OPEN ACCESS

Clinician interaction with a machine learning algorithm for the assessment of patients with possible acute heart failure: a qualitative study

Amy V Ferry [1], Kuan Ken Lee,[1] Dimitrios Doudesis [1], Chris Tuck [1], Alasdair J Gray,[2] Nicholas L Mills [1,2]

Handling editor Alex Novak

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/emermed-2025-215220).

[1]BHF Centre for Cardiovascular Science, The University of Edinburgh, Edinburgh, UK
[2]Usher Institute of Population Health Science and Informatics, The University of Edinburgh, Edinburgh, UK

Correspondence to
Dr Amy V Ferry;
amy.ferry@ed.ac.uk

Received 21 May 2025
Accepted 27 March 2026

Check for updates

To cite: Ferry AV, Lee KK, Doudesis D, *et al*. *Emerg Med J* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/emermed-2025-215220

ABSTRACT

Background Machine learning (ML) could improve clinical decisions in patients with possible acute heart failure, but few studies have evaluated acceptance, and barriers or facilitators that lead to clinician engagement with these tools. In a qualitative study, we used anonymised clinical cases of breathless patients to explore barriers and facilitators to engagement with a clinical decision-support tool—'CoDE-HF'—that applies ML to estimate the probability of acute heart failure from natriuretic peptide concentrations and clinical variables.

Methods Emergency department clinicians across three acute care hospitals were invited to participate in 1:1 semi-structured interviews either face-to-face or by video call. Clinicians were asked to review five anonymised clinical cases and 'think aloud' about patient assessment strategies and interpretation of the Collaboration for the Diagnosis and Evaluation of Heart Failure (CoDE-HF) model outputs. Interviews were recorded, transcribed and coded. Codes were mapped onto the four domains of the unified theory of acceptance and use of technology model (performance expectancy, effort expectancy, social influences, facilitating conditions) which was used to identify barriers and facilitators to acceptance.

Results Facilitators to use were CoDE-HF's ability to promote objective communication between colleagues and its role in reprioritising acute heart failure in cases where a diagnosis may have been missed. The method of presentation of model output (statements relating to the positive or negative predictive value of the CoDE-HF output score and visual traffic light system for the low-probability, intermediate-probability or high-probability categories) was viewed as facilitators, though the absolute numerical score was more difficult to interpret. Access to a computer and clinical sample processing time were the only potential organisational issues identified as barriers.

Conclusion Clinicians reported that CoDE-HF could be a useful adjunct to clinical assessment of patients with breathlessness in the emergency department. Ease of model output interpretation is key to acceptance with interviews identifying a need to refine presentation of score information.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Machine learning (ML) has the potential to revolutionise healthcare by enhanced diagnostics, streamlining workflows and targeting and personalising treatments. Despite this, adoption of some ML solutions into clinical practice is slow to materialise.

WHAT THIS STUDY ADDS

⇒ We analysed how clinicians used information from an ML-facilitated clinical decision support tool to assess patients presenting with breathlessness to the emergency department to aid implementation of the system into clinical practice.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ We describe how ML can be translated into frontline clinical decision-making.

INTRODUCTION

Background

Diagnosis of acute heart failure is challenging because other life-threatening conditions produce similar symptoms and clinical signs. Furthermore, patients with heart failure frequently have multiple other comorbidities, both cardiac (eg, ischaemic heart disease, atrial fibrillation) and non-cardiac (eg, chronic obstructive pulmonary disease (COPD), renal failure) which may complicate the clinical presentation. The initial assessment of these patients in the emergency department is vital as it influences clinical decisions, such as admission to hospital, referral to specialist services for investigations and the initiation of evidence-based therapies.

Importance

National and international guidelines recommend the use of natriuretic peptide testing to aid in the diagnosis of acute heart failure.[1 2] N-terminal pro-B-type natriuretic peptide (NT-proBNP) and B-type natriuretic peptide (BNP) levels are known to be influenced by age, obesity and renal failure. At the very least, adjustments must be made for age and sex for interpretation.[3] Therefore, solutions to improve the interpretation of natriuretic peptides and the diagnosis of patients with acute heart failure are needed to improve patient outcomes and utilisation of healthcare resources.

Our group developed Collaboration for the Diagnosis and Evaluation of Heart Failure (CoDE-HF),[4] a clinical decision-support tool (CDST) using machine learning (ML) and natriuretic peptide levels that aims to improve the care of patients with

