	Denominator
Unique individual	A person counted once during the study year irrespective of the number of visits	Catchment population
Encounter	A recorded clinical contact at a participating PHC during the study period	Encounters (48,650)
AI-assisted triage	Encounter in which at least one AI module output (qXR, ECG triage or MatRisk) was available to the clinician	Encounters
PHC-level resolution	Encounter completed at the PHC without specialist or tertiary referral	Encounters
Referral	Encounter in which specialist or tertiary-level care was advised	Encounters
STEMI case	ST-elevation myocardial infarction diagnosed on clinical and ECG criteria [23]	STEMI cases
Thrombolysis	Administration of fibrinolytic therapy to a STEMI case	STEMI cases
Door-to-needle time	Minutes from arrival at the facility to start of fibrinolytic therapy	STEMI cases treated
Consultation turnaround time	Minutes from consultation request to completion of the consultation	Consultations
Out-of-pocket expenditure	Mean patient expenditure per episode of care (travel, consultation, investigations, medicines), in Indian rupees	Episodes
Crude STEMI mortality	Deaths among STEMI cases as a proportion of cases; follow-up window not specified in the programme records	STEMI cases

Definitions are the standard forms applied in the analysis. Where programme records did not specify a window (for example mortality follow-up), this is stated as a limitation.

2.6 Ethics and consent

The study was approved by the Institutional Ethics Committee before commencement (reference IEC/PIMS/2023/114). Informed consent was obtained in Telugu, Hindi or English according to participant preference. No identifiable data are presented.

2.7 Sample size and precision

No a priori sample size calculation was performed because all consecutive encounters during the study year were included in this programme evaluation. Precision is reported through confidence intervals. For the STEMI comparison we additionally report, post hoc, the smallest mortality that would have been detectable with 80% power (two-sided $\alpha = 0.05$) against the historical 14.1% with 312 and 305 cases; observed (post hoc) power is not interpreted because it is a function of the p-value.

2.8 Statistical analysis

Categorical variables are presented as counts and percentages with Wilson score 95% CIs, which perform well for proportions near the boundaries [24]. For the STEMI contrast we calculated the risk ratio (log-normal CI), odds ratio, risk difference with Newcombe hybrid-score 95% CI [25], relative risk

reduction, number needed to treat (reciprocal of the risk difference) and Pearson χ^2 (with Yates correction and Fisher exact test as sensitivity checks). Before-and-after differences in turnaround time, door-to-needle time and expenditure are absolute and relative changes. Reported pooled sensitivity and specificity were reproduced as supplied; we derived likelihood ratios, Youden index, predictive values at assumed prevalences and the sample size implied by the CI width ($n \approx z^2p(1-p)/h^2$, h = half-width). To probe robustness to unmeasured confounding of the crude STEMI contrast we computed the E-value, the minimum strength of association that an unmeasured confounder would need with both period and death to explain away the observed risk ratio [26]. All aggregate figures were independently re-computed and reconciled (Appendix, Table A1).

No inferential model adjusted for repeated encounters, PHC clustering, case mix or secular trends because individual-level data were unavailable. p-values for the STEMI contrast therefore describe crude association only. Analyses used Python 3 (SciPy).

2.9 Data governance

Data handling followed institutional requirements and the Digital Personal Data Protection Act, 2023 [15]. A submission-grade AI evaluation should additionally document data sources, missingness, case definitions, index-test thresholds, reference standards, blinding, software versions, adverse events and audit trails; the gaps against this standard are mapped in Section 5.3.

3. Results

3.1 Encounter volume and centre profile

The programme recorded 48,650 encounters among 34,120 unique individuals (25.1% of the catchment population; 1.43 encounters per individual), equivalent to 358 encounters per 1,000 catchment population (Figure 3). Choppadandi contributed 17,420 encounters (35.8%), Gangadhara 15,880 (32.6%) and Manakondur 15,350 (31.6%) (Table 2, Figure 4). Utilisation rates ranged from 340 per 1,000 (Manakondur) to 376 per 1,000 (Gangadhara); the distribution of encounters differed from that expected under catchment-proportional use ($\chi^2 = 78.8$, df = 2, $p < 0.001$), although repeat encounters limit formal inference. The weighted mean age across centres (41.5 years) and female proportion (53.8%) reconciled with the reported overall values. Per 1,000 catchment population there were 297 AI-assisted encounters and 52.5 referrals.

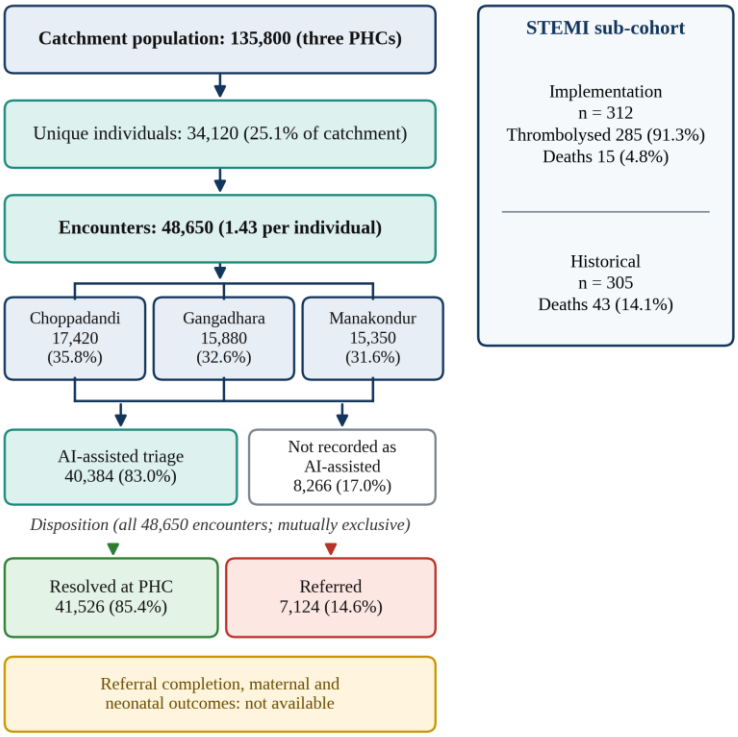


Figure 3. Flow of the catchment population, individuals, encounters, service disposition and STEMI sub-cohort. This is an aggregate encounter-flow diagram rather than a patient-level CONSORT/STROBE flow; exclusions, missingness and follow-up were not available.

