NBER WORKING PAPER SERIES

AI-ENHANCED HANDHELD ECGS TO SCREEN FOR PRIOR MYOCARDIAL INFARCTION IN RURAL INDIA

Jenny Wang
Alexander Schubert
Nikhil Kanakamedala
Madeline McKelway
Luke Messac
Cyrus Reginald
Frank Schilbach
T.S. Selvavinayagam
Girija Vaidyanathan
Esther Duflo
Ziad Obermeyer

Working Paper 34690
http://www.nber.org/papers/w34690

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2026

We acknowledge funding from Chan Zuckerberg Biohub San Francisco, Gordon and Betty Moore Foundation, Griffin Catalyst, J-PAL, and Schmidt Futures. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w34690

AI-Enhanced Handheld ECGs to Screen for Prior Myocardial Infarction in Rural India
Jenny Wang, Alexander Schubert, Nikhil Kanakamedala, Madeline McKelway, Luke Messac, Cyrus Reginald, Frank Schilbach, T.S. Selvavinayagam, Girija Vaidyanathan, Esther Duflo, and Ziad Obermeyer
NBER Working Paper No. 34690
January 2026
JEL No. I1, I14, O12

ABSTRACT

Efforts to move care out of hospitals and into the community face a hard constraint: while most medications can be delivered anywhere, diagnosis is still mostly trapped inside the formal health system—clinics, hospitals, and the people and machines inside them. This is especially binding in low- and middle-income countries due to distance, cost, and limited health system capacity. Low-cost data collection devices (e.g., handheld electrocardiograms (ECGs), retinal cameras) are promising, but do not solve the bottleneck: their outputs still require expert interpretation. Artificial intelligence (AI) could turn these raw signals into actionable outputs at scale, but most medical AI is trained in high-income settings and deteriorates in different populations and settings. We assemble a new dataset in rural Tamil Nadu, India that links low-cost device signals to hospital-grade reference testing. We use it to build a proof-of-concept screening tool: an algorithm that predicts the result of a hospital-based reference-standard test (regional wall motion abnormality on echocardiography), using only data from a handheld ECG device (cost: $60). This would allow targeting of secondary prevention medications—high-dose statins, antiplatelet agents, and beta blockers—proven to be cost-effective in both high- and low-resource settings. We evaluate predictive performance and cost-effectiveness of such screening, benchmarking against common clinical risk scores (including the American Heart Association 10-year risk score). In a held-out test set, our model identifies subgroups with rates high enough to make screening very cost-effective: the top 2.5% have a 9.3% rate of evidence consistent with prior heart attack—about six times the base rate—delivering benefits at US$1,999 per life-year saved. The model also flags high-yield patients who would be missed by standard clinical scores, underscoring the limits of transporting Western-derived risk tools to low-and middle-income settings. To enable extensions to other conditions in the dataset, we make all data freely available for research.

Jenny Wang
Massachusetts Institute of Technology
wangjm@mit.edu

Alexander Schubert
University of California, Berkeley
Computational Precision Health
and University of California, San Francisco
alexander_schubert@berkeley.edu

Nikhil Kanakamedala
Massachusetts Institute of Technology
nikhil.mitindia@gmail.com

Madeline McKelway
Dartmouth College
madeline.d.mckelway@dartmouth.edu

Luke Messac
Harvard University
lmessac@bwh.harvard.edu

Cyrus Reginald
Abdul Latif Jameel Poverty
Action Lab (J-PAL)
creginald@povertyactionlab.org

Frank Schilbach
Massachusetts Institute of Technology
Department of Economics
and NBER
fschilb@mit.edu

T.S. Selvavinayagam
Government of Tamil Nadu
drsvinayagam@gmail.com

Girija Vaidyanathan
Government of Tamil Nadu
gigiv_40@yahoo.com

Esther Duflo
Massachusetts Institute of Technology
Department of Economics
and NBER
eduflo@mit.edu

Ziad Obermeyer
University of California, Berkeley
School of Public Health
and NBER
zobermeyer@berkeley.edu

AI-enhanced handheld ECGs to screen for prior myocardial infarction in rural India

Jenny Wang[†] (MIT), Alex Schubert[†] (UC Berkeley), Nikhil Kanakamedala[*] (CEGIS), Madeline McKelway* (Dartmouth), Luke Messac[*] (Harvard), Cyrus Reginald[*] (J-PAL South Asia), Frank Schilbach[*] (MIT, NBER), T. S. Selvavinayagam[*] (Gov't of Tamil Nadu), Girija Vaidyanathan[*] (Gov't of Tamil Nadu), Esther Duflo[*‡] (MIT), Ziad Obermeyer[*‡] (UC Berkeley)

[†,‡]: equal contribution
[*]: alphabetical order

Background. Low-cost data collection devices, like single-lead electrocardiograms (ECGs), can provide advanced diagnostic information in community settings, but require physicians to interpret the complex data they produce. Artificial intelligence (AI) could overcome this bottleneck, but data needed to train AI are lacking, especially in low- and middle-income countries where diagnostic needs are greatest. We introduce a novel dataset and use it to train a proof-of-concept AI tool that screens patients for prior myocardial infarction (MI) based on a single-lead ECG, in order to target secondary prevention.

Methods. We enrolled participants aged ≥40 years in primary health centres across rural Tamil Nadu, India. We collected data from a range of low-cost devices and also performed reference-standard diagnostic testing. We linked handheld single-lead ECG waveforms to echocardiography results, and trained a deep neural network to predict echocardiographic regional wall motion abnormality, consistent with prior MI, as interpreted by a cardiologist. In an independent hold-out, we assessed positive predictive value (PPV), sensitivity, and cost-effectiveness of community-based screening for prior MI, and compared our algorithm to the American Heart Association (AHA) 10-year risk score.

Results. Among 5,301 participants, the base rate of prior MI was 1.5%. The ECG-based model achieved AUC 0.77 (95% CI 0.69–0.84), significantly higher than the AHA risk score (AUC 0.68; p=0.02) for prior MI. Among the riskiest 2.5% of participants, 9.3% had evidence of prior MI (analogous to PPV)—6x higher than the base rate. Community-based screening for prior MI using ECGs would cost US$1,999 per life-year saved, far less than using AHA risk scores ($15,414). ECGs identified a sizable number of patients who, despite low AHA risk scores, had high rates of prior MI, highlighting opportunities to improve on risk scores developed in Western populations.

Conclusions. AI applied to single-lead handheld ECGs can identify community-dwelling adults with prior MI at rates that allow cost-effective screening, and outperform traditional risk scores. Device–AI combinations enable a new approach to public health, which could be extended to other conditions captured in our open-access dataset and relevant to LMICs.

