

Original Paper

Alignment Between Cardiologists and AI-Driven Diagnostic Systems: Mixed Methods Study

Mahdi Mahdavi^{1,2}, MSc, PhD; Sarah White³, PhD; Sandeep S Hothi^{4,5}, PhD; Chris Flood^{6,7}, PhD; Rosica Panayotova⁸, PhD; Daniel Frings⁹, PhD

¹Department of Management and Enterprise, Faculty of Business and Law, University of Roehampton, London, United Kingdom

²Bernard Lown Scholars in Cardiovascular Health Programme, Harvard T.H. Chan School of Public Health, Boston, MA, United States

³School of Health and Medical Sciences, City St George's, University of London, London, United Kingdom

⁴Department of Cardiology, Royal Wolverhampton NHS Trust, Wolverhampton, United Kingdom

⁵College of Medical and Dental Sciences, Institute of Cardiovascular Sciences, University of Birmingham, Birmingham, United Kingdom

⁶School of Nursing and Midwifery, London South Bank University, London, United Kingdom

⁷Central London Community Healthcare NHS Trust London, United Kingdom

⁸Stockport NHS Foundation Trust, Stockport, United Kingdom

⁹College of Health and Life Sciences, London South Bank University, London, United Kingdom

Corresponding Author:

Daniel Frings, PhD
College of Health and Life Sciences
London South Bank University
103 Borough Road
London SE1 0AA
United Kingdom
Phone: 44 20 7815 7815
Email: dan.frings@lsbu.ac.uk

Abstract

Background: The clinical value of artificial intelligence (AI)-based diagnostic systems depends not only on their accuracy but also on how well their outputs integrate with clinicians' judgments in practice. Critical knowledge gaps remain regarding diagnostic concordance between AI and clinicians in stress echocardiography interpretation, patient characteristics predicting discordance, and how cardiologists respond when AI recommendations conflict with their clinical diagnoses.

Objective: This study examined the diagnostic alignment between an AI-driven stress echocardiography system (EchoGo Pro [EGP]) and cardiologists' diagnoses of coronary artery disease (CAD), identified predictors of concordance and AI scan rejection, and explored cardiologists' decision-making strategies when disagreements arise.

Methods: We conducted mixed methods research. The quantitative study analyzed concordance between EGP and cardiologists using data from 854 participants with suspected CAD in the multicenter PROTEUS randomized controlled trial. Logistic regression identified predictors of agreement, disagreement, and scan rejection, adjusting for age, sex, smoking status, BMI, and cardiovascular risk factors (hypertension, hypercholesterolemia, diabetes, family history of CAD, and prior CAD events). To gain deeper insight into discordance, we conducted a qualitative study analyzing survey responses from 61 UK consultant cardiologists recruited via Qualtrics, exploring their perceptions of AI tools, the risks of following discordant AI recommendations, and their typical responses to AI-clinician disagreement.

Results: EGP and cardiologists agreed in 60% (512/854) of the cases, but agreement was significantly lower among patients with hypertension (OR 0.58, 95% CI 0.38-0.89; $P=.01$), diabetes (OR 0.56, 95% CI 0.35-0.90; $P=.02$), and pre-existing CAD (OR 0.48, 95% CI 0.30-0.77; $P=.002$). EGP rejected 26.1% (222/854) of the scans due to insufficient image quality, with rejection significantly more common in male patients ($\beta=0.35$; $P=.03$) and those with a family history of CAD. If a positive CAD diagnosis was assigned when either cardiologists or EGP identified CAD, the proportion of positive cases increased from 17.9% (153/854) to 22.1% (189/854), potentially identifying additional at-risk patients. Survey respondents (50/60, 85% male; 26/57, 46% aged 40-49 years; 39/61, 64% White) required 65% to 69% confidence in their initial diagnosis to justify disregarding contradictory AI recommendations. The survey findings revealed cardiologists treated AI recommendations as advisory rather than definitive. When facing discordance, they retained confidence in their judgment and sought corroboration through additional testing, data review, or second opinions rather than deferring to AI. Paradoxically, cardiologists with higher

confidence in AI tools required greater confidence in their own diagnosis to disregard AI recommendations ($\beta=7.73$; $P=.02$). Cardiologists attributed discordance primarily to AI's inability to incorporate patient history, comorbidities, and broader clinical context.

Conclusions: EGP shows promise as an adjunctive tool but struggles with multimorbid patients and exhibits high, uneven rejection rates. Cardiologists use AI to prompt scrutiny, not replace judgment. Future systems need to integrate wider patient data with imaging and minimize bias through representative training to avoid exacerbating inequities.

Trial Registration: ClinicalTrials.gov NCT05028179; <https://clinicaltrials.gov/study/NCT05028179>

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2023-079617

J Med Internet Res 2026;28:e83541; doi: [10.2196/83541](https://doi.org/10.2196/83541)

Keywords: cardiac imaging; echocardiography; stress; clinical reasoning; artificial intelligence; cardiology

Introduction

The alignment between artificial intelligence (AI) assessments and human interpretations of health conditions influences clinicians' confidence in diagnostic accuracy and the perceived use of AI systems in practice [1-3]. AI-driven tools have shown considerable promise in image processing in health care [4,5]. However, the clinical value of these tools depends not only on their accuracy against a reference standard but also on how their outputs integrate with clinicians' judgments [6].

Studies on AI applied to health care suggest that clinicians are generally optimistic about AI but remain cautious, emphasizing transparency, bias, and responsibility, and often describe AI as an advisory rather than authoritative tool [6,7]. When AI and clinician judgments converge, this alignment can support timely and cost-effective care by reinforcing decision-makers' confidence and potentially reducing the need for further testing [8]. However, when discordance arises, the extent to which clinicians rely on AI versus their own judgment and the factors driving such discrepancies are less well understood [7]. Therefore, understanding how AI outputs are interpreted in cases of discordance, the risks associated with following AI recommendations, and the conditions under which clinicians rely on AI rather than their own judgment is essential to fully harness the benefits of AI systems.

Stress echocardiography (SE) is a widely performed clinical investigation for the identification of coronary artery disease (CAD) [9]. It generates images of left ventricular wall motion acquired before and during exercise and/or pharmacological stress to make a clinical judgment on whether CAD is present, and which areas of the heart are affected. However, due to a degree of subjectivity in the test's interpretation, clinical judgments may vary between clinicians [10]. Interobserver variability can lead to inconsistent classification of ischemia and variation in the patient journey, including referral for invasive angiography [11,12]. Additionally, SE is a resource-intensive test, requiring trained specialists to both perform and interpret the test. As a result, the demand for SE imaging is outpacing the capacity of cardiac specialists in many countries, particularly those with aging populations including the United Kingdom [13].

In SE tests, AI-based models have demonstrated high diagnostic accuracy for the detection and risk stratification of CAD, often achieving performance comparable to expert readers [6,10,14]. Their consistent and automated performance can enhance the accuracy of CAD diagnosis, while also reducing reporting time, providing a cost-effective second opinion, and minimizing interobserver variability [15,16].