Table 2. Centre profile and utilisation, 1 January–31 December 2024

PHC	Catchment population	Encounters (n)	Share of encounters, % (95% CI)	Encounters per 1,000 catchment	Mean age (years)	Female (%)
Choppadandi	48,500	17,420	35.8 (35.4–36.2)	359.2	41.2	53.4
Gangadhara	42,200	15,880	32.6 (32.2–33.1)	376.3	39.8	56.1
Manakondur	45,100	15,350	31.6 (31.1–32.0)	340.4	43.5	51.8
Total / overall	**135,800**	**48,650**	**100.0**	**358.2**	**41.5 ± 13.5**	**53.8**

Percentages are encounter-level shares; Wilson 95% CIs describe sampling precision for encounter shares only and do not account for repeat visits. Centre-specific age is given as the mean; the overall mean age is 41.5 ± 13.5 years. Demographic summaries are presented as supplied and are encounter-level unless confirmed otherwise.

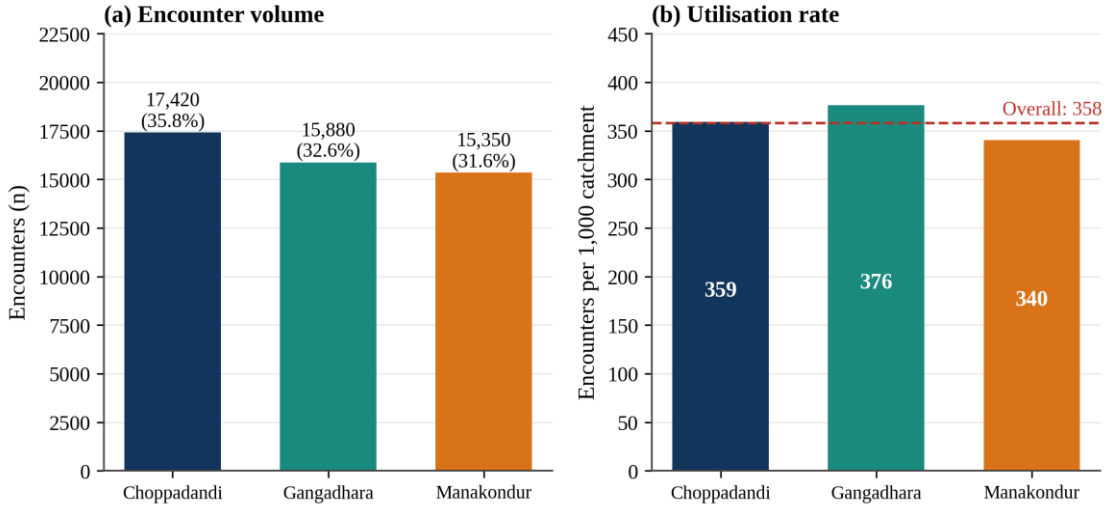


Figure 4. (a) Encounter volume by PHC with share of total encounters. (b) Utilisation rate per 1,000 catchment population; the dashed line marks the three-centre average.

3.2 Service utilisation and disposition

AI-assisted triage was recorded in 40,384/48,650 encounters (83.0%; 95% CI 82.7–83.3). PHC-level resolution occurred in 41,526/48,650 (85.4%; 85.0–85.7) and specialist or tertiary referral in 7,124/48,650 (14.6%; 14.3–15.0); resolution and referral are mutually exclusive and sum to 100% (Table 3, Figure 5). These indicators describe recorded disposition; they do not establish referral appropriateness, completion or safety.

Table 3. Service indicators with Wilson 95% confidence intervals

Indicator	Numerator / denominator	% (95% CI)	Interpretive note
Unique individuals	34,120 / 135,800	25.1	Person-level penetration of catchment
Encounters per individual	48,650 / 34,120	1.43 (mean)	Arithmetic mean only
AI-assisted triage	40,384 / 48,650	83.0 (82.7–83.3)	Use indicator, not accuracy
PHC-level resolution	41,526 / 48,650	85.4 (85.0–85.7)	Definition to be confirmed
Specialist/tertiary referral	7,124 / 48,650	14.6 (14.3–15.0)	Completion not available
STEMI thrombolysis	285 / 312	91.3 (87.7–94.0)	Implementation period only

CI, confidence interval (Wilson score method).

