OPEN ACCESS

Check for updates

Development and validation of a decision support tool for the diagnosis of acute heart failure: systematic review, meta-analysis, and modelling study

Kuan Ken Lee,[1,*] Dimitrios Doudesis,[1,2,*] Mohamed Anwar,[1,*] Federica Astengo,[1] Camille Chenevier-Gobeaux,[3] Yann-Erick Claessens,[4] Desiree Wussler,[5,6] Nikola Kozhuharov,[5,7] Ivo Strebel,[5] Zaid Sabti,[5] Christopher deFilippi,[8] Stephen Seliger,[9] Gordon Moe,[10] Carlos Fernando,[10] Antoni Bayes-Genis,[11] Roland R J van Kimmenade,[12] Yigal Pinto,[13] Hanna K Gaggin,[14,15] Jan C Wiemer,[16] Martin Möckel,[17] Joost H W Rutten,[18] Anton H van den Meiracker,[19] Luna Gargani,[20] Nicola R Pugliese,[21] Christopher Pemberton,[22] Irwani Ibrahim,[23] Alfons Gegenhuber,[24] Thomas Mueller,[25] Michael Neumaier,[26] Michael Behnes,[27] Ibrahim Akin,[27] Michele Bombelli,[28] Guido Grassi,[29] Peiman Nazerian,[30] Giovanni Albano,[30] Philipp Bahrmann,[31] David E Newby,[1] Alan G Japp,[1] Athanasios Tsanas,[2] Anoop S V Shah,[1,32] A Mark Richards,[22,33] John J V McMurray,[34] Christian Mueller,[5] James L Januzzi,[14,15] Nicholas L Mills,[1,2] on behalf of the CoDE-HF investigators

For numbered affiliations see end of the article

*Contributed equally

Correspondence to: N L Mills
nick.mills@ed.ac.uk
(or @HighSTEACS on Twitter;
ORCID 0000-0003-0533-7991)
Additional material is published online only. To view please visit the journal online.

ABSTRACT

OBJECTIVES
To evaluate the diagnostic performance of N-terminal pro-B-type natriuretic peptide (NT-proBNP) thresholds for acute heart failure and to develop and validate a decision support tool that combines NT-proBNP concentrations with clinical characteristics.

DESIGN
Individual patient level data meta-analysis and modelling study.

SETTING
Fourteen studies from 13 countries, including randomised controlled trials and prospective observational studies.

PARTICIPANTS
Individual patient level data for 10 369 patients with suspected acute heart failure were pooled for the meta-analysis to evaluate NT-proBNP thresholds. A decision support tool (Collaboration for the Diagnosis and Evaluation of Heart Failure (CoDE-HF)) that combines NT-proBNP with clinical variables to report the probability of acute heart failure for an individual patient was developed and validated.

MAIN OUTCOME MEASURE
Adjudicated diagnosis of acute heart failure.

RESULTS
Overall, 43.9% (4549/10 369) of patients had an adjudicated diagnosis of acute heart failure (73.3% (2286/3119) and 29.0% (1802/6208) in those with and without previous heart failure, respectively). The negative predictive value of the guideline recommended rule-out threshold of 300 pg/mL was 94.6% (95% confidence interval 91.9% to 96.4%); despite use of age specific rule-in thresholds, the positive predictive value varied at 61.0% (55.3% to 66.4%), 73.5% (62.3% to 82.3%), and 80.2% (70.9% to 87.1%), in patients aged <50 years, 50-75 years, and >75 years, respectively. Performance varied in most subgroups, particularly patients with obesity, renal impairment, or previous heart failure. CoDE-HF was well calibrated, with excellent discrimination in patients with and without previous heart failure (area under the receiver operator curve 0.846 (0.830 to 0.862) and 0.925 (0.919 to 0.932) and Brier scores of 0.130 and 0.099, respectively). In patients without previous heart failure, the diagnostic performance was consistent across all subgroups, with 40.3% (2502/6208) identified at low probability (negative predictive value of 98.6%, 97.8% to 99.1%) and 28.0% (1737/6208) at high probability (positive predictive value of 75.0%, 65.7% to 82.5%) of having acute heart failure.

CONCLUSIONS
In an international, collaborative evaluation of the diagnostic performance of NT-proBNP, guideline recommended thresholds to diagnose acute heart failure varied substantially in important patient subgroups. The CoDE-HF decision support tool incorporating NT-proBNP as a continuous measure and other clinical variables provides a more consistent, accurate, and individualised approach.