Funding: Chan Zuckerberg Biohub San Francisco, Gordon and Betty Moore Foundation, Griffin Catalyst, J-PAL, Schmidt Futures.

There is growing interest in shifting health care out of the hospital. This can improve access, reduce delays, and lower costs, delivering a triple win for patients, providers, and health systems.[1] While drugs can be prescribed anywhere, an obvious drawback of community-based care is the lack of those diagnostic technologies available only in-hospital. Lack of access to diagnosis is a particular problem in low- and middle-income countries (LMICs),[2–4] due to logistical and financial barriers and limited health care system resources.[5]

Low-cost diagnostic devices—e.g., handheld electrocardiograms (ECGs) and retinal cameras, point-of-care blood tests, pulse oximeters—are starting to change this calculus, allowing health workers with minimal training or even patients to acquire advanced physiological data. This can be valuable in some settings,[6,7] but human expertise remains a bottleneck: the complex data produced by devices must still be interpreted and acted upon by physicians, limiting impact.

Layering artificial intelligence (AI) on top of low-cost devices could solve this problem, and there is growing evidence that this approach can achieve better diagnosis in high-income settings.[8–10] But a key barrier to broadening this vision globally is the lack of data from LMICs. Most AI is trained on data from tertiary health systems in high-income countries,[11] and on the rare occasions where AI is applied elsewhere, performance can drop sharply.[12]

We introduce a new dataset pairing data from low-cost devices with reference-standard clinical diagnostics. The data were collected from participants in rural India, and chosen based on relevance to disease burden in LMICs. They lay the groundwork for a new kind of AI, that can predict hospital-grade diagnostic test results using only data from low-cost devices.

We demonstrate the promise of this approach by training an algorithm to flag patients with prior myocardial infarction (MI), based on a handheld ECG. Underdiagnosis of MI is a problem in all countries: even in high-income countries: e.g., the United States, 40% to 80% of MIs are 'silent.'[13,14] Patients with undiagnosed MIs cannot benefit from secondary prevention—high-dose statins, antiplatelet agents, and beta blockers—a proven, cost-effective intervention in both high- and LMICs.[15,16] A key reason many MIs remain undiagnosed is because reference-standard tests (e.g., echocardiography, MRI) are impractical for screening, particularly in low-resource settings. ECGs are far cheaper, but their sensitivity and specificity for prior MI, at least based on human interpretation, is low.[17,18] We train a deep learning model and measure its ability to identify those with prior MI, as measured by echocardiography, inspired by recent work in high-income settings.[19] We also evaluate the tool's potential cost-effectiveness when used to screen elders for prior MI in their homes, via community health workers. Finally, we compare it to more traditional risk predictors, e.g., the American Heart Association (AHA) 10-year risk score.[20]

This algorithm is a proof-of-concept: one of many possible applications of this dataset, which we make publicly and freely available for non-profit research via an open science platform.[21] It serves as a concrete example of how AI trained on new data can create value: by increasing access to high-quality information in the community, at a fraction of the cost of the hospital.

Methods

We collected data across primary health centers (PHCs) in five districts in Tamil Nadu, working closely with the Government of Tamil Nadu, which authorized the work via G.O. (Ms.) No. 58 PD&SI (TC-III), dated 03.04.2023. The study received ethical clearance from the MIT institutional review board (COUHES protocol 2210000779), as well as the Institute for Financial Management and Research. Approval was also obtained from the Directorate of Public Health Scientific Advisory Committee (Letter No. DPHPM/SAC/2022/082, dated 30.12.2022).

Data collection

District and PHC Selection. Data were collected via 'health camps' in five districts: Dharmapuri, Greater Chennai, Kanyakumari, Tiruvannamalai, and Trichy. These districts were selected based on our prior work,[2] to ensure diversity in Human Development Index and old-age dependency ratios. Within each district, four primary health centers were selected, based on: adequate space for equipment and patient flow, availability of diagnostic rooms (for ECG, X-ray, echocardiogram), availability of key staff (medical officers, lab technicians, data entry operators), and sufficient outpatient footfall.

Participant Selection and Consent. Participants were recruited using voluntary response. Individuals aged ≥ 40 who visited the PHC during camp hours were offered enrollment. Each consenting participant received an information sheet and signed a consent form. A unique study ID was printed on both the consent form (which remained with the participant) and the case sheet (retained by the PHC after verification).

Data Collection. Health camps were conducted between July and September 2023. In each PHC, data collection spanned eight days, one day for staff training and seven for participant testing, across two weeks. The daily target enrollment for each PHC was 45–50 participants.

Study personnel trained PHC staff to implement the testing protocol prior to the camps, covering consent protocols, questionnaire administration, and mobile device use. The Department of Public Health trained district microbiologists on blood sample handling and analysis. Study field staff and Medical Officers supervised daily operations. Spot checks by study personnel ensured participant understanding, consent adherence, and completion of tests. All devices were calibrated prior to use following usual practice. PHC staff logged all data on paper case sheets (Appendix A), except for questionnaire responses which were entered digitally via SurveyCTO.

Each participant underwent both low-cost and reference-standard tests. The key low-cost test for this study was a handheld mobile ECG (AliveCor KardiaMobile), a small (8.2 x 3.2 x 0.35 cm) FDA-cleared device that records a single-lead ECG (equivalent to lead I) and relays the tracing to a tablet or phone. It is typically used by patients to record ECGs during symptoms at home and send to their physician for review. In addition to the ECG, we collected physical measurements, pulse oximetry, grip and endurance tests, retinal photography, tonometry, cognition and smoking questionnaires, and blood glucose testing. Reference-standard tests included 12-lead ECG, echocardiogram, chest X-ray, and blood tests (Hb, HbA1C, lipid panel, creatinine).

Most tests were conducted by PHC staff under the supervision of the medical officer. Three reference-standard tests (12-lead ECG, echocardiography, X-ray) were performed by a private vendor, who also provided registered specialist interpretation (cardiology and radiology) via structured reports. Blood samples were analyzed at multiple sites, based on availability: PHCs (Hb), district medical colleges (HbA1c), and the State Public Health Lab (lipids, creatinine). Technicians performing established reference-standard tests were instructed to flag any findings that appeared urgent, for immediate review by a primary health care clinician on site, which happened in at least one case (an acute pathology on echocardiography).

Data from case sheets were digitized daily, and a 10% random sample was re-digitized for quality control, yielding an error rate <0.5%. Survey forms included built-in logic checks. Blood data were merged post-hoc to complete the dataset. Prior to analysis, data were de-identified in accordance with usual practice, removing names, report dates and times, and other identifying variables (based on US Safe Harbor criteria).

Report Generation and Distribution. Health reports containing results and physician interpretations were compiled by study staff and couriered to PHCs within three weeks (see Appendix A for a sample report). Participants were reminded by phone to collect reports, using their consent form as ID. Medical Officers provided consultations upon report collection. A receipt was signed by participants and retained at the PHC.