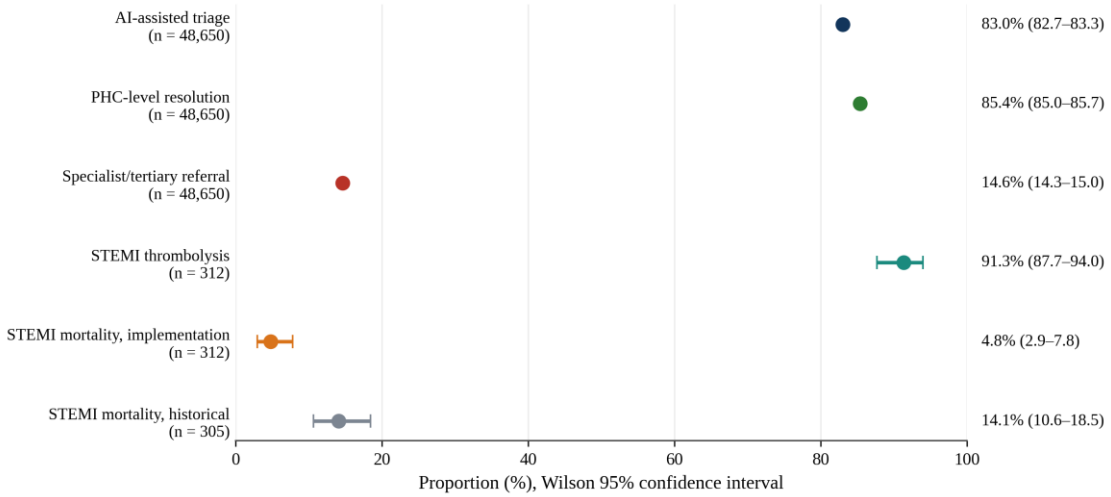


Figure 5. Forest-style display of key proportions with Wilson 95% confidence intervals. Narrow intervals for service indicators reflect large denominators, not validity of measurement.

3.3 STEMI pathway

Among 312 STEMI cases during implementation, 285 received thrombolysis (91.3%; 95% CI 87.7–94.0) and 15 died (4.8%; 2.9–7.8). Historical records showed 43 deaths among 305 cases (14.1%; 10.6–18.5). The crude risk ratio was 0.34 (95% CI 0.19–0.60), the odds ratio 0.31 (0.17–0.57) and the risk difference −9.3 percentage points (95% CI −14.0 to −4.7); Pearson $\chi^2 = 15.63$ (df = 1, $p < 0.001$; Yates-corrected $\chi^2 = 14.56$, $p < 0.001$; Fisher exact $p < 0.001$). The crude relative risk reduction was 65.9%, corresponding to a number needed to treat of approximately 11 (95% CI 7–21) (Table 4, Figure 6). Door-to-needle time was 38 minutes during implementation versus 142 minutes before, a 104-minute (73.2%) reduction.

This contrast is unadjusted and hypothesis-generating. Case definitions, eligibility, severity, follow-up duration, outcome ascertainment and concurrent changes in referral patterns or treatment protocols were not demonstrably comparable between periods, so the estimate cannot be interpreted as an attributable survival benefit.

Table 4. STEMI outcomes: implementation period versus historical comparison

Measure	Implementation	Historical	Contrast (95% CI)
STEMI cases (n)	312	305	Comparability not established
Thrombolysis, n/N (%)	285/312 (91.3)	Not supplied	Implementation only
Deaths, n/N (%)	15/312 (4.8)	43/305 (14.1)	—
Risk ratio			0.34 (0.19–0.60)
Odds ratio			0.31 (0.17–0.57)
Risk difference (percentage points)			−9.3 (−14.0 to −4.7)
Relative risk reduction			65.9%
Number needed to treat (descriptive)			11 (7–21)
χ^2 test (df = 1)			15.63; $p < 0.001$
Door-to-needle time (min)	38	142	104 min shorter (73.2%)

Unadjusted, crude comparison. Newcombe hybrid-score CI for the risk difference; log-normal CI for risk and odds ratios. NNT = 100/|risk difference in percentage points|; its CI is the reciprocal of the risk-difference CI limits.

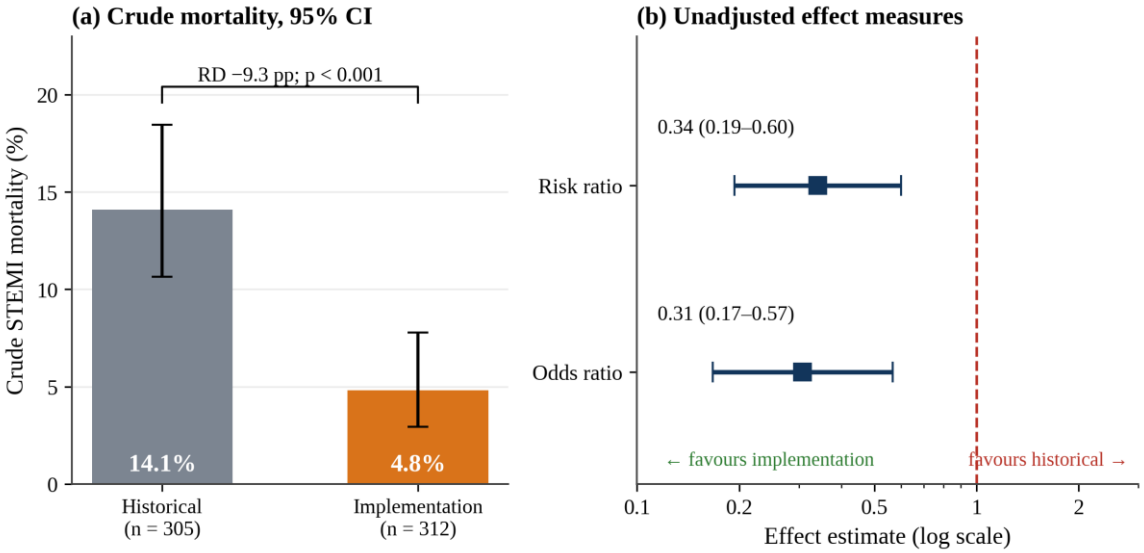


Figure 6. STEMI mortality. (a) Crude mortality with Wilson 95% CIs. (b) Unadjusted risk and odds ratios with 95% CIs on a logarithmic scale; values below 1 favour the implementation period. The comparison is uncontrolled.