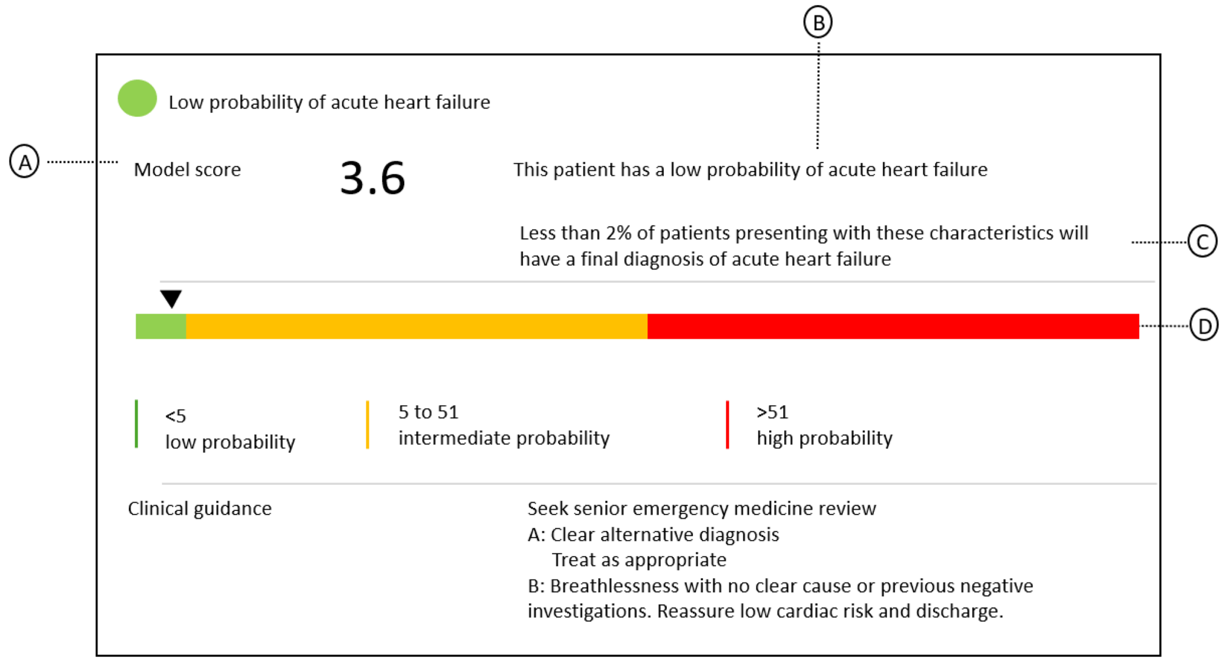

Figure 1 Example of the CoDE-HF algorithm output. (A) Model score represented on a scale of 0–100 corresponding to an individual patient's probability of acute heart failure. (B) Statement of either low, intermediate or high probability of acute heart failure. (C) Statement of the positive predictive value (this is given as a negative predictive value in cases of low probability. (D) Visual representation in a traffic light system. CoDE-HF, Collaboration for the Diagnosis and Evaluation of Heart Failure.

suspected acute heart failure through more accurate diagnosis. The tool was developed using data from a systematic review and patient-level meta-analysis to minimise the risk of bias. CoDE-HF uses an ML algorithm to compute a score (0–100) that corresponds to an individual patient's probability of acute heart failure.[4] Patients are categorised as having low, intermediate or high probability of acute heart failure based on scores that identify the highest proportion of patients as having a low or high probability of acute heart failure with optimal performance to rule out (98% negative predictive value and 90% sensitivity) and rule in (75% positive predictive value and 90% specificity) acute heart failure, respectively. The output of the CoDE-HF model is made up of four pieces of information: (1) a statement declaring acute heart failure is effectively ruled in/out or the patient has an intermediate probability, (2) a score on a scale of 0–100 that corresponds to the probability of diagnosis, (3) a statement detailing the positive or negative predictive value corresponding to the score and (4) a horizontal bar detailing the CoDE-HF score and the boundaries between the low, intermediate and high probabilities categories represented by green, yellow and red bars, respectively. The clinician is also given clinical guidance about appropriate management based on the probability (figure 1). The weighting given to the model components is also available to view.

We previously demonstrated that CoDE-HF is able to rule in and rule out acute heart failure more accurately than approaches using natriuretic peptide thresholds alone.[4 5] CoDE-HF could facilitate patient management and outcomes through better targeting of heart failure therapies and expediting the investigation and treatment of other potential differential diagnoses in those who do not have acute heart failure.

Goals of investigation

Despite studies reporting the diagnostic accuracy of some ML enabled CDSTs to match that of expert clinicians, the uptake and sustained use of ML in healthcare is limited.[6] We therefore aimed to explore the acceptance, and barriers and facilitators that led to clinician engagement with CoDE-HF when used to assess anonymised clinical cases. The next phase of development will be to test CoDE-HF in a clinical trial.

METHODS

Study design and setting

Clinician participants were identified by advertisement through local networks. Eligible clinicians were those who assessed breathless patients in the emergency department or acute receiving unit. All clinicians coming forward agreed to participate. Recruitment continued until saturation was achieved signifying when no new findings emerge from continued data collection.[7] This was determined by debriefing with the study team (authors NLM and KKL who are cardiology clinicians and algorithm developers). Recruitment occurred in February and March 2024.

Participants viewed an online supplemental video 1 introducing the CoDE-HF algorithm, detailing how it was derived and data on performance. This can be viewed as a supplemental file (online supplemental appendix 1). Semi-structured interviews were conducted face to face or by video call (MS Teams) by an experienced cardiology research nurse (AVF) with expertise in qualitative methods. Interviews followed a topic guide (online supplemental appendix 2) piloted with colleague clinicians during the study design phase. Participants were presented with five anonymised clinical cases selected from a prospective cohort study (ClinicalTrials.gov: NCT05518396) in whom the diagnosis of heart failure was independently adjudicated according to international guidelines.[8] Cases were chosen by the study team to represent all possible outputs of the CoDE-HF algorithm. Clinical information presented included the presenting history, medical history, medication, examination findings, the ECG and chest X-ray, NT-proBNP concentrations and the CoDE-HF score. Information was revealed in a stepwise manner. A summary of the clinical characteristics of the five cases is provided in online supplemental table S1, and the outputs from the CoDE-HF algorithm for each case reviewed by clinicians is presented in figure 2. Participants were asked to 'think aloud' as they described their approach to patient assessment, interpretation of the CoDE-HF score and how this may influence their diagnosis and management. After each interview, AF recorded reflections in a research diary, and debriefing with the study team was employed to promote researcher reflexivity and serve as an audit trail for the development of the analysis, therefore increasing the trustworthiness of the study.[9 10] Topics of discussion concerned the selection of participants coming forward and how results were unfolding.