Statistical analysis

Sample re-weighting to match the general population. Our dataset, drawn from a convenience sample of adults aged ≥40 attending PHCs, has compositional differences to the general population which we eventually intend to screen using the tool. We thus re-weight the sample to match the age and sex distribution of the Tamil Nadu population aged 40 and older based on 2011 Census data to provide a more representative picture of how the AI tool would perform in screening the intended target population. Appendix B shows that summary statistics and model performance are similar between the original sample and the re-weighted population.

Model development. We use a two-stage process to train the model: pre-train in a large dataset from a high-income setting, then fine-tune in data we collect from Tamil Nadu.

The pretrained model uses a dataset of 12-lead ECGs drawn from an electronic health record dataset in Sweden, described previously.[22] We extract Lead I from the 12-lead ECGs, and apply standard transformations[23] to approximate waveforms acquired from a handheld device. We convert 10-second tracings to a median beat, by segmenting using standard methods.[24] We train a deep neural network (18-layer ResNet) to predict whether the patient had ever been assigned an International Classification of Disease diagnosis of MI prior to the date of the ECG.

This pre-trained model is the starting point for a new model trained on single-lead ECG tracings from Tamil Nadu, similarly processed into median beats: an 18-layer ResNet predicting any regional wall motion abnormality (RWMA) on echocardiography. Our model outputs a likelihood of RWMA based on the waveform alone. We then use logistic regression, predicting RWMA using

this raw waveform model output along with age and sex (in the training set), to form our final predictions. For full architecture details, see Appendix B.

We train the model in a 40% random sample (training set) from our main dataset, and measure performance in the remaining 60% (holdout set). This split was based on a power calculation yielding an 80% likelihood of identifying a significant difference in rate of prior MI between a high-risk group—the top 20% of model-predicted risk—and a low-risk group—the bottom 80%— under conservative model performance assumptions based on the AHA 10-year risk score.

Model evaluation. All results in this manuscript come from the 60% Tamil Nadu hold-out set, which the model has never seen. We assess discrimination using the area under the receiver operating characteristic curve (AUC) for prior MI, measured by any RWMA on echocardiography. To measure performance in discrete high-risk groups, we apply thresholds that flag the riskiest 2.5%, 5%, 7.5%, … of the sample as high-risk, and calculate positive predictive value (PPV) and sensitivity, with 95% confidence intervals (CIs) generated by bootstrapping.

We then simulate a hypothetical screening intervention, shown in Figure 1, where community health workers collect handheld ECGs during visits to elders' homes, using the algorithm in real time to generate predictions on the likelihood of prior MI. High-risk individuals are referred to local health centers for echocardiography. If this confirms prior MI via identification of RWMA, a package of medications for secondary prevention is initiated. This workflow outlined in Figure 1.

We estimate the cost of each workflow component, as well as the health benefits of secondary prevention, to estimate the cost-effectiveness of ECG-based screening for prior MI. We begin by estimating the cost of referring a patient for confirmatory testing, C_{refer}, both direct (physician and testing costs) as well as indirect (patient time and travel) costs. For simplicity we assume all referred patients comply. In addition to referral costs, patients with a positive confirmatory test $Y = 1$ also incur the cost of treatment with medications for secondary prevention, C_{med}. In positive patients, we estimate life-years gained from secondary prevention, L, using age-specific benefit-risk tradeoffs, accounting for bleeding risks and medication effects on preventing future atherosclerotic cardiovascular disease. We consider a range of life-year valuations, V_L, from both WHO guidelines and GDP-adjusted estimates of the value of a statistical life.[25,26]

To determine the referral threshold at which screening is cost-effective, we set the benefits of screening (left) equal to costs (right):

$$Pr(Y = 1)L \cdot V_L = C_{refer} + Pr(Y = 1)C_{med}$$

This equation illustrates that the fraction of screened patients with prior MI, $Pr(Y = 1)$, is a key determinant of the cost-effectiveness of screening: costs C_{refer} are incurred for all patients referred for testing, but (net) benefits $L \cdot V_L - C_{med}$ only accrue to those who screen positive, $Y = 1$, and receive effective secondary prevention. Rearranging terms, the threshold at which

screening is cost-effective can be expressed as the marginal prevalence of prior MI at which costs equal net benefits:

$$Pr(Y = 1) = \frac{C_{refer}}{L \cdot V_L - C_{med}}$$

Because each input to our cost-effectiveness calculations has uncertainty, we sample across plausible ranges of each when calculating $Pr_m(Y = 1)$, which generates a distribution of marginal prevalence rates for cost-effective screening. Our main analysis uses a conservative benchmark at the 95th percentile of $Pr_m(Y = 1)$ to define cost-effectiveness, and we use the cost-effectiveness parameter values that correspond to this benchmark. Full assumptions, calculations, and analyses of alternative screening policies are detailed in Appendix D.

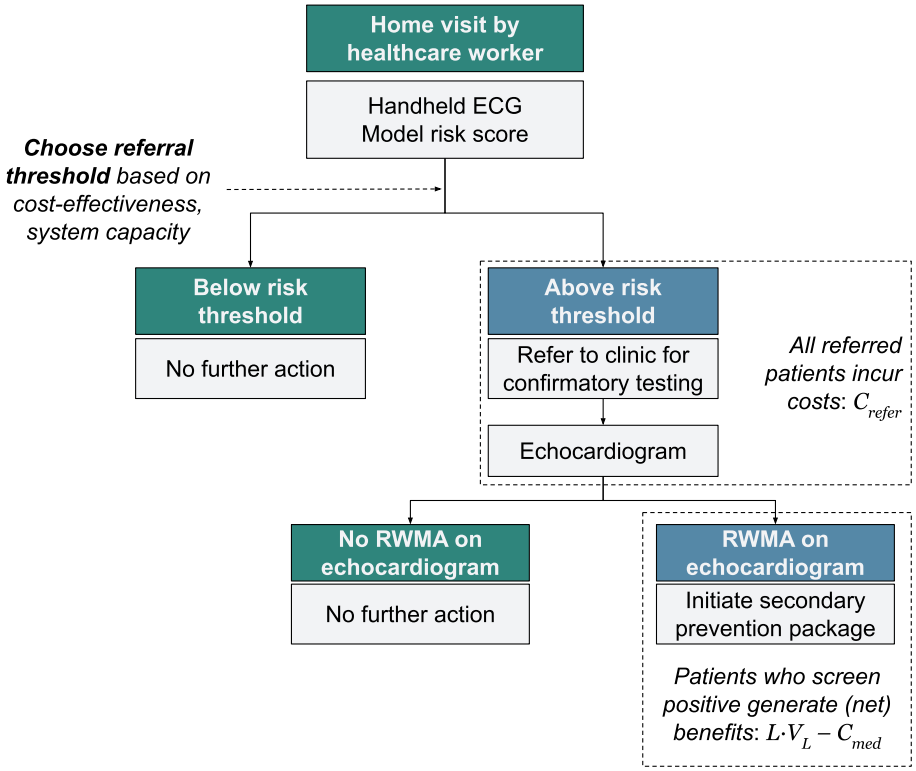

Figure 1. Hypothetical Workflow—Screening for Prior MI*: Handheld ECGs are taken by community health workers. If a patient is flagged as high-risk, she is referred to a health center for confirmatory testing. If prior MI is confirmed via echocardiogram, secondary prevention medications are initiated.*