3.4 Sensitivity analysis and precision of the STEMI contrast

The E-value for the crude risk ratio of 0.34 was 5.31: an unmeasured confounder would need to be associated with both the study period and mortality by a risk ratio of at least 5.31 each, above and beyond measured covariates, to fully explain away the point estimate; for the upper confidence limit (RR 0.60) the E-value was 2.72 (Table 5, Figure 7) [26]. These values indicate that moderate confounding alone is unlikely to account for the entire contrast. However, the E-value does not address selection bias, non-comparable case definitions, differences in outcome ascertainment or incomplete follow-up in the historical cohort, which are the principal threats here. With 312 and 305 cases the study had 80% power to detect a mortality of 7.1% or lower against the historical 14.1%, i.e. an absolute reduction of at least 7.0 percentage points; the observed 9.3-point difference exceeds this threshold, but the lower confidence limit of the risk difference (4.7 points) shows that smaller true effects remain compatible with the data.

Table 5. Sensitivity analysis and precision of the crude STEMI mortality contrast

Measure	Value	Interpretation
Crude risk ratio (95% CI)	0.34 (0.19–0.60)	Lower mortality in implementation period
E-value, point estimate	5.31	Confounder RR with period and death needed to nullify
E-value, upper CI limit	2.72	Confounder strength needed to move CI to include 1
Minimum detectable mortality (80% power)	7.1%	Versus historical 14.1%; n = 312 vs 305
Observed absolute reduction	9.3 pp (4.7–14.0)	Exceeds detectable threshold; lower bound compatible with smaller effects

E-value computed as RR + $\sqrt{RR(RR-1)}$ *where* $RR = 1/RR$ for protective associations. pp, percentage points.

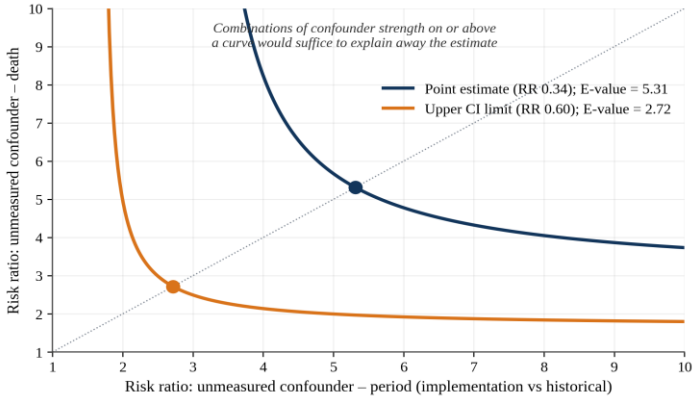


Figure 7. E-value sensitivity plot for the crude STEMI risk ratio. Curves show combinations of confounder associations (with period and with death) that could fully explain the observed estimate (navy) or shift the upper confidence limit to the null (orange). Filled circles mark the E-values.

3.5 Operational efficiency and patient expenditure

Mean consultation turnaround time was 11.4 ± 1.5 minutes during implementation versus 48.2 ± 6.1 minutes before, an absolute reduction of 36.8 minutes (76.4%). Mean out-of-pocket expenditure was ₹585 versus ₹1,850 per episode, a difference of ₹1,265 (68.4% lower) (Table 6, Figure 8). Multiplying ₹1,265 by 48,650 encounters yields an indicative arithmetic figure of ₹61,542,250 (≈ ₹6.15 crore) in patient-level savings. This extrapolation assumes the episode-level difference applies uniformly to every encounter and is not a cost-effectiveness estimate: programme costs, distributions and uncertainty were unavailable [22].

Table 6. Before-and-after operational and economic indicators

Indicator	Before	During implementation	Absolute change	Relative change
Consultation turnaround (min), mean ± SD	48.2 ± 6.1	11.4 ± 1.5	−36.8	−76.4%
STEMI door-to-needle (min)	142	38	−104	−73.2%
Out-of-pocket expenditure (₹ per episode)	1,850	585	−1,265	−68.4%
Indicative aggregate patient saving (₹)	—	—	61,542,250 (≈ ₹6.15 crore)	Arithmetic extrapolation

The standardised difference in turnaround time (≈ 8.3 pooled SDs) is implausibly large for field data; the SDs may reflect between-centre rather than between-encounter variation and should be verified. No inferential test is reported because sample sizes for the pre-implementation values were unavailable.