EchoGo Pro (EGP), an AI-driven SE analysis system, has been developed for cardiac assessments and identifying potential ischemia [17]. EGP automatically analyzes SE images to estimate the probability of CAD and produces a structured binary risk report intended to support clinician decision-making. Initial validation studies have reported encouraging diagnostic accuracy and suggested that incorporating EGP into SE workflows may improve the detection of severe CAD compared with conventional interpretation alone [17].

Although systematic reviews consistently show that AI systems can achieve diagnostic performance comparable to health care professionals [3,6,18], most studies focus on accuracy metrics rather than on how clinicians respond in practice. Emerging evidence indicates that clinicians rarely respond to discordance in a simple accept-reject manner. Instead, they may ignore, selectively incorporate, negotiate, or conditionally defer to AI recommendations depending on their initial level of confidence in their own judgment, the perceived severity of the patient's condition, and how clearly and convincingly the AI system explains its recommendation [19-21].

Yet, existing evidence is limited in scope and often based on noncardiology contexts, with little known about clinician behavior when AI outputs conflict with their own judgment, particularly in time-sensitive settings such as cardiac care [22]. Evidence is also scarce on which patient characteristics are associated with AI-clinician discordance, which factors predict AI rejection of scans, how cardiologists weigh AI outputs against their own judgment, and how perceived risk influences their willingness to follow AI recommendations [23,24]. Understanding how AI-clinician interaction unfolds in cardiac care, where rapid decisions can be lifesaving, is therefore critical for safe and effective implementation [24].

This mixed methods study integrates quantitative and qualitative analyses. The quantitative study uses clinical trial

data to assess concordance between EGP and clinicians, examine patient characteristics associated with concordant and discordant cases, and identify demographic and clinical profiles linked to AI scan rejection or diagnostic divergence. The qualitative component explores cardiologists' broader perceptions and experiences with AI tools, providing insight into attitudes, barriers, and contextual factors influencing adoption and trust. Together, these findings inform future development and implementation strategies for AI in cardiology. The analysis plan was preregistered on the Open Science Framework.

Methods

Overview of Research Design

We conducted a mixed methods study with 2 components, following a prespecified protocol published earlier [25]. The first was a secondary analysis of diagnostic agreement between EGP and cardiologists' assessments of CAD in SE images. This analysis used data from 854 participants in the PROTEUS trial (for a broader evaluation of EGP, see below) [26]. The second component comprised a qualitative survey of consultant cardiologists, featuring open-ended questions designed to explore their perceptions and responses to disagreements with AI tools. Notably, this component considered AI tools in general rather than focusing specifically on EGP. Reporting of the agreement analysis adhered to the Guidelines for Reporting Reliability and Agreement Studies [27]. The survey component followed the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist for cross-sectional studies [28].

The Experimental Study

The PROTEUS trial was a randomized, multicenter, 2-armed study aimed at participants referred to SE clinics for the investigation of suspected CAD. The details of the PROTEUS research design are published by Woodward et al [26]. The trial randomized participants into a control group (SE assessment without EGP report) or an intervention group (SE assessment with EGP report). In the intervention group, SE images were analyzed by EGP, and a binary report produced by EGP indicating the probability of severe CAD was provided to the clinician alongside their own judgment [26]. Images failing quality criteria were flagged with a rejection report.

Cases were categorized as either concordant or discordant based on the alignment between EGP and clinician interpretation. Clinician interpretation was classified as positive for CAD if a stress-induced wall motion abnormality was observed, negative for CAD if no wall motion abnormality was induced, or inconclusive if the interpretation was abandoned or deemed indeterminate for inducible ischemia. The EGP risk rating classified cases as high risk of CAD (coronary stenosis $\geq 70\%$), low risk (coronary stenosis $\leq 70\%$), or uninterpretable if the image quality was insufficient for assessment.

Concordance was defined as agreement between the clinician and the EGP in either of the following scenarios: (1) a positive clinician interpretation and a high EGP risk rating or (2) a negative clinician interpretation and a low EGP risk rating. Discordance was defined as (1) a positive clinician interpretation paired with a low EGP risk rating or (2) a negative clinician interpretation paired with a high EGP risk rating.

The control variables consisted of age, sex, smoking status, BMI, and cardiovascular risk factors (CRFs). The data were collected through a combination of self-report questionnaires, physical measurements, and laboratory assessments. Age was categorized into the following groups: 18 to 39, 40 to 59, 60 to 80, and ≥ 81 years of age. Smoking status was self-reported and comprised current smokers, ex-smokers, or nonsmokers. BMI was calculated from measured height and weight, which were obtained during clinical examination. CRFs included hypertension, hypercholesterolemia, diabetes mellitus, positive family history of CAD, and history of CAD (including prior clinical events). Family history of CAD was self-reported, while other CRFs were determined through a combination of medical history review, physical measurements, and laboratory results. Specifically, hypertension was defined as measured blood pressure greater than or equal to 140/90 mm Hg or current use of antihypertensive medications [29]. Diabetes mellitus was diagnosed based on laboratory values (hemoglobin A_{1c} $\geq 6.5\%$ [48 mmol/mol] or fasting blood sugar >126 mg/dL) or current use of antidiabetic medications [30]. Hypercholesterolemia and history of CAD were ascertained from medical records and confirmed through patient interview.

Descriptive statistics were used for demographics and clinical characteristics. Means and SDs were reported for normally distributed continuous variables, and frequencies and percentages were reported for categorical variables. We excluded cases where EGP reports were not used due to arriving late or being rejected ($n=222$, 26%). Logistic regression was used to identify the predictors of concordance or discordance and factors associated with AI image rejection. We used R for statistical analysis.

The Survey

We collected survey responses from consultant cardiologists with expertise in SE (as a convenience sample) using a secure online survey platform and recruitment panel (Qualtrics). Demographic items were adapted from the UK Office for National Statistics framework [31], and screening questions ensured participant eligibility.

We included explanatory variables in the regression models to explore factors influencing cardiologists' diagnostic confidence and their responses to AI recommendations. Years of experience as a cardiologist were grouped into 4 categories: 1 to 5, 6 to 10, 11 to 15, and 16 or more. The number of monthly SE reviews was categorized as follows: 0 to 20 tests, 21 to 50 tests, 51 to 100 tests, and more than 100 tests. Use of AI in the cardiac care pathway assessed exposure to AI in clinical practice. Participants were asked whether they used AI tools in the diagnosis or management of cardiac

patients (yes or no). Confidence in AI was measured using a 5-point Likert scale (1 = very low, 5 = very high). Respondents also indicated the level of confidence (on a scale from 1% to 100%) they would require in their initial diagnosis to disregard AI advice.