STUDY REGISTRATION
PROSPERO CRD42019159407.

WHAT IS ALREADY KNOWN ON THIS TOPIC

The diagnosis of acute heart failure can be challenging because patients often present with non-specific symptoms

Most national and international guidelines recommend N-terminal pro-B-type natriuretic peptide (NT-proBNP) testing to aid in the diagnosis of acute heart failure

NT-proBNP testing has not been universally implemented owing to concerns about diagnostic performance in clinically important patient subgroups

WHAT THIS STUDY ADDS

The guideline recommended NT-proBNP thresholds for acute heart failure had relatively poor diagnostic performance in important patient subgroups

A validated decision support tool that uses statistical modelling to combine NT-pro-BNP as a continuous measure with clinical variables has been developed

This tool ruled in and ruled out acute heart failure more accurately than did any approach using NT-proBNP thresholds alone and performed consistently across all subgroups

Introduction

Nearly 1 million people are living with heart failure in the UK, and the prevalence is projected to rise by approximately 50% over the next 25 years owing to the ageing population.[1] Decompensated acute heart failure accounts for 5% of all unplanned hospital admissions.[2] The accurate and timely diagnosis of acute heart failure can be challenging, and both national and international guidelines recommend natriuretic peptide testing to aid in the diagnosis.[3-8] Despite these recommendations, N-terminal pro-B-type natriuretic peptide (NT-proBNP) testing has not been universally implemented, in part owing to concerns about its clinical utility in a real world setting. Studies investigating the diagnostic performance of NT-proBNP have mainly been conducted in relatively small, selected patient cohorts, limiting the generalisability of study findings across clinically important subgroups, such as older patients and those with renal disease or obesity, characteristics that are becoming increasingly prevalent in patients with heart failure.[9-11] Statistical modelling approaches that incorporate patients' characteristics to provide a more individualised assessment may have more consistent diagnostic performance across patient subgroups.[12]

Although many models have been developed to predict prognosis in patients with heart failure, very few have been developed to aid in the diagnosis of acute heart failure.[13-19] Previous attempts have many strengths but have incorporated subjective variables, such as the clinician's estimation of pre-test probability or the patient's description of symptoms. Furthermore, they have incorporated NT-proBNP as a binary variable, which does not take into account the dynamic and non-linear interaction between NT-proBNP and other clinical variables. Previous attempts at developing and validating diagnostic scores have also included a limited number of patients from a single healthcare setting, which has precluded the assessment of performance within subgroups and limited external generalisability.
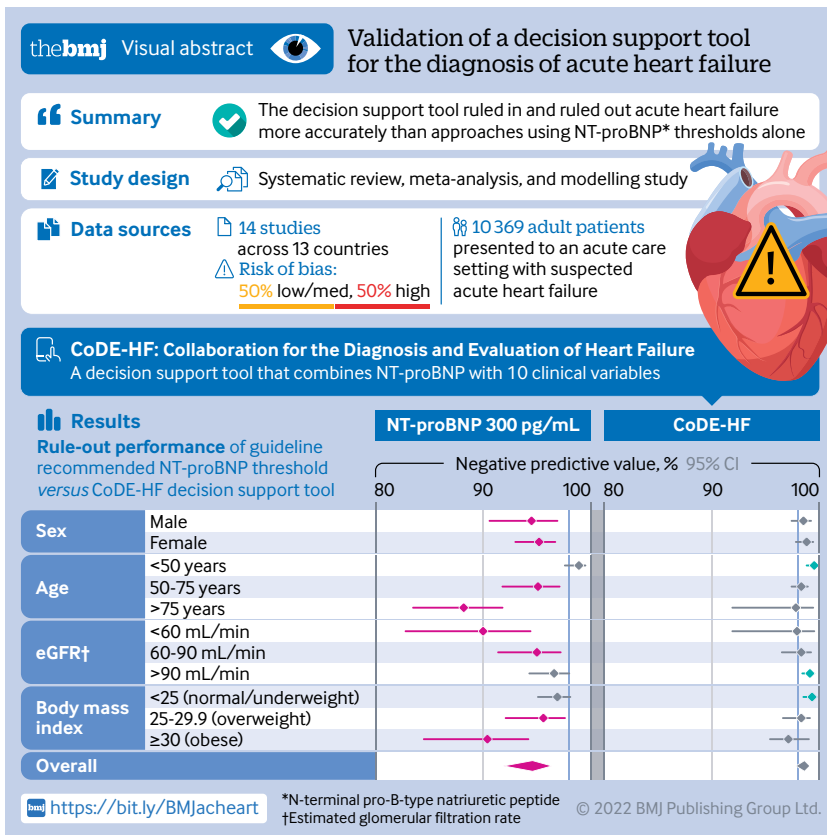
In this collaborative international analysis, we evaluated the diagnostic performance of guideline recommended NT-proBNP thresholds for acute heart failure across patient subgroups. We subsequently developed and validated a decision support tool for patients with suspected acute heart failure that uses statistical modelling to combine NT-proBNP concentrations with clinical characteristics.