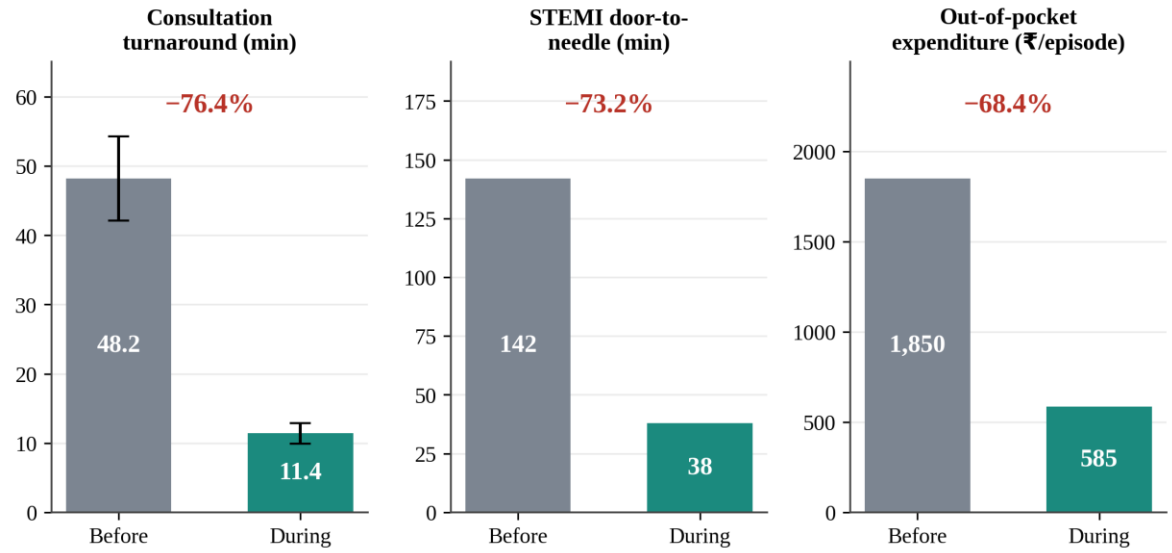


Figure 8. Before-and-after comparison of consultation turnaround time (error bars, ±1 SD), STEMI door-to-needle time and out-of-pocket expenditure, with relative reductions.

3.6 Reported diagnostic performance

Pooled sensitivity was reported as 94.8% (95% CI 93.9–95.6) and specificity as 92.6% (91.8–93.4). From these values, the positive likelihood ratio is 12.8, the negative likelihood ratio 0.06 and the Youden index 0.87. The CI widths imply approximately 2,620 diseased and 4,110 non-diseased reference-standard cases, far fewer than 48,650 encounters, consistent with accuracy being assessed in a tested subset. Task-specific denominators, thresholds, reference standards and 2×2 tables for chest radiography, ECG triage and maternal risk scoring were not available; these figures are therefore unverified aggregate summaries and not validated estimates for any single task [19,21]. Predictive values depend strongly on prevalence: assuming the reported accuracy, the positive predictive value would be 20.7%, 40.3% and 58.7% at prevalences of 2%, 5% and 10%, with negative predictive values above 99% (Table 7, Figure 9).

Table 7. Reported pooled diagnostic performance and derived measures

Measure	Value	Basis
Sensitivity, % (95% CI)	94.8 (93.9–95.6)	Reported
Specificity, % (95% CI)	92.6 (91.8–93.4)	Reported
Positive likelihood ratio	12.8	Derived: Se/(1−Sp)
Negative likelihood ratio	0.06	Derived: (1−Se)/Sp
Youden index	0.87	Derived: Se+Sp−1
Implied diseased / non-diseased cases (n)	≈ 2,620 / ≈ 4,110	Back-calculated from CI width
PPV / NPV at 2% prevalence, %	20.7 / 99.9	Illustrative
PPV / NPV at 5% prevalence, %	40.3 / 99.7	Illustrative
PPV / NPV at 10% prevalence, %	58.7 / 99.4	Illustrative

Se, sensitivity; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value. Predictive values assume the pooled estimates apply to the target condition and prevalence stated, and are for illustration only.

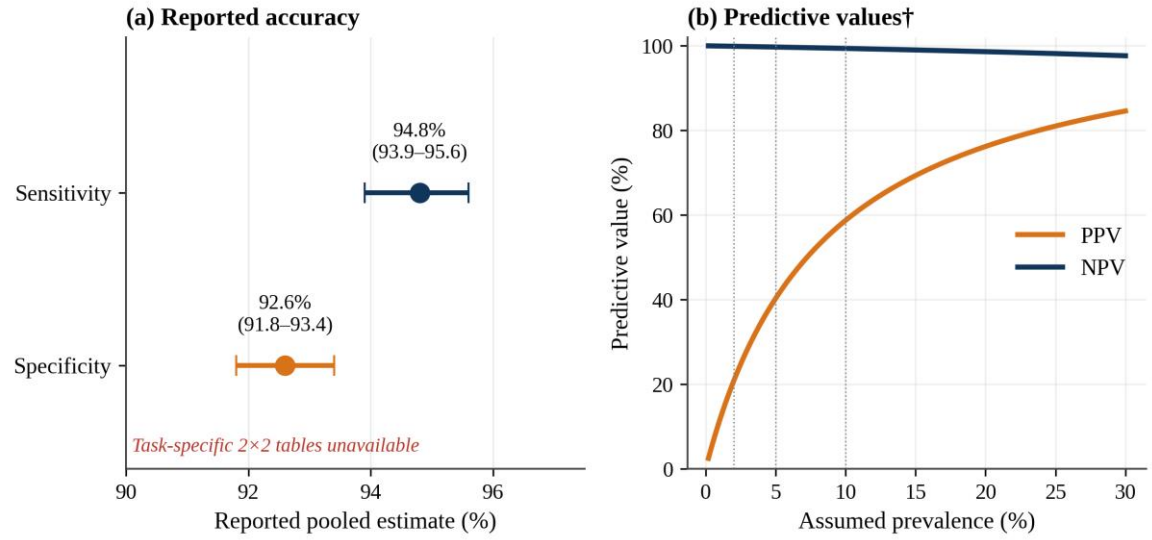


Figure 9. (a) Reported pooled sensitivity and specificity with 95% CIs. (b) †Illustrative positive and negative predictive values across assumed prevalence, computed from the reported pooled estimates; dotted lines mark 2%, 5% and 10% prevalence.

3.7 Maternal, neonatal and referral outcomes

Verified maternal deaths, live births, neonatal deaths, eligible pregnancy counts, referral denominators, completion windows and follow-up sources were not available. Maternal mortality, neonatal mortality and referral completion are therefore not estimated; they should be retrieved and validated rather than inferred from the existence of a risk-scoring or referral workflow.