Reporting of the study follows the Consolidated Criteria for Reporting Qualitative Research (COREQ) guideline.[11]

Patient and public involvement

No patient or public involvement.

Data analysis

Interviews were recorded with consent and transcribed with data collection and analysis occurring concurrently. An interpretive approach was used to analyse data thematically.[12] Transcripts were read multiple times by AVF using the constant comparative method to develop relevant codes and analytical themes which were mapped onto the domains of the Unified Theory of Acceptance and Use of Technology (UTAUT) framework[13] (online supplemental appendix 3). This model suggests four constructs impact acceptance and user behaviour; performance expectancy (the degree to which the clinician believes CoDE-HF will help them attain gains in job performance), effort expectancy (the degree of ease associated with using the system), social influence (the degree to which an individual believes that important others think that they should use the system) and facilitating conditions (the degree to which the clinician believes the organisational and technical infrastructure exists to support the use of the system).[11] NVivo (QRS International, Victoria, Australia) was used to facilitate data coding, retrieval and analysis. Quotes were selected to support the interpretation of the data and represent the range of participant responses observed. Representation from the variety of clinician groups recruited was ensured to reduce bias in the interpretation of data. If opposing views were evident, both views were represented.

RESULTS

17 multidisciplinary participants from three acute care hospitals ranging in years of experience formed the study sample (table 1). Interviews ranged from 64 to 92 min.

Performance expectancy: the degree to which the clinician believes that the system will help them attain gains in job performance (assessment of the breathless patient)

Eight major themes were identified in this domain and evidence is given as quotes in table 2 and online supplemental table S2.

Role of machine learning in decision-making

All participants accepted ML could benefit medicine giving examples such as reducing human error and standardising assessment. A concern that newly qualified doctors may lack the experience and clinical acumen to effectively appraise the CDST output was raised.

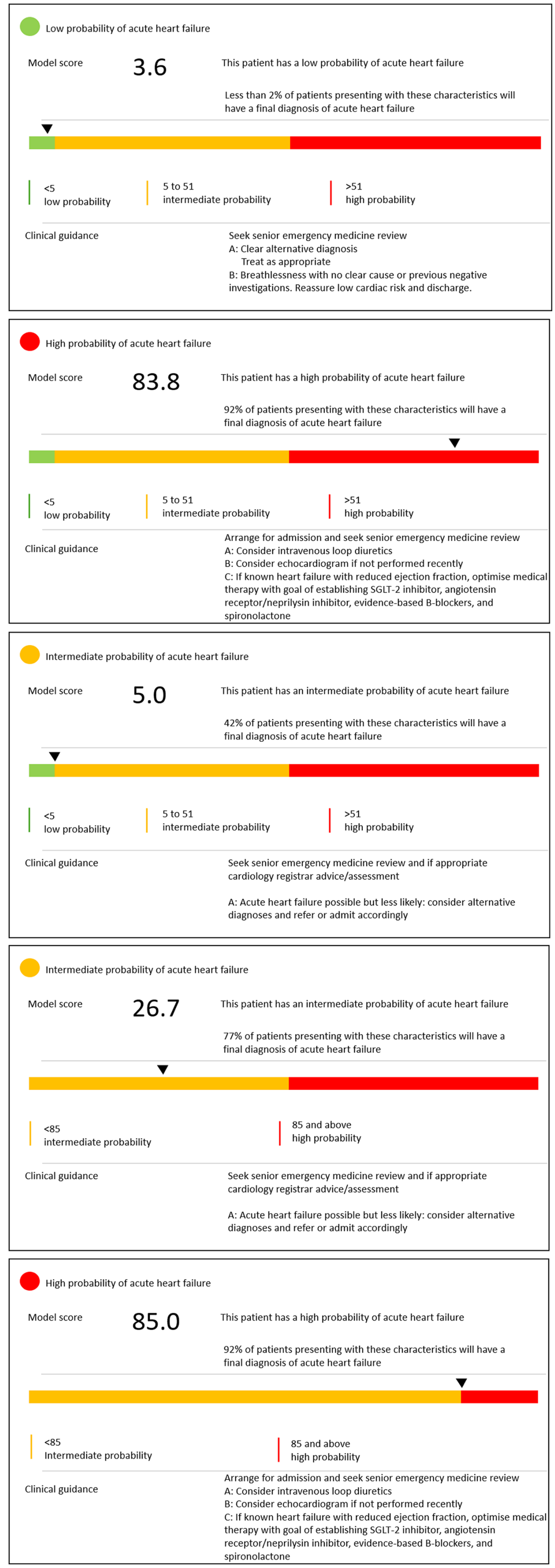

Figure 2 CoDE-HF output screens for each of the five anonymised clinical cases. CoDE-HF, Collaboration for the Diagnosis and Evaluation of Heart Failure; SGLT-2, sodium-glucose cotransporter 2.