Comparison to other risk predictors. We implement two alternative risk predictors that could be used to screen for ischemic heart disease risk in our setting. It is important to note that there are no established tools that are specific to *prior* MI risk, so we assume that predictors of *future* MI risk are correlated enough with past risk to provide a good baseline for screening.

First, we implement the AHA ASCVD 10-year Risk Calculator,[20] which predicts probability of future MI or stroke in the next 10 years, using sex, race, age, smoking, diabetes, blood pressure (systolic, and medications), and cholesterol (total and HDL). We are able to implement it because

our data collection included laboratory studies on all participants. Laboratory studies are nontrivial to implement in LMICs, but we choose the AHA model over the simpler but less accurate WHO Cardiovascular Risk Charts, because it allows us to upper-bound the predictive power of traditional risk predictors, and compare them to ECG-based approaches. Second, at the other extreme, we simulate a simple policy of age-based screening, in which all adults above a certain age are referred for further testing. This is easy to implement, but less accurate. Together, these two predictors lie at opposite extremes of more traditional statistical approaches, and thus serve as useful comparators to ECG-based risk predictions. To simplify comparison, our calculations do not account for the cost of the screening visit, or the cost of generating any of the risk scores (e.g., the ECG and tablet for the ECG, collection of vital signs and laboratory studies for the AHA risk score): screenings are not typically standalone programs, but rather added onto existing home visit and public health programs, and accounting for the marginal cost of adding a screening is beyond the scope of our simplified analysis.

To compare the cost-effectiveness of screening with different risk predictors, we sample from the distribution of $Pr(Y = 1|H_r = 1)$: the prevalence of MI in a given model r's high-risk group H_r, and calculate the cost-effectiveness of screening, using that model to define r. This yields a simulation-based p-value for cost-effectiveness comparisons: the fraction of draws in which the cost-effectiveness of one risk predictor is greater (less) than the other.

Results

A total of 5,930 individuals were enrolled, of which 343 were missing echocardiography, 281 were missing handheld ECG data, and 5 were missing age. Table 1 presents summary statistics for the remaining 5,301 individuals, split between training (40%) and holdout (60%) groups. The prevalence of any regional wall motion abnormality (RWMA) on echocardiography, as reported by the interpreting cardiologist, was 1.4% in the hold-out set.[1] The table also shows the five most common findings on echocardiography, and interpretations of the 12-lead ECG suggestive of prior MI. All metrics were broadly similar between the training and holdout samples. Statistics are reweighted by age and sex; statistics for the unweighted raw data are in Appendix B.

	All	***Training***	***Holdout***
n	5301	2252	3049
Traditional Risk Factors			
Age (mean)	54.2 (0.15)	53.8 (0.22)	54.5 (0.20)
Male	39.6 (0.67)	38.9 (1.03)	40.2 (0.89)
Smoker	9.4	9.4	9.3

[1] This is comparable to prevalence of prior MI shown in similar populations: Gopinath et al.[27] report a rate of 1.4% in Delhi; Rao et al.[28] arrive at a rate of 5.2% in urban areas and 3.1–4.4% in rural areas; Gupta et al.[29] report 1.5-4.0% in urban and 0.5-2% in rural areas. All these estimates come from studies that define prior MI based on ECG criteria (Q-waves), which has both false positives and false negatives compared to our ultrasound-based measure (RWMA).

	(0.40)	(0.62)	(0.53)
Systolic BP>140	25.3 (0.60)	25.0 (0.91)	25.6 (0.79)
Diastolic BP>90	21.4 (0.56)	22.5 (0.88)	20.6 (0.73)
Hypertensive: SBP>140 or DBP>90	31.8 (0.64)	31.9 (0.98)	31.8 (0.84)
Diabetic: HbA1c>6.5	35.6 (0.66)	35.7 (1.01)	35.6 (0.87)
Total Cholesterol (mg/dL, mean)	205.8 (0.67)	207.2 (1.03)	204.8 (0.89)
HDL (mg/dL, mean)	43.6 (0.16)	44.0 (0.24)	43.2 (0.21)
LDL (mg/dL, mean)	118.9 (0.44)	118.7 (0.67)	119.0 (0.58)
Ultrasound Findings			
Prior MI (RWMA)	**1.5** **(0.17)**	**1.7** **(0.27)**	**1.4** **(0.21)**
Aortic Valve Sclerosis	19.8 (0.55)	17.8 (0.81)	21.3 (0.74)
Diastolic Dysfunction	19.3 (0.54)	19.0 (0.83)	19.6 (0.72)
Mitral regurgitation, Mild	8.0 (0.37)	7.1 (0.54)	8.6 (0.51)
Left Ventricular Hypertrophy, Concentric	7.5 (0.36)	6.7 (0.53)	8.1 (0.50)
Aortic Regurgitation, Mild	5.5 (0.31)	4.7 (0.45)	6.2 (0.44)
ECG Findings (12-lead)			
Prior Infarction†	10.4 (0.44)	10.9 (0.66)	9.9 (0.60)
Q Wave	2.4 (0.22)	2.7 (0.34)	2.2 (0.29)
Poor R Wave Progression	2.7 (0.24)	2.9 (0.35)	2.6 (0.32)

†Cardiologist interpretation could state “prior infarction” without details, or could specifically note Q waves or Poor R Wave Progression; we combined these all together to form the “prior infarction” category.

Table 1. Summary statistics*. All values percentages unless noted; standard errors in parentheses.*

At our preferred cost-effectiveness parameter values, we find that a 2.8% rate of prior MI is the minimum at which benefits of systematic testing outweigh costs in our context. Thus, the 1.4% rate of prior MI in the holdout set is too low to justify universal screening with echocardiography. In order to make screening cost-effective, a predictor must isolate a subset of the full population with a prior MI rate of at least 2.8%, analogous to the model’s positive predictive value (PPV).