In addition, open-ended questions explored perceived risks of following AI in scenarios where AI tools contradicted clinicians' judgment. Participants provided free-form text describing the potential risks of following the AI recommendation and further actions in response to two scenarios: (1) AI tools contradict the clinician's diagnosis of CAD being present and (2) AI tools contradict the clinician's diagnosis of CAD not being present. Of the 88 responses received, 61 met the inclusion criteria and were retained for analysis.

We used a mixed methods approach to analyze survey responses. Quantitative data were analyzed using regression models to examine associations between clinician characteristics (eg, years of experience, number of SE reviews per month, use of AI, and confidence in AI tools) and their reported confidence in diagnostic decision-making in hypothetical scenarios. In parallel, qualitative thematic analysis was conducted on open-text responses. This involved content analysis to identify common themes related to perceived risks of following AI recommendations and typical follow-up actions taken when AI tools and clinical assessments diverged.

The thematic analysis was based on the 6-phase framework of Braun and Clarke [32]: (1) familiarizing oneself with the data, (2) generating initial qualitative codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing a report. Two researchers independently coded responses and resolved discrepancies through discussion. Given the anonymous nature of the survey, member checking was not possible. However,

qualitative themes were reviewed by 3 consultant cardiologists from PROTEUS field sites to enhance validity.

Ethical Considerations

This study comprises 2 components: a clinical trial and a survey study. For the clinical trial, the protocol, informed consent form, participant information leaflet, and proposed advertising materials were reviewed by the Patient and Public Involvement and Engagement (PPIE) group and approved by the study sponsor and the North West—Preston Research Ethics Committee (REC) of the NHS Health Research Authority (reference: 21/NW/0199). The PROTEUS trial was registered on August 31, 2021, at ClinicalTrials.gov with registration number NCT05028179. For the survey study, ethical approval was obtained from the NHS Health Research Authority (IRAS number 315284) and the London South Bank University Ethics Panel (ETH2223-0164). All participants provided informed consent, and anonymized data were used for both study components. Survey participants were compensated for their participation via Qualtrics, with payments administered directly by the platform rather than the study team.

Results

The Experimental Study Results

Overview

Data from 854 participants in the intervention arm of the PROTEUS trial were used in the experimental analysis. Regional wall motion abnormalities were detected in 8% (71/893) of the participants, and 18% (156/893) received a clinical diagnosis of positive CAD (Table 1). Further details on participant demographics, lifestyle factors, and clinical history are available in Multimedia Appendix 1.

Table 1. Descriptive statistics and bivariate logistic regression results—primary outcome is concordance between clinician and EchoGo Pro (EGP) or divergence.

Patient characteristics	Convergent ^a (n=512)	Divergent ^b (n=113)	OR (95% CI)	P value
Sex, n				
Female	245	50	1	— ^c
Male	267	63	0.86 (0.57-1.30)	.49
Age group (y), n				
18-39	20	3	1	—
40-59	179	27	0.99 (0.28-3.57)	.99
60-80	280	79	0.53 (0.15-1.84)	.32
>81	33	4	1.24 (0.25-6.11)	.79
Smoking status, n				
Current smoker	52	11	1	—
Ex-smoker	216	41	1.11 (0.54-2.32)	.77
Nonsmoker	240	60	0.85 (0.42-1.72)	.64
BMI, mean (SD)	29.0 (10.8)	29.8 (6.2)	0.99 (0.98-1.01)	.45
Control variables				
Has cardiovascular risk factors?, n	406	102	0.41 (0.21-0.80)	.008

Patient characteristics	Convergent ^a (n=512)	Divergent ^b (n=113)	OR (95% CI)	P value
Number of cardiovascular risk factors, mean (SD)	1.5 (1.2)	1.9 (1.0)	0.77 (0.65-0.92)	.003
Hypertension, n	254	71	0.58 (0.38-0.89)	.01
Hypercholesterolemia, n	214	58	0.68 (0.45-1.02)	.07
Diabetes, n	90	31	0.56 (0.35-0.90)	.02
Positive family history of coronary artery disease, n	182	41	0.97 (0.63-1.48)	.88
Previous CAD ^d clinical events, n	44	14	0.66 (0.35-1.26)	.21
Pre-existing coronary artery disease, n	85	33	0.48 (0.30-0.77)	.002

^aConcordance refers to (1) both clinician and EchoGo Pro (EGP)-confirmed presence of coronary artery disease (CAD) or (2) both clinician- and EGP-confirmed absence of CAD.

^bDiscordance refers to (1) clinician-diagnosed CAD, but EGP did not confirm it, or (2) EGP-diagnosed CAD, but the clinician did not confirm it. The analysis focuses on data from the experimental study conducted specifically with the EGP tool.

^cNot applicable.

^dCAD: coronary artery disease.

Concordance and Discordance

Of the analyzed cases, 512 (60%) were classified as concordant (Table 1). Within this group, 38 (4.5%) cases demonstrated concordance on a positive CAD diagnosis, where both the clinician identified inducible ischemia and EGP reported a high risk of CAD. The remaining 474 (55.5%) cases were concordant on a negative diagnosis, with the clinician reporting no inducible ischemia and EGP indicating low risk.

A total of 113 (13.2%) cases were classified as discordant. Of these, 36 (4.2%) cases involved a clinician-assessed negative result for inducible ischemia, while EGP reported a high risk of CAD. Conversely, 77 (9%) cases involved a positive clinical interpretation for inducible ischemia, but EGP assessed the case as low risk.

If a positive diagnosis were assigned based on either a positive clinician interpretation or a high-risk EGP rating, the

total number of positive CAD diagnoses would increase from 153 (17.9%) to 189 (22.1%).

EGP rejected 222 (26.1%) cases due to insufficient image quality, whereas clinicians classified only 10 (1.2%) cases as inconclusive. Concordance among rejected cases was very low, with only 3 (0.4%) cases showing agreement between EGP rejection and clinician inability to interpret.

Factors Predicting Concordance or Discordance

Table 2 presents the comparison of concordance and discordance rates across sex, age groups, and smoking status. Bivariate logistic regression analyses indicated no statistically significant differences in concordance or discordance between male and female participants, across age groups, or among different smoking statuses.

Table 2. Concordance and discordance between clinician interpretation and EchoGo Pro (EGP) report (experimental study).

Clinician interpretation of presence or ischemia ^b	Artificial intelligence risk rating report ^a			Total, n (%)
	High, n (%)	Low, n (%)	Rejected, n (%)	
Positive for inducible ischemia	38 (4.5)	77 (9)	38 (4.5)	153 (17.9)
Negative for inducible ischemia	36 (4.2)	474 (55.5)	181 (21.2)	691 (80.9)
Inconclusive or abandoned	0 (0)	7 (0.8)	3 (0.4)	10 (1.2)
Total	74 (8.7)	558 (65.3)	222 (26.1)	854 (100)

^aArtificial intelligence risk rating report was classified as high risk of CAD (stenosis ≥70%) or low risk (stenosis <70%), or the stress echocardiography image was rejected due to poor quality.