Table 1 Interview participants

Participant	Clinical role	Years experience in current role
1	Advanced nurse practitioner	10+
2	Emergency medicine consultant	10+
3	Resident doctor (emergency medicine)	5
4	Physician associate (emergency medicine)	4
5	Resident doctor (emergency medicine)	6
6	Resident doctor (emergency medicine)	3
7	Resident doctor (emergency medicine)	5
8	Resident doctor (acute medicine)	3
9	Advanced nurse practitioner	5
10	Emergency medicine consultant	10+
11	Advanced nurse practitioner	1.5
12	Advanced nurse practitioner	2
13	Emergency medicine consultant	10+
14	Advanced nurse practitioner	7
15	Resident doctor (emergency medicine)	3
16	Resident doctor (emergency medicine)	7
17	Acute medicine consultant	8

Each domain of the UTAUT model will be presented in turn with the coding categories mapping onto that domain.
UTAUT, Unified Theory of Acceptance and Use of Technology.

Importance of clinical assessment

Clinical assessment remaining central to decision-making with CDSTs providing adjunctive information was a commonly expressed view. When the CoDE-HF output conflicted with their initial clinical impression, clinicians questioned their own judgement rather than that of the algorithm. This prompted clinical reassessment. We actively sought instances of differing interpretation or use of the algorithm according to clinical role or years of experience but did not identify such examples. For example, when it was noted that consultants were placing importance on clinical examination, accounts of less experienced clinicians were interrogated searching for examples of being led by the algorithm output, but this was not evident.

Aid to communication

CDSTs were valued for enabling objective communication (eg, Glasgow Coma Scale or Clinical Frailty Scale). Clinicians stated CoDE-HF could objectively support heart failure diagnosis discussions with senior or specialty colleagues in the same way. Some resident doctors feared intermediate scores might hinder referrals to cardiology, while low scores were seen as reassuring and could shape discussions with patients.

CoDE-HF influence over diagnosis

Clinicians were observed to reprioritise heart failure as a diagnosis due to CoDE-HF. Similarly, when acute heart failure

Table 2 Evidence for theme of performance expectancy

Performance expectancy: the degree to which the clinician believes that the system will help them attain gains in job performance (assessment of the breathless patient)	
Evidence	**Participant**
Role of machine learning in decision-making	
I'm not precious that clinicians are the only people or things that can diagnose patients or assist with that. Working in the ED is horrendously busy. You're stretched cognitively and shattered so you can use all the help you can get, and the health system needs all the help you can get so you can't be overly precious. The ideal situation of having a clinician that is well rested and has plenty of time to assess a patient and has a wealth of experience is the gold standard but that's not what we're working with.	Resident doctor, participant 16
I think you need to be a bit careful with junior colleagues because I suppose, a bit like social media, I think we're of a generation where we didn't grow up with this and we can let it go when needed. We can see both sides. Whereas I do wonder whether some of the younger generations don't know how to do that because they don't know when it's actually being a bit too invasive and when it isn't, and I think similarly with this, I think you need clinicians to have a good amount of that sense	Resident doctor, participant 5
Importance of clinical assessment	
Use of AI (artificial intelligence) is inevitable and it might be useful. I feel quite strongly that you can't really replace speaking to a human and examining them and using what you've been taught in medical school.	Emergency medicine consultant, participant 14
And it doesn't matter what is on these calculators. I'm looking at what's in front of me first and foremost. And it certainly doesn't replace clinical judgement, but it certainly helps give you other ideas to think about, and it certainly helps in where to go next.	Advanced nurse practitioner, participant 12
Aid to communication	
I think we often have a fear of referring people to cardiology because they obviously have to be quite unwell, there's a lot of gatekeeping going on as to who gets in and I think it (CoDE-HF) gives us a bit more of a backing as to who actually needs to go and see them.	Resident doctor, participant 15
I think the 3.6 comes across as a bit arbitrary but the less than two per cent of patients, I think that to me is quite meaningful. That's something where if a patient was saying to me, I'm just worried that this is heart failure, my dad had it, or something, I could say to them, well, look, we've looked at this algorithm or tool and we've used these various things, clinically, I don't think it is. But also, we've got this tool and it says that less two per cent patients with these symptoms have this diagnosis.	Resident doctor, participant 5
CoDE-HF influence over diagnosis	
And knowing that percentage is 77, it's quite common as well, so people with this kind of presentation, you will see it in, like, three-quarters of the patients there. So, I think that's quite helpful. I think we would still have treated it as acute heart failure, but I think the seriousness of it was more…is more solidified with the score.	Resident doctor, participant 8
I just think that 77 per cent, I would probably want that followed up, you know? I do think that has changed what I'm thinking about this.	ANP, participant 9
Over-reliance on the tool	
So I guess junior doctors and the like are also quite susceptible to liking a pathway, liking a protocol, because it makes our lives easier but the other side of that is not becoming too protocol driven and not creating more work for ourselves by missing something.	Resident doctor, participant 7
Emergency department assessment practices	
How closely linked diagnosis and management is particularly in the ED is questionable. The diagnosis comes a bit further down the road when you see response to different treatments.	Resident doctor, participant 16
I think nine times out of ten I decide what's wrong with the patient before I've done any tests. Like, listen to their chest and examine them and take a good history. 'Cause you treat patients…well you normally treat patients before their bloods come back.	Resident doctor, participant 6
Heart failure red flags	
Peripheral oedema is just yes or no. It doesn't take into account it's up to the thighs and not ankle oedema.	Resident doctor, participant 16
Benefit to patient flow	
That's extremely helpful in saying really heart failure's quite low here and there's less pressure to admit for inpatient echo.	Acute medicine consultant, participant 17