ECG-based risk predictor performance

As a preliminary metric of predictive performance, we calculate the AUC of our single-lead ECG-based risk predictions: 0.77 (95% bootstrapped CI: 0.69-0.84) for prior MI. This performance is comparable to recent work using 12-lead ECGs to predict RMWA on ultrasound (AUC: 0.72-78).[19] Our ECG-based predictions had a higher AUC than the AHA 10-year risk score and simple age-based screening (both AUC=0.68; vs. ECG-based predictions with DeLong's test for correlated ROC curves, p=0.02 and p=0.03, respectively).

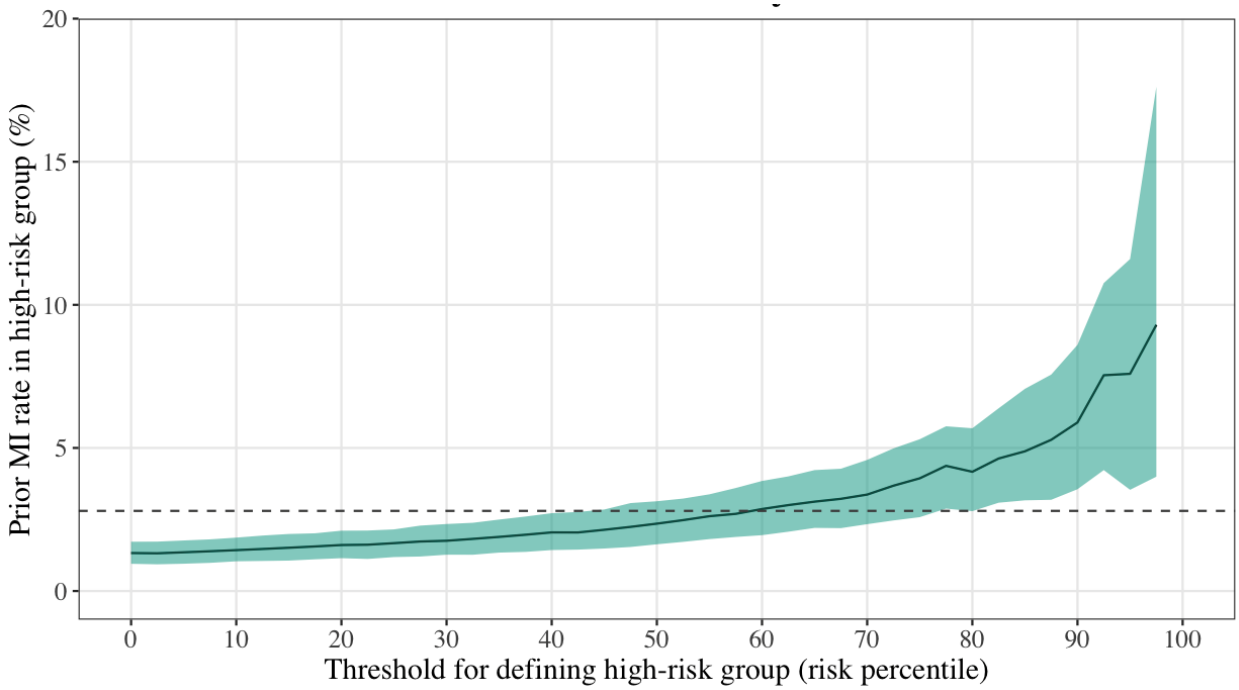

Figure 2. Positive predictive value of ECG predictions for prior MI*: Rate of prior MI (measured by RWMA on echocardiography) in high-risk group (y-axis, shaded 95% CI), vs. percentile threshold used to define high-risk group (x-axis). The far-right point shows the rate in the top 2.5% of the population. Horizontal dotted line indicates the cost-effective screening threshold of 2.8%.*

AUC measures discrimination throughout the risk distribution, but says little about a model's ability to identify an actionable high-risk group, which is critical for clinical and policy decisions. Figure 2 shows the rate of prior MI (i.e., PPV, y-axis) for different high-risk groups, defined by setting the threshold for "high risk" at different percentiles of model risk scores (x-axis). For example, the rightmost point sets the threshold at the 97.5th percentile of predicted risk, and shows the rate of prior MI in the highest-risk 2.5% of the sample: 9.3% (95% CI: 4-17.6%). This high-risk group accounts for 14.6% of all prior MI (i.e., sensitivity: 95% CI: 4.8-26.1%). Moving left, the threshold decreases in steps of 2.5%, and the next riskiest 2.5% of ECGs are added to the high-risk group. At each threshold, we compare the prevalence of prior MI between the high-risk groups vs. all others, and find that the high-risk rate of prior MI significantly exceeds the low-risk group from the 15th percentile onwards (this mirrors our power calculation, which used the 80th percentile). Starting at the 78th risk percentile, the prevalence of prior MI is significantly higher than 2.8%, indicating that screening at and above that risk percentile is cost-effective. The group with the highest prevalence of prior MI (based on the maximum lower bound of the 95% confidence interval) is at or above the 92.5th risk percentile, in whom prior MI prevalence is 7.5%

(95% CI: 4.2-10.8%); this group accounts for 35.4% of all prior MIs (95% CI: 22.9-50.0%). More details on model calibration and sensitivity are in the Appendix E.

Comparison of ECG performance to other screening tools

There are several widely-used ECG markers of prior MI, including Q-waves and R-wave abnormalities (e.g., poor R-wave progression). The consulting cardiologists who interpreted the ECGs identified any such abnormality consistent with prior MI in 10.9% of ECGs. In this group, the prevalence of prior MI was 1.9% (95% CI: 0.6%-3.7%), which is not significantly different from the base rate in the sample. Among the subset of these ECGs in whom Q waves were specifically noted (2.7% of the sample), there were *no* cases of prior MI.

Two other risk predictors may be useful for flagging those with prior MI: the AHA 10-year risk score and a simple age-based ranking. Panel A of Table 2 compares the PPV of these approaches. Both AHA and age-based risk predictors have low PPV at all risk thresholds, with the starkest gap at the highest end of predicted risk: for example, the PPV of the ECG-based risk scoring was 9.3% among the riskiest 2.5 percent, compared to PPVs of 0.8% and 1.4% for the AHA and age-based predicted risk, respectively. Inspecting the 95% CI on PPVs, the group with the highest lower-bound—i.e., where we can be most confident about a high rate of prior MI—is the riskiest 2.5% of the ECG-based risk score, with a lower bound of 4%.