Methods

Study population

We did a systematic review to identify studies that evaluated the diagnostic performance of NT-proBNP in patients with suspected acute heart failure. We updated a previous review by Roberts et al[1] with studies published up to 18 August 2021 by searching titles and abstracts on Embase, Medline, and the Cochrane Central Register of Controlled Trials with the key words "heart failure" and "natriuretic peptide" (supplementary text 1). Studies were eligible if they met the following pre-specified inclusion criteria: enrolled patients aged ≥18 years with suspected acute heart failure in an acute care setting, measured NT-proBNP in blood samples obtained during the patients' initial assessment on the day of the hospital attendance, and adjudicated diagnosis of acute heart failure by using an acceptable reference standard. Two investigators (KKL and MA) independently screened all studies identified in the systematic literature search, and a third (NLM) adjudicated conflicts by using a pre-specified protocol (PROSPERO register: CRD42019159407).

We contacted the corresponding authors of all eligible cohorts to request anonymised individual patient level data on NT-proBNP concentrations, adjudicated diagnosis of acute heart failure, demographics (age, sex, ethnicity), past medical history (heart failure, ischaemic heart disease, diabetes, hypertension, hyperlipidaemia, smoking, asthma, chronic obstructive pulmonary disease, chronic kidney disease), physiological variables during the initial assessment (heart rate and blood pressure), and clinical haematology and biochemistry profiles. We checked accuracy, definitions of variables, and completeness with all corresponding authors before harmonisation. All studies were conducted in accordance with the Declaration of Helsinki and with ethical approval to permit sharing of individual patient level data to conduct this meta-analysis. Two investigators (KKL and MA) independently assessed risk of bias for each study by using the Quality Assessment of Diagnostic Accuracy Studies version 2

the bmj Visual abstract — Validation of a decision support tool for the diagnosis of acute heart failure

Summary — The decision support tool ruled in and ruled out acute heart failure more accurately than approaches using NT-proBNP* thresholds alone

Study design — Systematic review, meta-analysis, and modelling study

Data sources — 14 studies across 13 countries. Risk of bias: 50% low/med, 50% high. 10 369 adult patients presented to an acute care setting with suspected acute heart failure

CoDE-HF: Collaboration for the Diagnosis and Evaluation of Heart Failure — A decision support tool that combines NT-proBNP with 10 clinical variables

Results — Rule-out performance of guideline recommended NT-proBNP threshold versus CoDE-HF decision support tool

		NT-proBNP 300 pg/mL	CoDE-HF
		Negative predictive value, % 95% CI	
Sex	Male		
	Female		
Age	<50 years		
	50-75 years		
	>75 years		
eGFR†	<60 mL/min		
	60-90 mL/min		
	>90 mL/min		
Body mass index	<25 (normal/underweight)		
	25-29.9 (overweight)		
	≥30 (obese)		
Overall			

https://bit.ly/BMJacheart

*N-terminal pro-B-type natriuretic peptide
†Estimated glomerular filtration rate

© 2022 BMJ Publishing Group Ltd.

(QUADAS-2) tool,[20] with conflicts resolved by a third (NLM).

NT-proBNP threshold analysis

We derived meta-estimates with 95% confidence intervals of the sensitivity, specificity, negative predictive value, and positive predictive value of the guideline recommended NT-proBNP rule-out threshold (300 pg/mL)[5 8] and age specific rule-in thresholds (450, 900, and 1800 pg/mL for patients aged <50, 50-75, and >75 years, respectively)[7] for acute heart failure by using a two stage approach, with estimates calculated separately within each study and then pooled across studies in a binomial-normal random effects model using the DerSimonian and Laird method.[21] We further evaluated the performance of these thresholds in pre-specified subgroups stratified by age, sex, ethnicity, body mass index, renal function, anaemia, and the presence of comorbidities (previous heart failure, hypertension, hyperlipidaemia, diabetes mellitus, atrial fibrillation, chronic obstructive pulmonary disease). Using the same approach, we subsequently evaluated the diagnostic performance of NT-proBNP concentrations across a range of concentrations to determine a rule-out threshold that would identify the highest proportion of patients as having a low probability of a negative predictive value ≥98% and a rule-in threshold that would identify the highest proportion of patients as having a high probability of a positive predictive value ≥75%.