ANP, advanced nurse practitioner; CoDE-HF, Collaboration for the Diagnosis and Evaluation of Heart Failure; ED, emergency department.

suspicion was low, CoDE-HF aided 'rule-out', guiding focus to other diagnoses and treatments. The statements of 'acute heart failure is effectively ruled out' and 'less than 2% of patients will have a final diagnosis of heart failure' provided reassurance that their clinical suspicion of an alternative diagnosis of COPD was correct.

Over-reliance on the tool

Some clinicians thought CoDE-HF might result in a focus on heart failure over other acute conditions linked to or contributing to heart failure. However, consultant interviews indicated confidence that routine senior review would prevent this issue.

Emergency department assessment practices

Emergency department clinicians described discipline-specific assessment practices. 'Treatment trials' were common practice, focusing on identifying life-threatening conditions and managing symptoms, with diagnosis often occurring later. Reliance on NT-proBNP for calculation of CoDE-HF meant treatment trials would already have started, before the results of laboratory tests were available. The appropriateness of starting goal-directed medical therapy, such as sodium-glucose cotransporter 2 inhibitors (SGLT-2 inhibitors), in the emergency department setting was also questioned.

Heart failure 'red flags'

Interviews identified key red flags which raise suspicion of heart failure. These were orthopnoea, peripheral oedema, history of heart failure or ischaemic heart disease, and timing of symptom onset. CoDE-HF does not use some of these features though this did not appear to affect confidence. Interviews revealed a history of heart failure and ischaemic heart disease raised suspicion of acute heart failure, while acute symptom onset reduced it. This information was expected to be gathered during examination and would form part of the subjective pretest probability against which the model output would be appraised.

Benefit to patient flow

Clinicians reported CoDE-HF could justify a referral for echocardiography, assist with patient flow by reducing unnecessary admissions for inpatient echocardiography if heart failure is unlikely, and support discharge decisions due to the rule out of acute heart failure.

Effort expectancy: the degree of ease associated with using the system

Two major themes were identified in this domain. Evidence is given as quotes in table 3.

Trust in the system

When directly questioned about trust associated with an ML algorithm and the requirement to understand the derivation of the model to employ it in clinical assessment, clinicians did not report any issues with trusting a new system providing the tool had been validated. Provision of information relating to the weighting of variables within the model was supplied which was viewed as key information when interpreting the model output. One clinician expressed reservations with artificial intelligence (AI), citing automated ECG interpretations as often unhelpful and obliging the clinician to disprove the interpretation. Participants did not appear to view an ML-derived algorithm as different to any other CDST they had worked with previously. The consensus was that CoDE-HF should partner and not replace clinical judgement.

Integration of score into decision-making

Clinicians found the traffic light system useful, associating green with safety and confidence in ruling out acute heart failure. The statement derived from the positive predictive value was described as 'tangible information'. Only one interviewee appeared to focus on the actual score and concentrated on which category it fell into. Most other clinicians stated the CoDE-HF number did not hold significant meaning to them at this time but acknowledged it may gain relevance with continued use. The CoDE-HF score caused confusion when clinicians aimed to correlate the number to the positive predictive value statement (they are related, but not equivalent). If the model output fitted with their clinical suspicion, clinicians interpreted the score more favourably. Study participants felt that the clinical guidance offered for patients with an intermediate probability needed refining. Some clinicians felt the positive predictive values of 42% and 77% for the two intermediate cases did not correlate with the statement 'acute heart failure is possible but less likely, consider alternative diagnoses and refer or admit accordingly'.

Social influence: the degree to which an individual believes that important others think that they should use the system

Local, national and international guidelines had the strongest influence on CDST use. Other sources were research active colleagues and discussion on respected podcasts (evidence given as quotes in table 4).

Facilitating conditions: the degree to which the clinician believes the organisational and technical infrastructure exists to support the use of the system

Computer access and laboratory processing time were identified as organisational and technical barriers. Clinicians suggested using a mobile app and point of care NT-proBNP assay as possible solutions (evidence given as quotes in table 4).

DISCUSSION

AI and ML have the potential to transform healthcare,[14] if successfully integrated into clinical practice.[6] Reporting often neglects human interaction with the product in favour of algorithm development.[15] This study aimed to explore the acceptance, and barriers and facilitators that led to clinician engagement with CoDE-HF.

We report five key findings. First, clinicians acknowledged that harnessing ML algorithms could help reduce cognitive load and agreed that CoDE-HF could provide a useful adjunct to decision-making. All clinicians in the study stated they were willing to engage with AI in healthcare, which is a view supported in the literature,[14 16] yet the clinical impact of these systems is slow to materialise.[17–20] Second, lack of trust was not widely raised by participants relating to CoDE-HF even on direct questioning. The issue of trust in AI in healthcare has been widely reported as being a barrier to adoption, with factors of explainability, transparency, interpretability, usability and education identified as factors influencing the development of trust.[6 20–23] We observed clinicians deferring trust to the process of validation and the large datasets used to derive models rather than adopting independent responsibility for an assessment of trust. This is a key finding as use of an algorithm should come with an understanding of the patient population it is derived from and consideration of any bias that may be introduced.