Cost-effectiveness of screening with different risk predictors

PPV is clinically useful, but does not address cost-effectiveness of screening with a given risk predictor, based on costs of referral and testing, and benefits of secondary prevention. Panel B of Table 2 compares cost-effectiveness of screening with a risk predictor, as measured by the cost (in US dollars) per life-year saved. ECG-based risk performs the best across all risk thresholds: $1,999 dollars per life year in the top 2.5% ($p<0.001$ based on bootstrapped comparisons of the ECG- and age-based screening), $2,243 dollars per life year in the top 5% ($p=0.003$), and $2,624 dollars per life year in the top 10% ($p<0.001$). For comparison, WHO's cost-effectiveness thresholds based on India's GDP range from $3,500 to $10,500 USD per life-year (Appendix Table D3). More data on PPV, life-years, and cost-effectiveness are in Appendix D.

	Prior MI rate in high-risk groups, by risk predictor		
High-risk group size	*ECG-based*	*AHA 10y risk*	*Age*
Panel A: Prevalence of prior MI in high-risk group (PPV)			
Riskiest 2.5%	9.3% (4, 17.6)	0.8% (0, 4.7)	1.4% (0, 5.6)
Riskiest 5%	7.6% (3.5, 11.6)	3.8% (0.9, 6.4)	2.7% (0.4, 6)
Riskiest 10%	5.9% (3.6, 8.6)	3.7% (1.9, 5.8)	2.8% (1, 5)

Riskiest 15%	4.9% (3.2, 7.1)	3.1% (1.8, 4.6)	3% (1.6, 4.7)
Riskiest 30%	3.4% (2.3, 4.6)	2.5% (1.6, 3.5)	2.6% (1.7, 3.6)
Panel B: Cost per life-year saved (in USD) in high-risk group			
Riskiest 2.5%	$1,999 (1,423, 3,467)	$15,414 (4,684, 16,104)	$13,125 (4,348, 14,539)
Riskiest 5%	$2,243 (1,715, 3,698)	$4,168 (2,725, 12,516)	$6,649 (3,729, 28,841)
Riskiest 10%	$2,624 (2,070, 3,858)	$3,767 (2,765, 6,287)	$6,511 (4,308, 15,449)
Riskiest 15%	$2,937 (2,286, 4,080)	$4,209 (3,122, 6,598)	$5,223 (3,757, 9,133)
Riskiest 30%	$3674 (2,892, 5,017)	$4,730 (3,650, 6,800)	$5,135 (3,953, 7,267)

Table 2: Comparison of PPV and cost-effectiveness of screening with ECG-based risk, AHA 10-year risk score, and age-based screening*. The percent of population screened refers to the percent among those aged 40 and over. 95% bootstrap CIs displayed in parentheses.*

Population differences between those with high-risk ECGs and traditional risk models.

Our cost-effectiveness analysis highlights an important downside of screening policies that explicitly use age (whether alone or as part of a risk score like the AHA 10-year risk): older patients, by construction, have fewer remaining life-years to save, limiting the benefit of screening them. We see that for the age-based predictor, even though more stringent screening leads to identifying a higher fraction of positive cases, it can still be more expensive per life-year saved-–for example, moving from screening the riskiest 10% to 5% by age increases the PPV from 3.1 to 3.8%, but screening becomes *less* cost-effective: the cost per life-year increases from $3,767 to $4,168. The additional advantage of ECGs being able to identify younger patients can be seen most clearly in comparing the cost-effectiveness of ECG-based vs. age-based screening. Even at similar PPV levels, the ECG-based score has far higher cost-effectiveness than the AHA score or age: for example, the PPVs among the riskiest 30% of individuals are 3.4% for ECG-based and 2.6% for age-based screening, but the cost-effectiveness values are far lower for ECG-based screening: $3,674 per life year vs. $5,135 per life year). The average cost per life-year saved is $4,290 (with 95% confidence interval of $2,575 to $11,635) for patients flagged by cardiologist-flagged ECG manifestations of prior MI.

	AHA 10-year Score: Low Risk			***AHA 10-year Score: High Risk***		
Value	*ECG Low Risk*	*ECG High Risk*	*Difference: ECG High vs Low Risk*	*ECG Low Risk*	*ECG High Risk*	*Difference: ECG High vs Low Risk*
	(1)	(2)	(3)	(4)	(5)	(6)
N	2277	252	---	257	200	---
Proportion of sample	80.9	7.1	---	7.1	4.9	---

Prior MI (RWMA)	**0.7%**	**5.0%**	**4.3% (p=0.01)**	**1.6%**	**6.4%**	**4.8% (p=0.02)**
AHA 10year Risk	5.4%	11.1%	5.7 (p<0.01)	29.5%	34.5%	5.0 (p<0.01)
Age	51.0	61.2	10.3 (p<0.01)	63.9	70.1	6.1 (p<0.01)
Male	33.9%	62.1%	28.2% (p<0.01)	67.1%	72.7%	5.6% (p=0.24)
Smoker	7.4%	11%	3.5% (p=0.09)	24.1%	17.8%	-6.3% (p=0.10)
Systolic BP >140	20.2%	22.1%	2.0% (p=0.49)	63.7%	54.3%	-9.3% (p=0.06)
Diastolic BP >90	18.2%	15.6%	-2.6% (p=0.29)	44.2%	34.2%	-10.0% (p=0.03)
Hypertensive (BP>140/90)	26.7%	25.8%	-0.8% (p=0.78)	68.5%	61.3%	-7.2% (p=0.13)
Diabetic (HbA1c>6.5)	31.6%	27.9%	-3.5% (p=0.27)	74.4%	60.6%	-13.9% (p<0.01)
Total Cholesterol (mean)	205.2	194.4	-10.8 (p<0.01)	220.2	204.1	-16.1 (p<0.01)
HDL (mean)	43.5	42.4	-1.1 (p=0.13)	40.6	41.6	1.0 (p=0.36)
LDL (mean)	119.4	111.8	-7.7 (p<0.01)	123.5	120.2	-3.3 (p=0.32)

Table 3: Cardiovascular risk factors of low- vs. high-risk patients, as categorized by AHA 10-year risk vs. ECG-based predictions. *Columns 1-2 show those with low and intermediate AHA risk scores (88% of the sample), by low (Column 1) vs. high (Column 2) ECG-based risk. Column 3 compares Columns 1 vs. 2. Columns 4-5 show those with high AHA risk scores (i.e., >20% 10-year risk: 12% of the sample), by low (Column 4) vs. high (Column 5) ECG-based risk. Column 6 compares Columns 4 vs. 5. The ECG high-risk group is set to the riskiest 12% of the sample, to match the fraction with high AHA 10-year risk. Hold-out patients with non-missing ECG and AHA risk scores are included (N=2,986).*

This finding raises the possibility that ECG-based screening may identify high-risk individuals who lack traditional risk factors like age or other known comorbidities. To investigate this, we compare characteristics of the ECG high-risk group to those with high AHA 10-year risk. Among our (reweighted) sample, 11.6% are considered high-risk (>20%) by the AHA 10-year risk score. To compare these high-risk patients to a high-risk ECG group, we similarly take the riskiest 11.6% of ECGs and classify them as high-risk. Comparing mean age between these two equally-sized groups shows that the ECG high-risk group is significantly younger than the AHA high-risk group (62.0 vs. 64.4, p=0.01).