4. Discussion

4.1 Principal findings

In this 12-month evaluation across three rural PHCs, an AI-enabled telemedicine pathway recorded 48,650 encounters among one quarter of the catchment population, used AI-assisted triage in 83.0% of encounters and resolved 85.4% of encounters at primary level while maintaining a district specialist link. Consultation turnaround, STEMI door-to-needle time and patient expenditure were all lower than pre-implementation values, and crude STEMI mortality was lower than in a historical cohort (RR 0.34; 95% CI 0.19–0.60). These findings are relevant to service design but, given the uncontrolled design and aggregate data, do not establish comparative effectiveness or safety, consistent with guidance that separates implementation performance from diagnostic validity and causal effect [8,18,20].

4.2 STEMI pathway in context

Rapid fibrinolysis within a pharmaco-invasive strategy is an established route to reperfusion where primary percutaneous intervention is not timely [23,27]. India's Tamil Nadu STEMI programme showed that hub-and-spoke networks combining ECG transmission, early thrombolysis and transfer can be implemented in public systems [28]. The 91.3% thrombolysis proportion and 104-minute reduction in door-to-needle time in our series are directionally consistent with this model. Nevertheless, a door-to-needle time of 38 minutes remains longer than guideline targets for fibrinolysis (ESC: ideally within 10 minutes of diagnosis, a metric with a different time zero) [23], leaving room for further pathway optimisation. The 9.3-percentage-point crude mortality difference is large relative to effects typically expected from a care-pathway change and probably reflects, in part, differences in case mix, referral patterns, documentation and follow-up between periods. The E-value of 5.31 suggests that moderate confounding alone is unlikely to explain the whole contrast [26], yet selection and measurement differences between cohorts could. A prespecified cohort, a fixed time zero and outcome window, complete follow-up, and an interrupted time-series or controlled design are needed to isolate any contribution of the digital pathway [29].

4.3 Diagnostic performance and AI governance

Chest radiograph screening, ECG triage and maternal risk scoring have different target conditions, reference standards and costs of error. Independent evaluations have shown that AI chest-radiography algorithms can reach performance suitable for triage in high-burden settings, and that deep-learning ECG interpretation can approach cardiologist-level performance, but both depend on local calibration and threshold choice [5,6]. Pooling these tasks into a single sensitivity and specificity obscures task-specific performance and precludes verification. A defensible accuracy analysis should provide separate 2×2 tables, predictive values at local prevalence, confidence intervals, thresholds, indeterminate results, reader expertise and blinding [19,21].

Equity and bias also warrant prospective monitoring: algorithms can encode systematic bias through proxy measures and unrepresentative training data [30]. Implementers should audit performance by age, sex, language and centre, and track automation bias, false reassurance, connectivity failures and cybersecurity risks. India's SAHI framework and BODH benchmarking platform now provide a national route for independent validation of such tools [16], aligned with WHO and ICMR emphasis on external validation, human oversight and post-deployment monitoring [11–13].

4.4 Patient expenditure

The 68.4% lower out-of-pocket expenditure plausibly reflects avoided travel, wages lost and outside diagnostic visits, in keeping with the access-related rationale for telehealth [1,2]. However, current calculations omit programme costs (staffing, connectivity, equipment, maintenance, training) and downstream care. A formal evaluation following CHEERS 2022 should state perspective, time horizon, cost categories, uncertainty and incremental outcomes [22]; the ₹6.15 crore figure should be read as an upper-bound arithmetic illustration, not a net saving.

4.5 Generalisability and health-system implications

The evaluation covered three PHCs in one district with a single specialist hub, so transferability depends on connectivity, specialist availability, medical-officer capacity and referral logistics. The model fits within India's existing eSanjeevani and teleconsultation architecture [9,10], and its governance needs map onto the SAHI framework, whose emphasis on validation, data stewardship and post-deployment monitoring is directly relevant to any scale-up [16]. Because the AI-assisted subset (83.0%) and the PHC-resolution proportion (85.4%) were recorded at encounter level, they should not be generalised to person-level coverage without accounting for repeat visits.

4.6 Ethical, legal and data-protection considerations

AI-supported care in public facilities raises issues of informed consent, explainability, accountability for errors and equitable access. Consent in the patient's preferred language, preservation of clinician override, documentation of AI outputs in the clinical record and independent audit of model performance are minimum safeguards [11,12]. Processing of personal health data must follow the purpose-limitation, consent and security obligations of the Digital Personal Data Protection Act, 2023 [15]. Vendor-supplied algorithms should be accompanied by documentation of training data, versions and update policies, which were not available for this evaluation.

4.7 Strengths and limitations

Strengths include a full 12-month implementation period, three PHCs with a district specialist hub, explicit separation of individuals from encounters, reconciliation of all aggregate figures, and transparent reporting of what cannot be concluded. Limitations are: (i) observational design without a concurrent control; (ii) aggregate rather than individual-level data, precluding adjustment for repeated encounters, PHC clustering and case mix; (iii) historical comparators of uncertain comparability for STEMI and other indicators; (iv) pooled diagnostic estimates without task-specific contingency

tables, thresholds or software versions; (v) a possible measurement artefact in turnaround-time dispersion; (vi) no programme costing; (vii) absent maternal, neonatal and referral-completion endpoints; and (viii) possible selection into the AI-assisted subset (83.0%), whose clinical characteristics were not compared with the remaining 17.0%.

4.8 Implications for practice and policy

Five principles follow: define the unit of analysis before calculating percentages; distinguish descriptive association from causal effect; avoid pooling heterogeneous diagnostic tasks; define outcome denominators and follow-up windows before reporting maternal, neonatal or referral outcomes; and separate observed findings from proposed quality-improvement measures. Digital health performance should be judged by clinical safety, equity, patient experience, continuity of care and cost, not by volume or speed alone [3,11].