Table 3 Evidence for theme of effort expectancy

Effort expectancy: the degree of ease associated with using the system	
Evidence	**Participant**
Trust of the system	
I think that…like blood pressure is one of these things where clinically there's a million and one things that an impact on your blood pressure so how much in your own head you put on that is often difficult. And that's where some scores can come in handy 'cause it obviously has got enough patients to be able to actually bring out the clinical significance of a drop in blood pressure. Yeah, no, so that is good.	Resident doctor, participant 16
I think it's good to know what factors are considered so you can see…you can trust that it's considering all the different factors.	Advanced nurse practitioner, participant 19
At the front door you're probably not interested [in the derivation of the tool], you just want to know if the score is validated.	Emergency medicine consultant, participant 2
I wouldn't know how the tool worked that out, and that's probably quite reflective of how bad I am at this. But, no, yeah, I think…this is presented in the same sort of way as MDCalc would be, and that's familiar and I'm trusting the tool at this stage.	Emergency medicine consultant, participant 13
I suppose the interesting question is, the score is putting weighting on certain elements of the patient's presentation history and examination, and so am I. But probably what I'd put weighting on is not purely scientific and based on numbers, it's based on, I don't know, what I did last time	Resident doctor, Participant 5
I've always just put my trust in the validation processes and the weighting I assume just comes from huge datasets, but it's definitely something I think is interesting. Yes, I've never seen how things are weighted before, I always just assumed they are.	Resident doctor, Participant 7
Integration of score into decision-making	
It's additional information…you know, it's not taking away from your skills and your clinical decision-making, it should be supporting and helping rather than taking away from your clinical judgement.	Advanced nurse practitioner, participant 9
And I guess we've been using this for years in terms of, it reminds me of like the scoring systems, like the PERC criteria, Well's score, it's just a bit like that, it just gives you a bit more of a guide and something to guide what you're thinking rather than strict yes or no, this is what we do.	Resident doctor, participant 15
I quite like the green, yellow and red bar as, like, a visual guide. And then…but I think the score on its own it would be a bit difficult to interpret, so I think it's quite good having it in combination because it…yeah, like, with any of these clinical decisions it gives you a, kind of…an answer but then you always want to know what that actually then means. Less than two per cent of the patients will have a diagnosis of acute heart failure - that's useful in combination with the bar.	Resident doctor, participant 6
I find the discrepancy between the model score and the predicted outcome score, that discordance is really difficult to get your head around, especially in the acute department, a lot of people are going to see five and think…especially with the visual depiction there with the arrow at the five per cent mark, but actually just under half the patients will have a firm diagnosis of acute heart failure, I think that's quite misleading.	Resident doctor, participant 7
So, I would be quite happy to agree with this. She is not a typical heart failure picture. Forty-two per cent. It's just slightly less than half. She isn't a typical presentation. I think I would want to go down the list of looking for other things. I'd probably want to explore other differentials that would have been on my list there. I wouldn't… I have to say I would probably knock heart failure quite the way far down at that point. I would still keep it, just given the shape of her heart in her chest and her breathlessness lying flat, but I would want to explore other avenues at the moment.	Advanced nurse practitioner, participant 12
Possible but less likely…but you said that 77 per cent of these people with the characteristics will have a final diagnosis of heart failure…I don't understand…I don't totally understand that. The percentages you can understand. Consider alternative diagnosis. Yeah, we've realised…I think we've come to the point that we know…it's not 100 per cent sure anyway, but I think it's a significant issue given his other things…that he's not going home. You know.	Emergency medicine consultant, participant 2
It doesn't pass the 5 am test. You know, the junior, the ST3 that's on in a reg job overnight, at five o'clock…you know, to me that's the test these decision-making tools have to pass. It's not us that have been kicking around for decades. So I think that that is quite confusing, 'cause you're basically being presented with A, this patient probably does not have cardiac failure, but then it also then says, but there's also a 42 per cent chance that she does. Well, what does that mean?	Emergency medicine consultant, participant 10

PERC, Pulmonary embolism rule-out criteria; ST3, Specialty Trainee Year 3.

The CoDE-HF interface was designed with intended clinical users focusing on interpretability, and before the interviews, participants had watched a video describing external derivation and validation of the model, therefore satisfying many of these criteria. This may have aided the development of trust by avoiding the 'black box' approach which is unacceptable to clinicians.[17] Provision of this information can be viewed as a facilitator to engagement with CoDE-HF.

CoDE-HF is intended to support diagnosis aiming to accommodate clinician autonomy and facilitate workflow rather than disrupt it, which are also features known to increase acceptability of AI-facilitated assessment.[17 21] Again, the attention the interface development team paid to design has served to facilitate engagement with the algorithm. When CoDE-HF was disruptive, it was in a positive situation due to a discrepancy between the CDST and clinician gestalt. This situation prompted

Table 4 Evidence for themes of social influence and facilitating conditions

Social influence	
I think guidelines, first and foremost, like, when a guideline's updated or changed, or, if you listen to a podcast and they recommend a certain tool, that's usually where I would go first-line. I always listen to RCEM (Royal College of Emergency Medicine podcast).	Resident doctor, Participant 3
Facilitating conditions	
Trying to get a computer sometimes is quite difficult, so being able to whip out my phone and be like, there's MDCalc, it's on my app – great.	Advancer nurse practitioner, participant 12
I guess haemoglobin you can get from a gas, but the NT-proBNP and the eGFR take an hour or so to come back, by which point, if someone's really unwell with heart failure, we've treated it already.	Resident doctor, Participant 6

eGFR, estimated glomerular filtration rate; NT-proBNP, N-terminal pro-B-type natriuretic peptide.

reassessment rather than discounting either the score or clinical judgement, confirming the importance of clinical assessment.