^bClinician interpretation of the presence of ischemia was classified as a positive coronary artery disease if they diagnosed that wall motion abnormality was present, negative coronary artery disease if no wall motion abnormality was present, or the interpretation was inconclusive or abandoned for inducible ischemia. The analysis focuses on the data from the experimental study conducted specifically with the EchoGo Pro tool.

Having more CRFs or pre-existing CAD was associated with lower odds of concordance. Patients in concordant cases had fewer cardiovascular risk factors than those in discordant cases (mean 1.5, SD 1.2 vs mean 1.9, SD 1.0). Each additional cardiovascular risk factor was associated with lower odds of concordance between EGP and clinician interpretations (OR 0.77, 95% CI 0.65-0.92; $P=.003$). The strongest associations with discordance were observed in patients with any cardiovascular risk factor (OR 0.41, 95% CI 0.21-0.80; $P=.008$), pre-existing CAD (OR 0.48, 95% CI 0.30-0.77; $P=.002$), and diabetes (OR 0.56, 95% CI

0.35-0.90; $P=.02$). However, multivariate logistic regression suggested that no sociodemographic variable was associated with discordance or concordance.

EGP Rejecting Scans

We ran a multivariable logistic regression to examine whether sex, age group, smoking status, BMI, and cardiovascular risk factors were associated with the likelihood of SE image rejection by EGP (Table 3). Sex was the only demographic variable significantly associated with rejection: scans from male participants were more likely to be rejected than those

from female participants ($\beta=0.35$, $P=.03$). Among clinical variables, hypercholesterolemia and family history of CAD showed significant associations with scan rejection. Individuals with hypercholesterolemia were less likely to

have their scans rejected ($\beta=-0.36$; $P=.04$), whereas those with a positive family history of CAD were more likely to experience rejection ($\beta=0.33$; $P=.049$).

Table 3. Multivariate logistic regression results—primary outcome is artificial intelligence (AI) rejection of scan or conclusive scan^a.

Patient characteristics	Coefficient (SE)	P value
Intercept	-1.28 (0.56)	.02
Age groups (y)		
18-39	— ^b	—
40-59	0.02 (0.46)	.97
60-80	0.31 (0.45)	.50
>81	0.69 (0.53)	.19
Sex		
Female	—	—
Male	0.35 (0.17)	.03
Smoking		
Current smoker	—	—
Ex-smoker	-0.20 (0.28)	.48
Nonsmoker	-0.11 (0.27)	.68
BMI	0.00 (0.01)	.78
Cardiovascular risk factors		
Hypertension	0.03 (0.17)	.87
Hypercholesterolemia	-0.36 (0.17)	.04
Diabetes	-0.04 (0.21)	.85
Positive family history of coronary artery disease	0.33 (0.17)	.049
Previous CAD ^c clinical events	0.20 (0.26)	.43

^aThe analysis focuses on the data from the experimental study conducted specifically with the EchoGo Pro tool.

^bNot applicable.

^cCAD: coronary artery disease.

Survey Results

Overview

The data from 61 survey participants were included in the analysis. Most were male (50/60, 85%), aged 40 to 49 years (26/57, 46%), and White (39/61, 64%). Professional experience varied, with the majority having more than 5 years of experience as cardiologists (Multimedia Appendix 1). Over half of the survey participants reported using AI tools in cardiac care (33/61, 54%), including auto-indexing software (33/61, 54%) and decision-support software (23/61, 38%). The number of monthly SE tests reviewed ranged from 1 to over 100, with nearly half reviewing 21 to 50 tests per month.

Scenario Analysis

The scenario analysis used survey responses in which AI tools were considered in general rather than focusing on a specific system such as EGP. In scenario A (the cardiologist

diagnosed CAD, but the AI tools contradicted this), the mean level of confidence in the initial diagnosis required to justify taking no further action was 69%. In scenario B (CAD was not identified by the cardiologist but was diagnosed by the AI tools), the corresponding mean confidence level was 65%.

Table 4 presents linear regression models examining how cardiologist characteristics relate to the confidence threshold required to disregard AI advice. Across both scenarios, this threshold was not significantly associated with years of experience as a cardiologist, the number of stress echocardiograms reviewed per month, or the current use of AI in the cardiac care pathway. However, in scenario A, higher confidence in AI tools was positively associated with the level of diagnostic confidence required to disregard an AI recommendation ($\beta=7.73$; $P=.02$). This suggests that cardiologists who are more trusting of AI systems are more cautious and require stronger conviction in their own clinical judgment before overriding AI-generated assessments.

Table 4. Regression models for scenario analysis: cardiologists' confidence in their initial diagnosis^a.

Predictor variable	Scenario A: cardiologist made a diagnosis of CAD ^b , but the AI ^c tools contradicted ^d		Scenario B: CAD was not identified by cardiologists, but was diagnosed by AI tools ^e	
	Coefficient (SE)	P value	Coefficient (SE)	P value

Predictor variable	Scenario A: cardiologist made a diagnosis of CAD ^b , but the AI ^c tools contradicted ^d		Scenario B: CAD was not identified by cardiologists, but was diagnosed by AI tools ^e	
Intercept	39.37 (12.07)	.002	55.49 (14.26)	<.001
Years of experience as a cardiologist				
1-5 (reference)	— ^f	—	—	—
6-10	-1.51 (7.39)	.84	0.45 (8.74)	.96
11-15	10.21 (7.48)	.18	-1.21 (8.85)	.89
>16	0.90 (7.19)	.90	-0.52 (8.50)	.95
Number of monthly stress test reviews				
0-20 tests (reference)	—	—	—	—
21-50 tests	3.61 (6.83)	.60	-11.57 (8.07)	.16
51-100 tests	4.13 (8.92)	.65	-8.72 (10.54)	.41
>101 tests	4.90 (7.83)	.53	-3.93 (9.26)	.67
AI in the cardiac care pathway				
No (reference)	—	—	—	—
Yes	-3.68 (5.06)	.47	2.88 (5.98)	.63
Confidence in AI tool ^g	7.73 (3.11)	.02	4.58 (3.67)	.22

^aThe analysis presented in this table is based on survey data where artificial intelligence tools are considered in general.

^bCAD: coronary artery disease.

^cAI: artificial intelligence.

^dConfidence in initial diagnosis for not taking any further actions: 69%.

^eConfidence in initial diagnosis for not taking any further actions: 65%.

^fNot applicable.

^gConfidence in AI was measured on a 5-point Likert scale (1=very low, 5=very high). Confidence in their own diagnosis was measured on a scale from 1% to 100%. Regression coefficients are unstandardized.

Perceived Risk of Following AI Recommendations

In scenario A, the most perceived risk of following AI recommendations, considered broadly rather than specific to EGP, was the possibility of inaccurate diagnoses or inappropriate treatments, which could potentially cause harm. Other prominent concerns included AI limitations, such as lack of patient-specific context, inadequate training, and the early-stage development of AI tools, as well as legal liability, with cardiologists remaining ultimately responsible for patient care. Some participants, however, perceived minimal risk, viewing AI as a nonbinding decision-support tool.