5. Recommendations for the next evaluation

5.1 Evidence readiness

Figure 10 summarises the maturity of each evidence domain in the present dataset, from internally consistent service counts to unavailable endpoints.

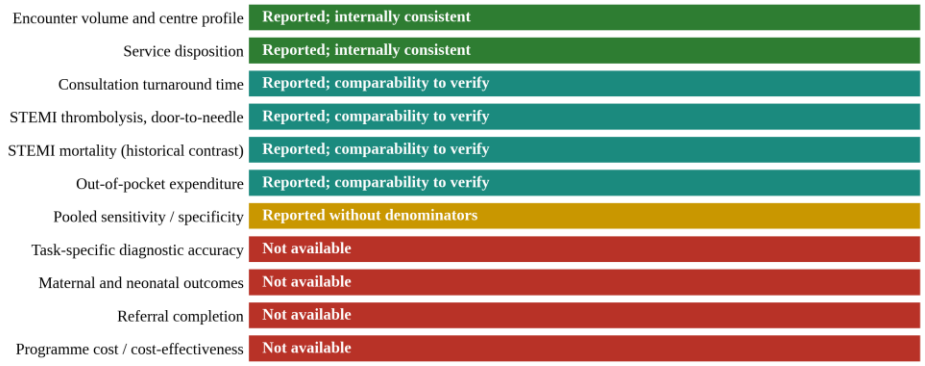


Figure 10. Evidence-readiness map of the present dataset by outcome domain.

5.2 Priority actions

1. Predefine primary and secondary outcomes, eligible populations, denominators and time windows in a registered protocol.
2. Create a participant-flow diagram distinguishing screened, eligible, enrolled, analysed and followed-up individuals, with reasons for exclusion and missingness.
3. Report each AI task separately, with software version, threshold, reference standard, blinding and complete 2×2 contingency tables.
4. Use a controlled or interrupted time-series design for effectiveness questions, with prespecified adjustment for case mix, PHC clustering and repeat encounters.
5. Capture referral issued, accepted, arrival at the receiving facility, specialist assessment, treatment and feedback within a defined interval.
6. Define maternal and neonatal outcomes using accepted programme definitions and reconcile deaths with approved records.
7. Conduct a prospective economic evaluation including programme and patient costs, uncertainty and sensitivity analyses.
8. Include patient experience, accessibility, equity, language preference, connectivity failures, overrides and adverse events.
9. Maintain a version-controlled data dictionary, audit trail and secure de-identification procedure.

5.3 Reporting-gap map

Table 8. Gap analysis against reporting guidance

Domain	Current status	Required for publication-grade evidence	Guidance
Observational reporting	Aggregate counts reconciled	Individual-level dataset, missingness, clustering	STROBE
AI diagnostic accuracy	Pooled Se/Sp only	Task-specific 2×2 tables, thresholds, reference standard, blinding	STARD-AI, CLAIM
Early clinical evaluation	Pathway described	Software versions, overrides, adverse events, human-factor data	DECIDE-AI
Effectiveness (STEMI)	Crude historical contrast	Time-series or controlled design with adjustment	Bernal et al.

Domain	Current status	Required for publication-grade evidence	Guidance
Economics	Patient expenditure only	Programme costs, perspective, uncertainty	CHEERS 2022
Maternal/neonatal/referral	Not quantifiable	Validated numerators, denominators, windows	National programme definitions

6. Conclusion

Across three rural PHCs, an AI-enabled telemedicine pathway recorded 48,650 encounters among 34,120 individuals, with AI-assisted triage in 83.0% and PHC-level resolution in 85.4% of encounters. Consultation turnaround, door-to-needle time and out-of-pocket expenditure were lower than before implementation. Among 312 STEMI cases, 91.3% were thrombolysed and crude mortality was 4.8% versus 14.1% historically (RR 0.34, 95% CI 0.19–0.60); this unadjusted contrast does not establish causation. Task-specific diagnostic validation, complete follow-up and controlled, adjusted evaluation are required before firm conclusions on clinical effectiveness, maternal and neonatal outcomes, referral completion or scalability.

Declarations

Ethics approval and consent to participate: Institutional Ethics Committee approval (IEC/PIMS/2023/114) was obtained before commencement; informed consent was obtained in Telugu, Hindi or English according to participant preference.

Consent for publication: Not applicable; no identifiable material is included.

Funding: Self-funded.

Competing interests: The authors declare no competing interests.

Data availability: De-identified data may be available from the corresponding author on reasonable request, subject to institutional and ethics approvals.

Author contributions: Atul Gopal Agrawal: conceptualisation, study design, methodology, supervision, investigation, interpretation and drafting. Ankita Agrawal: research support, interpretation, literature review and critical revision. Both authors approved the final manuscript.

Acknowledgements: The authors acknowledge the medical officers and healthcare personnel at the participating PHCs.

Abbreviations: AI, artificial intelligence; CI, confidence interval; ECG, electrocardiogram; NNT, number needed to treat; OR, odds ratio; PHC, primary health centre; PPV/NPV, positive/negative predictive value; RD, risk difference; RR, risk ratio; STEMI, ST-elevation myocardial infarction.