Thirdly, a main use identified for CoDE-HF was facilitating objective clinical assessment which promoted engagement. CDST-assisted assessment can standardise care between clinicians, therefore promoting equity of care.[22] This objectivity also increased confidence in communicating with colleagues, specialisms and patients. Other studies show that using objective scores for acute cardiac conditions in the emergency department enhances clinicians' ability to advocate for specialist review.[23]

Fourth, areas for improvement were identified relating to the cognitive load required to interpret CoDE-HF and the clinical guidance supplied. These findings could be a barrier to clinical adoption if not addressed, so will be refined in the next phase of development. CoDE-HF aims to provide actionable data to influence clinical decisions leading to new prescription of heart failure medications, effective use of tests and improved patient outcomes. The visual representation of low, intermediate and high probability of acute heart failure categories, the proximity to the category boundaries and the statement 'X% of patients with these characteristics will have a final diagnosis of acute heart failure' were valuable in decision making. This resulted in the reprioritisation of acute heart failure as a potential diagnosis. Clinicians noted that an individual's probability of acute heart failure based on the CoDE-HF score did not always align with the negative or positive predictive value of this score when it was applied to the population. They also noted that while the proximity of the score to the low and high probability thresholds was useful the absolute score was less so, particularly when this was not aligned with the negative or positive predictive value in the population. This could be addressed by illustrating an individual's score on a probability scale of 0–100 without reporting the absolute score or by including a better description of the diagnostic metrics to guide interpretation (eg, three out of four patients with a probability score above this value have a final diagnosis of acute heart failure). The difficulty in interpretation may suggest a lack of deep understanding about how the score is calculated. Some clinicians only required confirmation a score was validated before having confidence to use it whereas others had a deeper requirement to understand the weighting attributed to specific variables.

Lastly, the clinical guidance supplied had mixed responses. Some clinicians felt that it overstepped its remit in giving advice that is intuitive to an experienced clinician. To be accepted into clinical workflow, CDSTs must recognise the expertise of clinical users aiming to inform and assist but not replace the clinician.[17] Clinicians also commented that commencing medication such as SGLT-2 inhibitors was not the role of an emergency department clinician, a finding supported by a similar study investigating whether a CDST could help clinicians assessing patients with heart failure in this setting.[24] Other attempts at embedding CDSTs for heart failure assessment in the electronic health record have had mixed results. REVEAL-HF (Risk EValuation And its Impact on ClinicAL Decision Making and Outcomes in Heart failure) estimated the 1-year mortality for patients hospitalised with heart failure and failed to change patient outcomes or decision-making, concluding that clinicians require prescriptive support in addition to prognostic information.[25] This suggests there is a fine line between providing actionable, timely and appropriate information to influence care. BETTER CARE-HF (Building ElecTronic Tools to Enhance and Reinforce CArdiovascular REcommendations for Heart Failure) increased prescription of mineralocorticoid receptor antagonists for patients with heart failure with reduced ejection fraction using electronic health record alerts attributing the success to focusing on one class of medication and directing the alert to the appropriate clinician.[26] These studies highlight the importance of early stakeholder engagement and dedicated implementation teams to ensure that prediction tools deployed in the emergency department achieve clinical impact.[27]

We acknowledge some limitations to our study. It is possible that clinicians volunteering to participate were more likely to have positive opinions about the use of ML in medicine, though it was made clear at the beginning of each interview that we would welcome all comments and the interviewer was not part of the algorithm development team. Clinicians were all from the emergency department setting as the location of intended use of the tool, but we acknowledge the inclusion of clinicians from cardiology may have revealed additional insights. Inclusion of cases where the algorithm was engineered to give an incorrect answer may have given additional insight into how clinicians used the support tool in decision-making. Additionally, as the CoDE-HF product is still in the development stage, it was not possible to use it in the live clinical setting.

In summary, clinicians reported that CoDE-HF could be a useful tool in the assessment of patients with breathlessness in the emergency department. It could aid the assessment process by providing an objective means of communication and reprioritising heart failure as a possible diagnosis if used as an adjunct to clinical assessment. Clinicians had difficulty in interpreting the CoDE-HF metrics supplied indicating a need to refine the presentation of score information.

Contributors AVF, KKL and NLM conceived the study. AVF undertook recruitment of participants and managed the data, and analysed the data. AVF drafted the manuscript, and all authors contributed substantially to its revision. AVF is responsible for the overall content as guarantor.

Funding This study was supported by the British Heart Foundation through a Translational Award (TA/F/22/210039).