In scenario B, the risk of inaccurate diagnosis emerged as a key theme in participants' perceptions of the risks associated with following AI recommendations. Two additional prominent themes were the overuse of clinically ineffective but nonharmful services and the misuse of care—specifically, unnecessary treatments that could potentially harm patients or cause side effects. Participants expressed these concerns with statements such as: “Some treatments may not improve patient symptoms. Unnecessary intervention could be harmful,” and “Overtreatment, possible side effects of therapy, patient anxiety.”

Actions Taken Following AI Recommendations

Study participants were responsive to situations in which AI recommendations contradicted cardiologists' clinical judgment regarding the presence or absence of CAD. Across both scenarios, the most frequently reported response was to

initiate additional investigations. Another prominent theme was the pursuit of a second opinion, consultation, or advice, reflected in subthemes such as seeking a second opinion, requesting advice, or consulting peer colleagues. A third major theme involved reviewing broader patient data, including diagnostic tests and medical records, to verify or confirm information through a more comprehensive assessment. In a small number of cases, participants reported taking no further action. Finally, a less common theme was the repetition of tests or re-examination of existing information, suggesting an intent to audit or reassess the initial diagnosis and underlying data.

Discussion

This mixed methods study combined PROTEUS trial data analysis with a cardiologist survey, examining how EGP's stress echo assessments aligned with clinician interpretations and exploring attitudes toward AI-human discordance. EGP showed substantial overall agreement with clinicians but generated meaningful discordance, concentrated among patients with multiple cardiovascular risk factors and established CAD. Concordance was similar across sex, age, and smoking status but lower in patients with hypertension, diabetes, or pre-existing CAD. Over one-quarter of scans were rejected due to poor image quality, with rejection more common in male patients and those with a family history of CAD, and less frequent in patients with hypercholesterolemia.

Survey respondents described AI as supplementary rather than primary decision support. When facing discordance,

cardiologists reported retaining confidence in their initial judgment and seeking corroboration through further investigations, clinical data review, or second opinions rather than deferring to AI. Those who expressed higher overall confidence in AI reported that they would require stronger confidence in their own diagnosis before disregarding an AI recommendation, particularly when the AI suggested disease they had not initially identified. Some expressed concerns that following discordant AI advice could trigger unnecessary investigations and overtreatment.

The concentration of discordance in patients with complex profiles highlights the limitations of image-only approaches [33]. This is clinically significant because patients with hypertension, diabetes, and pre-existing CAD face substantially higher risks of future events [34], yet accurate diagnosis in these critical groups is where EGP and clinicians most often diverged. EGP is trained on images [35], whereas clinicians integrate clinical history, comorbidities, and broader patient data. For example, patient-specific anatomical features can contribute to false positive findings by creating artifacts that mimic ischemia, highlighting how structural features influence diagnostic variability [36]. Consequently, future systems must move toward multimodal designs that incorporate anatomical features, clinical history, and comorbidities to better mirror diagnostic reasoning [23, 36,37].

These findings highlight both EGP's potential and its limitations. When used alongside expert interpretation, EGP can flag additional high-risk patients who might otherwise be missed. Given that missed CAD diagnoses cost approximately £4200 per patient versus £2300 for correct diagnoses [38], even modest gains in diagnostic yield may improve outcomes and reduce downstream costs. However, EGP alone often classified fewer positives than clinicians and failed to interpret rejected scans, reinforcing its current role as an adjunctive decision-support tool rather than a stand-alone solution.

We observed a nonsignificant trend toward increased false positives with EGP. At population scale, even minor increases affect service planning; a theoretical 1% rise in false positives across the United Kingdom's 61,000 annual scans [35] would trigger approximately 610 unnecessary investigations, costing roughly £1.4 million. Beyond financial burdens, false positives generate patient anxiety and avoidable invasive procedures. Consequently, the trade-offs between enhanced sensitivity and the risks of over-diagnosis warrant careful consideration during AI upscaling.

A further challenge relates to EGP's relatively high rate of scan rejection, which undermines its reliability and usability. Frequent noninterpretable outputs are likely to erode clinicians' trust and limit the perceived usefulness of the system. Importantly, rejection was not uniform. Although mechanisms cannot be determined directly from our data, these differences are consistent with broader evidence that imbalances in training datasets, such as sex disparities or underrepresentation of certain clinical risk profiles, can lead to biased performance of AI models across demographic and

clinical groups [39-41]. Addressing these inequities requires more representative training data, routine postdeployment monitoring across subgroups, and model refinement where necessary.

The survey findings clarify how such tools are likely to be used. Cardiologists framed AI as supportive rather than authoritative, responding to disagreements by seeking further tests, reviewing additional data, or obtaining second opinions, suggesting a tendency toward confirmation-oriented behavior rather than automation bias [22,42]. While this caution mitigates overreliance risks, frequent discordance may paradoxically increase downstream testing and costs. Clear guidance on managing AI-clinician disagreement and warranting additional investigations is therefore essential for implementation.

This study has several notable strengths. Leveraging data from the multicenter PROTEUS trial provided a robust and standardized framework for evaluating EGP across diverse clinical sites, reducing the influence of local practice patterns and supporting internal validity. The sizeable sample allowed the examination of agreement, disagreement, and scan rejection across clinically relevant subgroups. By explicitly modeling predictors of AI-clinician concordance and AI rejection, we begin to identify which patients may be better or worse served by current image-based systems. Finally, the mixed methods design offers a holistic understanding of both EGP's technical behavior and how clinicians interpret and act on its recommendations in practice.

A few limitations should be borne in mind. First, the high scan rejection rate reduced the effective sample size for concordance analyses and may have introduced selection bias if rejected scans systematically differed from accepted scans in unmeasured ways. Second, although we included a range of demographic and clinical characteristics, other potentially important determinants of image quality and diagnostic decision-making, such as sonographer expertise, variation in acquisition protocols, and differences in operators' familiarity with EGP, were unavailable but likely influence both AI performance and clinician interpretations [43,44]. Finally, both study components were conducted within a specific national health care context; EGP's performance and acceptance may differ in other health systems or populations.

Beyond EGP, these findings have broader implications for AI implementation in cardiovascular imaging. Discordance concentrated in patients with hypertension, diabetes, and prior CAD reflects the limitations of image-only models that cannot incorporate symptoms, medications, and comorbidities as clinicians do. Future AI systems must integrate these clinical variables. The uneven rejection rates by sex and clinical profile require ongoing monitoring for bias after deployment. Additionally, cardiologists' concerns about legal liability when AI recommendations conflict with their judgment must be addressed through clear clinical protocols and regulatory guidance. Further research is needed to establish evidence-based approaches for managing AI-clinician disagreement and ensuring equitable performance across patient populations.