Table 3 further tabulates patient characteristics in the four groups: high vs. low-or-intermediate risk, for ECG vs. AHA 10-year risk predictors. Among those with low or intermediate AHA 10-year risk, 8% appear high-risk according to the ECG model. These patients do in fact have high rates of prior MI, 5.0% (vs. 0.7% in those considered low-risk by both AHA and ECG, p=0.01), but lack many traditional risk factors: relative to patients considered low-risk by both scores, there are no significant differences in many modifiable risk factors—smoking, blood pressure, HbA1c, HDL—

and their total cholesterol and LDL are *lower* ($p<0.01$, $p<0.01$, respectively). The only traditionally high-risk factors present in these patients is that they are more likely to be older (61.2 vs. 51 years old, $p<0.01$), and male (62.1% vs. 33.9%, $p<0.01$). The rate of prior MI in this previously unsuspected high risk group is higher than the 59% of patients with high AHA risk score who have low-risk ECGs (5.9% vs. 2.3%, respectively, $p=0.08$). Appendix Table E1 confirms these results via regression analysis of prior MI based on both ECG-based risk and AHA 10-year scores: neither age alone, nor AHA 10-year risk, is significant for predicting prior MI once ECG-based risk is included. Identification of patients that are missed by traditional risk scoring is particularly important in the South Asian context, given that patients with MI in the INTERHEART study were lower-risk than expected based on traditional risk factors.[30]

Discussion

We trained and evaluated a model that detects prior MI using single-lead ECG data, suitable for use in community settings. It performed comparably to 12-lead ECG-based approaches and outperformed traditional, widely-used risk scores on PPV. This performance is encouraging because our training sample was small relative to most machine learning training datasets: ongoing data collection efforts are likely to substantially improve performance relative to this first version of the model.

Our findings have implications for screening policies, especially in low-resource settings, where most MIs go undiagnosed and untreated. In the general population, base rates of conditions like prior MI are low, so screening everyone is inefficient. But of course, screening no one misses patients who would derive large benefits, e.g., from secondary prevention. AI triage makes screening feasible and cost-effective, by cheaply identifying high-risk groups in whom further testing with echocardiography is highly cost-effective. We show that thresholds for referral can be selected using established economic cost-benefit frameworks, enabling policy makers to adapt the algorithm to health system priorities and constraints. This enables a practical, scalable way to bring ECG-based diagnostics closer to where people live, building on prior work.[31]

Our model outperforms existing risk predictors on cost-effectiveness. Tools like the AHA 10-year risk score, while appealing given its widespread use and validation, face practical limitations: it requires comprehensive patient histories and data elements like laboratory studies that are frequently unavailable or incomplete in remote areas with limited healthcare access. Similarly, machine learning tools that rely on electronic health record data are non-starters in environments without such data. Simpler screening strategies based on age are inaccurate, and risk prioritizing treatment for those who benefit the least from diagnosis of prior heart attack. A device-based approach with AI interpretation addresses both limitations by producing robust, high-quality predictions from inexpensive measurements across all age groups and demographics. Indeed, we find some evidence that ECG-based risk identifies new, previously unsuspected high-risk groups that appear low-risk on existing tools, while also correctly reclassifying many traditionally high-risk patients as low-risk.

In India specifically, ECG-based screening with AI could be easily integrated into existing public health architecture, Programs like the National Program for Prevention and Control of Cancer, Diabetes, Cardiovascular Diseases and Stroke (NPCDCS) already provide free at-home screening for hypertension and diabetes, with referral pathways to nearby clinical facilities for those in need of further testing and treatment. An ECG-based predictor could be incorporated into the toolkit of community health workers to broaden the coverage of diagnostic testing. Individuals who are referred and then diagnosed with prior MI and other cardiac conditions may then be covered under insurance schemes like Ayushman Bharat PM-JAY and other programs that cover secondary or tertiary care. In summary, our approach complements ongoing government efforts, filling an important gap in NCD prevention and early detection.

The potential of low-cost devices enhanced with AI extends beyond screening for prior MI. Our dataset includes a range of low-cost diagnostic data in addition to single-lead ECGs—retinal photography, vital signs, anthropometrics, strength and endurance tests, and blood glucose—paired with a number of clinical reference-standard tests: 12-lead ECGs, echocardiograms, chest X-rays, and comprehensive blood tests. It thus speaks to a broad set of health conditions with high global burden, including cardiovascular disease, chronic respiratory disease, diabetes, vision loss, and dementia. Many more AI models could be built from this dataset, thus bringing clinical-grade information to broad public health approaches. To support this goal, we make the full dataset, appropriately deidentified, publicly available for nonprofit use through Nightingale Open Science.[21] This allows other researchers to use it to develop, test, and deploy new algorithms across disease areas.

Several limitations merit discussion. First, the model was trained and validated on data from patients seeking health care in villages in Tamil Nadu, which may not generalize to other populations. It was also a walk-in sample that was likely healthier than our anticipated target population, and while we reweighted to account for this along the dimensions of age and sex, external validation in other regions and settings is needed. Second, our definition of prior MI is based on regional wall motion abnormalities on echocardiogram—an accepted proxy, but one that is less sensitive for smaller infarcts, particularly subendocardial ones, than cardiac MRI.[32–34] While we focused on regional and excluded global wall motion abnormalities, that are more consistent with cardiomyopathies than prior MI, the precision of these physician judgments is unknown. Third, although the model outperformed common risk scores, it is not yet a diagnostic tool. Deployment would require validation and monitoring performance in real-world settings. The generalizability of ECG-based risk measures to different populations must be tested, particularly given large differences in risk predicted from traditional risk factors across countries.[35] Fourth, our comparator models (e.g., the AHA risk score) are designed for broader cardiovascular risk prediction, not specifically to detect past events (though the two obviously should be correlated), and may be suboptimal baselines for detecting prior MI. Finally, while our cost-effectiveness analysis incorporates real data and uncertainty, compliance and downstream impacts—such as medication adherence, side effects, long-term health outcomes, and health system constraints—are not modeled.

Despite these limitations, our findings offer a concrete example of how AI can close key information gaps in the community, using data collected at very low cost. Paired with low-cost

hardware and health worker networks, such tools can reduce diagnostic delays, support earlier treatment, and make health systems more efficient. By sharing our data openly, we aim to catalyze further work that makes diagnostic intelligence as widespread as the diseases it aims to detect.

References

1. World Health Organization. Task shifting: rational redistribution of tasks among health workforce teams: global recommendations and guidelines. In: Task shifting: rational redistribution of tasks among health workforce teams: global recommendations and guidelines. pesquisa.bvsalud.org; 2007.