Competing interests NLM reports receiving honoraria from Abbott Diagnostics, Roche Diagnostics and Siemens Healthineers in the last 36 months outside the submitted work. KKL, DD and NLM are employed by the University of Edinburgh, who have filed a patent on the CoDE-HF algorithm (Patent Reference: PCT/GB2021/051470). All other authors have no competing interests to declare.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and favourable ethical opinion was provided by Edinburgh Medical School Research Ethics Committee and the local Caldicott Guardian (23-EMREC-028 and 23113, respectively).Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iDs
Amy V Ferry https://orcid.org/0000-0003-0500-4441
Dimitrios Doudesis https://orcid.org/0000-0001-6699-9476
Chris Tuck https://orcid.org/0000-0003-4994-6342

Nicholas L Mills https://orcid.org/0000-0003-0533-7991

REFERENCES

1 Acute Heart Failure. National institute for health and care excellence, NICE clinical guideline 187 [CG187]. 2014.
2 Ponikowski P, Voors AA, Anker SD, *et al*. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J* 2016;37:2129–200.
3 Christiansen MN, Køber L, Weeke P, *et al*. Age-Specific Trends in Incidence, Mortality, and Comorbidities of Heart Failure in Denmark, 1995 to 2012. *Circulation* 2017;135:1214–23.
4 Lee KK, Doudesis D, Anwar M, *et al*. Development and validation of a decision support tool for the diagnosis of acute heart failure: systematic review, meta-analysis, and modelling study. *BMJ* 2022;377:e068424.
5 Doudesis D, Lee KK, Anwar M, *et al*. Machine learning to optimize use of natriuretic peptides in the diagnosis of acute heart failure. *Eur Heart J Acute Cardiovasc Care* 2025;14:474–88.
6 Tonekaboni S, Joshi S, McCradden M, *et al*. What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use. *Proc Mach Learn Res* 2019;1–21.
7 Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Soc Sci Med* 2022;292:114523.
8 McDonagh TA, Metra M, Adamo M, *et al*. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J* 2021;42:3599–726.
9 Maritz J, Jooste K. Debriefing interviews and coaching conversations: Stategies to promote student reflexivity and action. *South African Journal of Higher Education* 2011;25:972–86.
10 Rolfe G. Validity, trustworthiness and rigour: quality and the idea of qualitative research. *J Adv Nurs* 2006;53:304–10.
11 Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;19:349–57.
12 Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006;3:77–101.
13 Venkatesh V, Morris MG, Davis GB, *et al*. User Acceptance of Information Technology: Toward A Unified View1. *MIS Q* 2003;27:425–78.
14 Perivolaris A, Adams-McGavin C, Madan Y, *et al*. Quality of interaction between clinicians and artificial intelligence systems. A systematic review. *Future Healthc J* 2024;11:100172.
15 Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44–56.
16 Fazakarley CA, Breen M, Leeson P, *et al*. Experiences of using artificial intelligence in healthcare: a qualitative study of UK clinician and key stakeholder perspectives. *BMJ Open* 2023;13:e076950.
17 Shortliffe EH, Sepúlveda MJ. Clinical Decision Support in the Era of Artificial Intelligence. *JAMA* 2018;320:2199–200.
18 Henry KE, Kornfield R, Sridharan A, *et al*. Human-machine teaming is key to AI adoption: clinicians' experiences with a deployed machine learning system. *NPJ Digit Med* 2022;5:97.
19 Choudhury A, Asan O. Human factors: bridging artificial intelligence and patient safety. *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care* 2020;9:211–5.
20 Liyanage H, Liaw S-T, Jonnagaddala J, *et al*. Artificial Intelligence in Primary Health Care: Perceptions, Issues, and Challenges. *Yearb Med Inform* 2019;28:41–6.
21 Shamszare H, Choudhury A. Clinicians' Perceptions of Artificial Intelligence: Focus on Workload, Risk, Trust, Clinical Decision Making, and Clinical Integration. *Healthcare (Basel)* 2023;11:2308.
22 Nair M, Andersson J, Nygren JM, *et al*. Barriers and Enablers for Implementation of an Artificial Intelligence-Based Decision Support Tool to Reduce the Risk of Readmission of Patients With Heart Failure: Stakeholder Interviews. *JMIR Form Res* 2023;7:e47335.
23 Gullick J, Wu J, Chew D, *et al*. Objective risk assessment vs standard care for acute coronary syndromes-The Australian GRACE Risk tool Implementation Study (AGRIS): a process evaluation. *BMC Health Serv Res* 2022;22:380.
24 Casey SD, Reed ME, LeMaster C, *et al*. Physicians' Perceptions of Clinical Decision Support to Treat Patients With Heart Failure in the ED. *JAMA Netw Open* 2023;6:e2344393.
25 Ahmad T, Desai NR, Yamamoto Y, *et al*. Alerting Clinicians to 1-Year Mortality Risk in Patients Hospitalized With Heart Failure: The REVEAL-HF Randomized Clinical Trial. *JAMA Cardiol* 2022;7:905–12.
26 Mukhopadhyay A, Reynolds HR, Phillips LM, *et al*. Cluster-Randomized Trial Comparing Ambulatory Decision Support Tools to Improve Heart Failure Care. *J Am Coll Cardiol* 2023;81:1303–16.
27 Chan SL, Lee JW, Ong MEH, *et al*. Implementation of Prediction Models in the Emergency Department from an Implementation Science Perspective-Determinants, Outcomes, and Real-World Impact: A Scoping Review. *Ann Emerg Med* 2023;82:22–36.