In this mixed methods study, EGP agreed with cardiologists in most cases, but discordance clustered in patients with multiple cardiovascular risk factors or established coronary disease, and many scans were rejected, particularly in some patient groups. Cardiologists described using AI as supportive input rather than a decision-maker, typically responding to AI-clinician disagreement by ordering further tests, reviewing additional data, or seeking second opinions. Overall, these

findings position EGP and similar tools as useful adjuncts that can highlight additional at-risk patients but cannot yet replace holistic clinical assessment. Future development should prioritize incorporating key clinical information alongside imaging, reducing and monitoring rejection and bias across subgroups, and embedding AI within clear clinical pathways so that disagreement is managed transparently and patient safety, efficiency, and equity are all improved.

Acknowledgments

The authors would like to thank Virgil Griffith and Sharanya Jayadev for their valuable support to this research. They are also grateful to the PROTEUS trial and Ultromics Ltd for providing access to the data that made this research possible. Finally, we extend our appreciation to all those who contributed to the completion of this study.

The authors declare the use of generative artificial intelligence (GAI) in the writing process. According to the GAIDeT taxonomy (2025), GAI tools were used to assist in proofreading and editing under full human supervision. Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Funding

This work was supported by NHSX, the former joint digital transformation unit of National health services (NHS) England and the Department of Health and Social Care (United Kingdom). This report is independent research funded by the National Health Services (AI Award in Health and Care, Technology Specific Evaluation Team). The views expressed are those of the authors and not necessarily those of the National Health Services, NHSX, or the Department of Health and Social Care.

Data Availability

The survey data generated and analyzed during this study are available from the corresponding author upon reasonable request, subject to a data-sharing agreement and approval from the relevant ethics committee. The experimental study data, collected as part of the PROTEUS trial, can be requested directly from Ultromics Ltd, in accordance with their data-sharing policies and applicable approvals.

Authors' Contributions

DF, CF, and MM conceptualized the study. SH and RP reviewed and commented on study materials. MM prepared the data and conducted the analyses, supported by SW. MM and DF prepared the initial draft of the manuscript. All authors contributed to the review and editing of the manuscript.

Conflicts of Interest

SH has received speaker honoraria from Ligence AI, has received research funding for research agreements with Ventripoint Medical Systems and Ligence AI, and has held a research agreement with funding from Ultromics. The other authors declare no conflicts of interest.

Multimedia Appendix 1

Additional descriptive statistics and subgroup analyses of participants in the experimental study and cardiologist survey. [[DOCX File \(Microsoft Word File\), 139 KB-Multimedia Appendix 1](#)]

References

1. Kompa B, Snoek J, Beam AL. Second opinion needed: communicating uncertainty in medical machine learning. *NPJ Digit Med*. Jan 5, 2021;4(1):4. [doi: [10.1038/s41746-020-00367-3](https://doi.org/10.1038/s41746-020-00367-3)] [Medline: [33402680](https://pubmed.ncbi.nlm.nih.gov/33402680/)]
2. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. *JAMA*. Dec 4, 2018;320(21):2199-2200. [doi: [10.1001/jama.2018.17163](https://doi.org/10.1001/jama.2018.17163)] [Medline: [30398550](https://pubmed.ncbi.nlm.nih.gov/30398550/)]
3. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. Jan 2019;25(1):44-56. [doi: [10.1038/s41591-018-0300-7](https://doi.org/10.1038/s41591-018-0300-7)] [Medline: [30617339](https://pubmed.ncbi.nlm.nih.gov/30617339/)]
4. Song Y, Ren S, Lu Y, Fu X, Wong KKL. Deep learning-based automatic segmentation of images in cardiac radiography: a promising challenge. *Comput Methods Programs Biomed*. Jun 2022;220:106821. [doi: [10.1016/j.cmpb.2022.106821](https://doi.org/10.1016/j.cmpb.2022.106821)] [Medline: [35487181](https://pubmed.ncbi.nlm.nih.gov/35487181/)]
5. Raghu D, Tripathy HK. Medical image analysis using deep learning algorithm convolutional neural networks. In: Pandian AP, Fernando X, Haoxiang W, editors. *Computer Networks, Big Data and IoT*. Vol. 2022:117. 245-256. [doi: [10.1007/978-981-19-0898-9_19](https://doi.org/10.1007/978-981-19-0898-9_19)]