2. Banerjee A, Duflo E, Grela E, et al. Depression and loneliness among the elderly in low- and middle-income countries. J Econ Perspect 2023;37(2):179–202.

3. Manne-Goehler J, Geldsetzer P, Agoudavi K, et al. Health system performance for people with diabetes in 28 low- and middle-income countries: A cross-sectional study of nationally representative surveys. PLoS Med 2019;16(3):e1002751.

4. Siddharthan T, Pollard SL, Quaderi SA, et al. Discriminative accuracy of chronic obstructive pulmonary disease screening instruments in 3 low- and middle-income country settings. JAMA 2022;327(2):151–60.

5. Kondal D, Jeemon P, Manimunda S, et al. Structured lifestyle modification interventions involving frontline health workers for population-level blood pressure reduction: Results of a cluster randomized controlled trial in India (DISHA study). J Am Heart Assoc 2022;11(6):e023526.

6. Marston HR, Hadley R, Banks D, Duro MDCM. Mobile self-monitoring ECG devices to diagnose arrhythmia that coincide with palpitations: A scoping review. Healthcare (Basel) 2019;7(3):96.

7. Wong KC, Klimis H, Lowres N, von Huben A, Marschner S, Chow CK. Diagnostic accuracy of handheld electrocardiogram devices in detecting atrial fibrillation in adults in community versus hospital settings: a systematic review and meta-analysis. Heart 2020;106(16):1211–7.

8. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet 2019;394(10201):861–7.

9. Perez MV, Mahaffey KW, Hedlin H, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019;381(20):1909–17.

10. Svennberg E, Engdahl J, Al-Khalili F, Friberg L, Frykman V, Rosenqvist M. Mass screening for untreated atrial fibrillation: The STROKESTOP study: The STROKESTOP study. Circulation 2015;131(25):2176–84.

11. Kaushal A, Altman R, Langlotz C. Geographic distribution of US cohorts used to train deep learning algorithms. JAMA 2020;324(12):1212–3.

12. Beede E, Baylor E, Hersch F, et al. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy [Internet]. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM; 2020. Available from: http://dx.doi.org/10.1145/3313831.3376718

13. Turkbey EB, Nacif MS, Guo M, et al. Prevalence and correlates of myocardial scar in a US cohort. JAMA 2015;314(18):1945–54.

14. Zhang Z-M, Rautaharju PM, Prineas RJ, et al. Race and sex differences in the incidence and prognostic significance of silent myocardial infarction in the Atherosclerosis Risk in Communities (ARIC) study. Circulation 2016;133(22):2141–8.

15. Megiddo I, Chatterjee S, Nandi A, Laxminarayan R. Cost-effectiveness of treatment and secondary prevention of acute myocardial infarction in India: a modeling study. Glob Heart 2014;9(4):391–8.e3.

16. Basu S, Bendavid E, Sood N. Health and economic implications of national treatment coverage for cardiovascular disease in India: Cost-effectiveness analysis: Cost-effectiveness analysis. Circ Cardiovasc Qual Outcomes 2015;8(6):541–51.

17. Horan LG, Flowers NC, Johnson JC. Significance of the diagnostic Q wave of myocardial infarction. Circulation 1971;43(3):428–36.

18. Sandler LL, Pinnow EE, Lindsay J. The accuracy of electrocardiographic Q waves for the detection of prior myocardial infarction as assessed by a novel standard of reference: Accuracy of Q waves for the detection of prior MI. Clin Cardiol 2004;27(2):97–100.

19. Rogers AJ, Bhatia NK, Bandyopadhyay S, et al. Identification of cardiac wall motion abnormalities in diverse populations by deep learning of the electrocardiogram. NPJ Digit Med 2025;8(1):21.

20. Goff DC Jr, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2014;63(25 Pt B):2935–59.

21. Mullainathan S, Obermeyer Z. Solving medicine's data bottleneck: Nightingale Open Science. Nature Medicine 2022;28:897–9.

22. Ashfaq A, Lönn S, Nilsson H, et al. Data Resource Profile: Regional healthcare information platform in Halland, Sweden. Int J Epidemiol 2020;49(3):738–9f.

23. Attia ZI, Harmon DM, Dugan J, et al. Prospective evaluation of smartwatch-enabled detection of left ventricular dysfunction. Nat Med 2022;28(12):2497–503.

24. Makowski D, Pham T, Lau ZJ, et al. NeuroKit2: A Python toolbox for neurophysiological signal processing. Behav Res Methods 2021;53(4):1689–96.

25. Carleton T, Jina A, Delgado M, et al. Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits. Q J Econ 2022;137(4):2037–105.

26. Griffiths M, Maruszczak M, Kusel J. The who-choice cost-effectiveness Threshold: a Country-level analysis of changes over time. Value Health 2015;18(3):A88.

27. Gopinath N, Kaul U, Chadha SL, Sood AK, Bhattacharya D, Radhakrishnan S. Asymptomatic coronary heart disease detected on epidemiological survey of urban population of Delhi. Indian Heart J 1992;44(2):95–8.

28. Rao M, Xavier D, Devi P, et al. Prevalence, treatments and outcomes of coronary artery disease in Indians: A systematic review. Indian Heart J 2015;67(4):302–10.

29. Gupta R, Mohan I, Narula J. Trends in coronary heart disease epidemiology in India. Ann Glob Health 2016;82(2):307–15.

30. Sharma M, Ganguly NK. Premature coronary artery disease in Indians and its associated risk factors. Vasc Health Risk Manag 2005;1(3):217–25.

31. Khunte A, Sangha V, Oikonomou EK, et al. Automated diagnostic reports from images of electrocardiograms at the Point-of-care. medRxiv 2024;2024.02.17.24302976.

32. Jaarsma C, Schalla S, Cheriex EC, et al. Incremental value of cardiovascular magnetic resonance over echocardiography in the detection of acute and chronic myocardial infarction. J Cardiovasc Magn Reson 2013;15(1):5.

33. Gardner BI, Bingham SE, Allen MR, Blatter DD, Anderson JL. Cardiac magnetic resonance versus transthoracic echocardiography for the assessment of cardiac volumes and regional function after myocardial infarction: an intrasubject comparison using simultaneous intrasubject recordings. Cardiovasc Ultrasound 2009;7(1):38.

34. Wenzel J-P, Albrecht J-N, Toprak B, et al. Head-to-head comparison of cardiac magnetic resonance imaging and transthoracic echocardiography in the general population (MATCH). Clin Res Cardiol 2025;1–15.

35. Kaptoge S, Pennells L, De Bacquer D, et al. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. Lancet Glob Health 2019;7(10):e1332–45.