Appendix: calculation audit

Table A1. Independent re-computation and reconciliation of aggregate figures

Check	Calculation	Result	Status
Centre encounters sum	17,420 + 15,880 + 15,350	48,650	Reconciled
Catchment sum	48,500 + 42,200 + 45,100	135,800	Reconciled
Disposition sum	41,526 + 7,124	48,650	Reconciled
Encounters per individual	48,650 / 34,120	1.4258 → 1.43	Confirmed
Penetration of catchment	34,120 / 135,800	25.1%	Derived
Gangadhara share	15,880 / 48,650	32.64% → 32.6%	Corrected (32.7% in earlier draft)
Choppadandi / Manakondur share	17,420 / 48,650; 15,350 / 48,650	35.8%; 31.6%	Confirmed
Weighted mean age	Σ(encounters × centre mean age) / 48,650	41.47 years	Consistent with 41.5
Weighted female proportion	Σ(encounters × centre %) / 48,650	53.78%	Consistent with 53.8
AI-assisted triage	40,384 / 48,650	83.01%	Confirmed
Thrombolysis	285 / 312	91.35%	Confirmed
STEMI mortality (implementation, historical)	15 / 312; 43 / 305	4.81%; 14.10%	Confirmed
Risk difference	4.81% − 14.10%	−9.29 pp	Confirmed (≈ −9.3)
Turnaround difference	48.2 − 11.4	36.8 min (76.4%)	Confirmed
Door-to-needle difference	142 − 38	104 min (73.2%)	Confirmed

Check	Calculation	Result	Status
Expenditure difference	1,850 − 585	₹1,265 (68.4%)	Confirmed
Aggregate extrapolation	1,265 × 48,650	₹61,542,250 (≈ ₹6.15 crore)	Confirmed
Implied n from Se CI	$z^2p(1-p)/h^2$, h = 0.0085	≈ 2,621	Derived
Implied n from Sp CI	$z^2p(1-p)/h^2$, h = 0.0083	≈ 4,113	Derived

REFERENCES

1. Dorsey ER, Topol EJ. State of telehealth. N Engl J Med. 2016;375(2):154–161. doi:10.1056/NEJMra1601705
2. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open. 2017;7(8):e016242. doi:10.1136/bmjopen-2017-016242
3. World Health Organization. WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening. Geneva: WHO; 2019.
4. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56. doi:10.1038/s41591-018-0300-7
5. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med. 2019;25(1):65–69. doi:10.1038/s41591-018-0268-3
6. Qin ZZ, Ahmed S, Sarker MS, et al. Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of five artificial intelligence algorithms. Lancet Digit Health. 2021;3(9):e543–e554. doi:10.1016/S2589-7500(21)00116-3
7. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17:195. doi:10.1186/s12916-019-1426-2
8. Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ. 2020;368:m689. doi:10.1136/bmj.m689
9. Ministry of Health and Family Welfare, Government of India. Telemedicine Practice Guidelines: Enabling Registered Medical Practitioners to Provide Healthcare Using Telemedicine. New Delhi: MoHFW; 25 March 2020.
10. Centre for Development of Advanced Computing (C-DAC). eSanjeevani: National Telemedicine Service. Government of India. https://www.cdac.gov.in/index.aspx?id=product_details&productId=eSanjeevaniNationalTelemedicineService
11. World Health Organization. Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Geneva: WHO; 2021. https://www.who.int/publications/i/item/9789240029200
12. Indian Council of Medical Research. Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare. New Delhi: ICMR; 2023.
13. World Health Organization. Regulatory Considerations on Artificial Intelligence for Health. Geneva: WHO; 2023.
14. World Health Organization. Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multi-Modal Models. Geneva: WHO; 2024. https://www.who.int/publications/i/item/9789240084759
15. Government of India. The Digital Personal Data Protection Act, 2023 (Act No. 22 of 2023). Gazette of India; 11 August 2023.
16. Ministry of Health and Family Welfare, Government of India. Launch of the Strategy for Artificial Intelligence in Healthcare for India (SAHI) and the Benchmarking Open Data Platform for Health AI (BODH). New Delhi: Press Information Bureau; 17 February 2026.
17. World Health Organization. Global Strategy on Digital Health 2020–2025. Geneva: WHO; 2021; extension to 2027 endorsed by the World Health Assembly, May 2025. https://www.who.int/news/item/23-05-2025-world-health-assembly-endorses-extension-of-the-global-digital-health-strategy-to-2027
18. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453–1457. doi:10.1016/S0140-6736(07)61602-X
19. Sounderajah V, Guni A, Liu X, et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. Nat Med. 2025;31:3283–3289. doi:10.1038/s41591-025-03953-8
20. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ. 2022;377:e070904. doi:10.1136/bmj-2022-070904
21. Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell. 2020;2(2):e200029. doi:10.1148/ryai.2020200029
22. Husereau D, Drummond M, Augustovski F, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement. BMJ. 2022;376:e067975. doi:10.1136/bmj-2021-067975
23. Ibanez B, James S, Agewall S, et al. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation. Eur Heart J. 2018;39(2):119–177. doi:10.1093/eurheartj/ehx393
24. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med. 1998;17(8):857–872.
25. Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat Med. 1998;17(8):873–890.
26. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med. 2017;167(4):268–274. doi:10.7326/M16-2607
27. Armstrong PW, Gershlick AH, Goldstein P, et al. Fibrinolysis or primary PCI in ST-segment elevation myocardial infarction. N Engl J Med. 2013;368(15):1379–1387. doi:10.1056/NEJMoa1301092
28. Alexander T, Mullasari AS, Joseph G, et al. A system of care for patients with ST-segment elevation myocardial infarction in India: the Tamil Nadu–ST-Segment Elevation Myocardial Infarction Program. JAMA Cardiol. 2017;2(5):498–505. doi:10.1001/jamacardio.2016.5977
29. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol. 2017;46(1):348–355. doi:10.1093/ije/dyw098

30. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447–453. doi:10.1126/science.aax2342