6. Liu B, Reis J, Sharma A, Wang W. Application of artificial intelligence in non-invasive cardiovascular imaging for coronary artery disease: a systematic review and meta-analysis. *Front Cardiovasc Med*. 2025;12:1664183. [doi: [10.3389/fcvm.2025.1664183](https://doi.org/10.3389/fcvm.2025.1664183)] [Medline: [41458983](https://pubmed.ncbi.nlm.nih.gov/41458983/)]
7. Shamszare H, Choudhury A. Clinicians' perceptions of artificial intelligence: focus on workload, risk, trust, clinical decision making, and clinical integration. *Healthcare (Basel)*. Aug 16, 2023;11(16):2308. [doi: [10.3390/healthcare11162308](https://doi.org/10.3390/healthcare11162308)] [Medline: [37628506](https://pubmed.ncbi.nlm.nih.gov/37628506/)]
8. El Arab RA, Al Moosa OA. Systematic review of cost effectiveness and budget impact of artificial intelligence in healthcare. *NPJ Digit Med*. Aug 26, 2025;8(1):548. [doi: [10.1038/s41746-025-01722-y](https://doi.org/10.1038/s41746-025-01722-y)] [Medline: [40858882](https://pubmed.ncbi.nlm.nih.gov/40858882/)]
9. Moss AJ, Williams MC, Newby DE, Nicol ED. The updated NICE guidelines: cardiac CT as the first-line test for coronary artery disease. *Curr Cardiovasc Imaging Rep*. 2017;10(5):15. [doi: [10.1007/s12410-017-9412-6](https://doi.org/10.1007/s12410-017-9412-6)] [Medline: [28446943](https://pubmed.ncbi.nlm.nih.gov/28446943/)]
10. de Siqueira VS, Borges MM, Furtado RG, Dourado CN, da Costa RM. Artificial intelligence applied to support medical decisions for the automatic analysis of echocardiogram images: a systematic review. *Artif Intell Med*. Oct 2021;120:102165. [doi: [10.1016/j.artmed.2021.102165](https://doi.org/10.1016/j.artmed.2021.102165)] [Medline: [34629153](https://pubmed.ncbi.nlm.nih.gov/34629153/)]
11. Shivaie S, Tohidi H, Loganathan P, Kar M, Hashemy H, Shafiee MA. Interobserver variability of coronary stenosis characterized by coronary angiography: a single-center (Toronto General Hospital) retrospective chart review by staff cardiologists. *Vasc Health Risk Manag*. 2024;20:359-368. [doi: [10.2147/VHRM.S431612](https://doi.org/10.2147/VHRM.S431612)] [Medline: [39157424](https://pubmed.ncbi.nlm.nih.gov/39157424/)]
12. Rovetto C, Niemann T, Almeida G, et al. Assessment of inter- and intrareader agreement among different stenosis quantification methods in coronary CT angiography. *Eur J Radiol*. May 2025;186:112059. [doi: [10.1016/j.ejrad.2025.112059](https://doi.org/10.1016/j.ejrad.2025.112059)] [Medline: [40127592](https://pubmed.ncbi.nlm.nih.gov/40127592/)]
13. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health*. Oct 2019;1(6):e271-e297. [doi: [10.1016/S2589-7500\(19\)30123-2](https://doi.org/10.1016/S2589-7500(19)30123-2)] [Medline: [33323251](https://pubmed.ncbi.nlm.nih.gov/33323251/)]
14. Holste G, Oikonomou EK, Tokodi M, Kovács A, Wang Z, Khara R. Complete AI-enabled echocardiography interpretation with multitask deep learning. *JAMA*. Jul 22, 2025;334(4):306-318. [doi: [10.1001/jama.2025.8731](https://doi.org/10.1001/jama.2025.8731)] [Medline: [40549400](https://pubmed.ncbi.nlm.nih.gov/40549400/)]
15. Alsharqi M, Woodward WJ, Mumith JA, Markham DC, Upton R, Leeson P. Artificial intelligence and echocardiography. *Echo Res Pract*. Dec 1, 2018;5(4):R115-R125. [doi: [10.1530/ERP-18-0056](https://doi.org/10.1530/ERP-18-0056)] [Medline: [30400053](https://pubmed.ncbi.nlm.nih.gov/30400053/)]
16. Massalha S, Clarkin O, Thornhill R, Wells G, Chow BJW. Decision support tools, systems, and artificial intelligence in cardiac imaging. *Can J Cardiol*. Jul 2018;34(7):827-838. [doi: [10.1016/j.cjca.2018.04.032](https://doi.org/10.1016/j.cjca.2018.04.032)] [Medline: [29960612](https://pubmed.ncbi.nlm.nih.gov/29960612/)]
17. Upton R, Mumith A, Beqiri A, et al. Automated echocardiographic detection of severe coronary artery disease using artificial intelligence. *JACC Cardiovasc Imaging*. May 2022;15(5):715-727. [doi: [10.1016/j.jcmg.2021.10.013](https://doi.org/10.1016/j.jcmg.2021.10.013)] [Medline: [34922865](https://pubmed.ncbi.nlm.nih.gov/34922865/)]
18. 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*. Mar 25, 2020;368:m689. [doi: [10.1136/bmj.m689](https://doi.org/10.1136/bmj.m689)] [Medline: [32213531](https://pubmed.ncbi.nlm.nih.gov/32213531/)]
19. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak*. Nov 30, 2020;20(1):310. [doi: [10.1186/s12911-020-01332-6](https://doi.org/10.1186/s12911-020-01332-6)] [Medline: [33256715](https://pubmed.ncbi.nlm.nih.gov/33256715/)]
20. Gaube S, Suresh H, Raue M, et al. Do as AI say: susceptibility in deployment of clinical decision-aids. *NPJ Digit Med*. Feb 19, 2021;4(1):31. [doi: [10.1038/s41746-021-00385-9](https://doi.org/10.1038/s41746-021-00385-9)] [Medline: [33608629](https://pubmed.ncbi.nlm.nih.gov/33608629/)]
21. Sivaraman V, Bukowski LA, Levin J, Kahn JM, Perer A. Ignore, trust, or negotiate: understanding clinician acceptance of AI-based treatment recommendations in health care. Presented at: CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems; Hamburg Germany. 2023.[doi: [10.1145/3544548.3581075](https://doi.org/10.1145/3544548.3581075)]
22. Snow T. From satisficing to artificing: the evolution of administrative decision-making in the age of the algorithm. *Data & Policy*. 2021;3:e3. [doi: [10.1017/dap.2020.25](https://doi.org/10.1017/dap.2020.25)]
23. Joseph J, Kartheeban K. Cognitive alignment in cardiovascular AI: designing predictive models that think with, not just for, clinicians. *Front Cardiovasc Med*. 2025;12:1651324. [doi: [10.3389/fcvm.2025.1651324](https://doi.org/10.3389/fcvm.2025.1651324)] [Medline: [40964520](https://pubmed.ncbi.nlm.nih.gov/40964520/)]
24. Hassan N, Slight R, Bimpong K, et al. Systematic review to understand users perspectives on AI-enabled decision aids to inform shared decision making. *NPJ Digit Med*. Nov 21, 2024;7(1):332. [doi: [10.1038/s41746-024-01326-y](https://doi.org/10.1038/s41746-024-01326-y)] [Medline: [39572838](https://pubmed.ncbi.nlm.nih.gov/39572838/)]
25. Mahdavi M, Thomas N, Flood C, et al. Evaluating artificial intelligence-driven stress echocardiography analysis system (EASE study): a mixed method study. *BMJ Open*. Oct 2, 2024;14(10):e079617. [doi: [10.1136/bmjopen-2023-079617](https://doi.org/10.1136/bmjopen-2023-079617)] [Medline: [39357985](https://pubmed.ncbi.nlm.nih.gov/39357985/)]

26. Woodward G, Bajre M, Bhattacharyya S, et al. PROTEUS study: a prospective randomized controlled trial evaluating the use of artificial intelligence in stress echocardiography. *Am Heart J*. Sep 2023;263:123-132. [doi: [10.1016/j.ahj.2023.05.003](https://doi.org/10.1016/j.ahj.2023.05.003)] [Medline: [37192698](https://pubmed.ncbi.nlm.nih.gov/37192698/)]
27. Kottner J, Audigé L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol*. Jan 2011;64(1):96-106. [doi: [10.1016/j.jclinepi.2010.03.002](https://doi.org/10.1016/j.jclinepi.2010.03.002)] [Medline: [21130355](https://pubmed.ncbi.nlm.nih.gov/21130355/)]
28. 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. *Ann Intern Med*. Oct 16, 2007;147(8):573-577. [doi: [10.7326/0003-4819-147-8-200710160-00010](https://doi.org/10.7326/0003-4819-147-8-200710160-00010)] [Medline: [17938396](https://pubmed.ncbi.nlm.nih.gov/17938396/)]
29. Choi WJ, Lee HS, Hong JH, Chang HJ, Lee JW. Comparison of the JNC7 and 2017 American College of Cardiology/American Heart Association guidelines for the management of hypertension in Koreans: analysis of two independent nationwide population-based studies. *Int J Environ Res Public Health*. Dec 16, 2019;16(24):1-13. [doi: [10.3390/ijerph16245134](https://doi.org/10.3390/ijerph16245134)] [Medline: [31888171](https://pubmed.ncbi.nlm.nih.gov/31888171/)]
30. ElSayed NA, Aleppo G, Aroda VR, et al. 2. Classification and diagnosis of diabetes: standards of care in diabetes-2023. *Diabetes Care*. Jan 1, 2023;46(Suppl 1):S19-S40. [doi: [10.2337/dc23-S002](https://doi.org/10.2337/dc23-S002)] [Medline: [36507649](https://pubmed.ncbi.nlm.nih.gov/36507649/)]
31. Demography question development for Census 2021. Office for National Statistics. 2024. URL: <https://www.ons.gov.uk/census/censustransformationprogramme/questiondevelopment/demographyquestiondevelopmentforcensus2021> [Accessed 2025-09-04]
32. Braun V, Clarke V, Newson L. Thematic analysis: a practical guide to understanding and doing. *QMIP Bulletin*. 2022;1(33):46-50. [doi: [10.53841/bpsqmip.2022.1.33.46](https://doi.org/10.53841/bpsqmip.2022.1.33.46)]
33. Migliaro S, Celotto R, Teliti R, Mariani S, Altamura L, Tomai F. Comparing AI-driven and heart team decision-making in multivessel coronary artery disease. *J Clin Med*. Jun 23, 2025;14(13):4452. [doi: [10.3390/jcm14134452](https://doi.org/10.3390/jcm14134452)] [Medline: [40648826](https://pubmed.ncbi.nlm.nih.gov/40648826/)]
34. Visseren FLJ, Mach F, Smulders YM, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. *Eur Heart J*. Sep 7, 2021;42(34):3227-3337. [doi: [10.1093/eurheartj/ehab484](https://doi.org/10.1093/eurheartj/ehab484)] [Medline: [34458905](https://pubmed.ncbi.nlm.nih.gov/34458905/)]
35. Zeltzer D, Herzog L, Pickman Y, et al. Diagnostic accuracy of artificial intelligence in virtual primary care. *Mayo Clin Proc Digit Health*. Dec 2023;1(4):480-489. [doi: [10.1016/j.mcpdig.2023.08.002](https://doi.org/10.1016/j.mcpdig.2023.08.002)] [Medline: [40206313](https://pubmed.ncbi.nlm.nih.gov/40206313/)]
36. Sonaglioni A, Nicolosi GL, Rigamonti E, Lombardo M, Gensini GF, Ambrosio G. Does chest shape influence exercise stress echocardiographic results in patients with suspected coronary artery disease? *Intern Emerg Med*. Jan 2022;17(1):101-112. [doi: [10.1007/s11739-021-02773-1](https://doi.org/10.1007/s11739-021-02773-1)] [Medline: [34052977](https://pubmed.ncbi.nlm.nih.gov/34052977/)]
37. Harada T, Shimizu T, Kaji Y, et al. A perspective from a case conference on comparing the diagnostic process: human diagnostic thinking vs. artificial intelligence (AI) decision support tools. *Int J Environ Res Public Health*. Aug 22, 2020;17(17):1-6. [doi: [10.3390/ijerph17176110](https://doi.org/10.3390/ijerph17176110)] [Medline: [32842581](https://pubmed.ncbi.nlm.nih.gov/32842581/)]
38. Johnson CL, Woodward W, McCourt A, et al. Real world hospital costs following stress echocardiography in the UK: a costing study from the EVAREST/BSE-NSTEP multi-entre study. *Echo Res Pract*. May 31, 2023;10(1):8. [doi: [10.1186/s44156-023-00020-1](https://doi.org/10.1186/s44156-023-00020-1)] [Medline: [37254216](https://pubmed.ncbi.nlm.nih.gov/37254216/)]
39. Vervoort D, Wang R, Li G, et al. Addressing the global burden of cardiovascular disease in women: JACC state-of-the-art review. *J Am Coll Cardiol*. Jun 25, 2024;83(25):2690-2707. [doi: [10.1016/j.jacc.2024.04.028](https://doi.org/10.1016/j.jacc.2024.04.028)] [Medline: [38897679](https://pubmed.ncbi.nlm.nih.gov/38897679/)]
40. Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: an assailable barrier to equitable digital health care. *Lancet Digit Health*. Apr 2021;3(4):e260-e265. [doi: [10.1016/S2589-7500\(20\)30317-4](https://doi.org/10.1016/S2589-7500(20)30317-4)] [Medline: [33678589](https://pubmed.ncbi.nlm.nih.gov/33678589/)]
41. Larrazabal AJ, Nieto N, Peterson V, Milone DH, Ferrante E. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proc Natl Acad Sci U S A*. Jun 9, 2020;117(23):12592-12594. [doi: [10.1073/pnas.1919012117](https://doi.org/10.1073/pnas.1919012117)] [Medline: [32457147](https://pubmed.ncbi.nlm.nih.gov/32457147/)]
42. Selten F, Robeer M, Grimmelikhuijsen S. 'Just like I thought': street-level bureaucrats trust ai recommendations if they confirm their professional judgment. *Public Adm Rev*. Mar 2023;83(2):263-278. [doi: [10.1111/puar.13602](https://doi.org/10.1111/puar.13602)]
43. Knackstedt C, Bekkers SCAM, Schummers G, et al. Fully automated versus standard tracking of left ventricular ejection fraction and longitudinal strain: the FAST-EFs multicenter study. *J Am Coll Cardiol*. Sep 29, 2015;66(13):1456-1466. [doi: [10.1016/j.jacc.2015.07.052](https://doi.org/10.1016/j.jacc.2015.07.052)] [Medline: [26403342](https://pubmed.ncbi.nlm.nih.gov/26403342/)]
44. Tajik AJ. Machine learning for echocardiographic imaging: embarking on another incredible journey. *J Am Coll Cardiol*. Nov 29, 2016;68(21):2296-2298. [doi: [10.1016/j.jacc.2016.09.915](https://doi.org/10.1016/j.jacc.2016.09.915)] [Medline: [27884248](https://pubmed.ncbi.nlm.nih.gov/27884248/)]

Abbreviations

- AI:** artificial intelligence
- CAD:** coronary artery disease
- CRF:** cardiovascular risk factor
- EGP:** EchoGo Pro

SE: stress echocardiography

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by Alicia Stone; peer-reviewed by Andrea Sonaglioni, John Michael Hoppe, Samuel Krasner; submitted 04.Sep.2025; final revised version received 13.Mar.2026; accepted 31.Mar.2026; published 20.May.2026

Please cite as:

Mahdavi M, White S, Hothi SS, Flood C, Panayotova R, Frings D

Alignment Between Cardiologists and AI-Driven Diagnostic Systems: Mixed Methods Study

J Med Internet Res 2026;28:e83541

URL: <https://www.jmir.org/2026/1/e83541>

doi: [10.2196/83541](https://doi.org/10.2196/83541)

© Mahdi Mahdavi, Sarah White, Sandeep S Hothi, Chris Flood, Rosica Panayotova, Daniel Frings. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 20.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.