Model development and validation

We developed and validated a decision support tool (Collaboration for the Diagnosis and Evaluation of Heart Failure (CoDE-HF); https://decision-support.shinyapps.io/code-hf/) by using statistical modelling to compute a value (0-100) that corresponds to an individual patient's probability of acute heart failure. Owing to significant differences in comorbidities and the prevalence of acute heart failure, we developed and validated models for patients with and without previous heart failure separately. We used NT-proBNP concentrations as a continuous measure and selected simple objective clinical variables that are known to be associated with acute heart failure, which were found to have the highest relative importance in our model training phase (age, estimated glomerular filtration rate, haemoglobin, body mass index, heart rate, blood pressure, peripheral oedema, chronic obstructive pulmonary disease, and ischaemic heart disease) (supplementary text 2).

We evaluated four different statistical models in the development of CoDE-HF: generalised linear mixed model, naïve Bayes, random forest, and extreme gradient boosting (XGBoost) (supplementary text 2).[22-24] To account for missing data across studies (supplementary figure A), we multiply imputed 10 datasets by using joint modelling multiple imputation with random study specific covariance matrices fitted with a Markov chain Monte Carlo algorithm.[25] We did multiple imputation for all variables included in the model except NT-proBNP.

We ran 10 iterations of 10-fold cross validation for each model and used the median score across the iterations and imputed datasets as the CoDE-HF score for each patient. We subsequently identified the score that would classify the highest proportion of patients as having a high or low probability of acute heart failure with optimal performance to rule in (75% positive predictive value and 90% specificity) and rule out (98% negative predictive value and 90% sensitivity) acute heart failure.

We assessed the performance of each model across a range of diagnostic metrics (area under the receiver operator curve, Brier score, proportion of patients achieving the optimal high and low probability criteria, and positive and negative predictive value across patient subgroups). The Brier score is a measure of both discrimination and calibration and is calculated by taking the mean squared difference between predicted probabilities and the observed outcome.[26] We selected the best performing model for the CoDE-HF decision support tool. We used a decision curve analysis and internal-external cross validation to evaluate the performance of CoDE-HF. In brief, this approach iteratively leaves out one study at a time for external validation and uses the remaining studies for model development.[27] We did not do imputation in the external validation dataset, so external validation was not done for studies in which most of the variables were completely missing (supplementary figure A). We used R version 4.1.2 for all analyses.

Patient and public involvement

Members of a patient and public panel were involved in the interpretation of results. There are plans to disseminate the results of the research to relevant patient communities.

Results

Study population

We contacted investigators from 30 eligible studies, of which 19 responded. Fourteen studies (12 prospective cohort studies and two randomised controlled trials) provided individual patient level data on 10 369 patients with suspected acute heart failure (mean age 69.3 years; 53.3% male) from 13 countries (table 1; supplementary figure B; supplementary tables A and B).[15 28-40] All studies were conducted in the emergency department except one that included patients admitted to the cardiology and pulmonology department (median number of patients in each study was 488 (interquartile range 322-1053)). Overall, 43.9% (4549/10 369) of patients had an adjudicated diagnosis of acute heart failure (median prevalence across studies 46% (31-54%)). Patients with a previous history of heart failure had a higher prevalence of acute heart failure than those without (73.3% (2286/3119) versus 29.0% (1802/6208)) (supplementary table C).

Guideline recommended and age specific NT-proBNP thresholds

Pooled meta-estimates of negative predictive value, sensitivity, positive predictive value, and specificity of

Table 1 | Baseline characteristics of patients stratified by diagnosis of acute heart failure. Values are numbers (percentages) unless stated otherwise

Characteristics	Overall (n=10 369)	Patients with acute heart failure (n=4549)	Patients without acute heart failure (n=5820)
Male sex	5531 (53.3)	2568 (56.5)	2963 (50.9)
Mean (SD) age, years	69.3 (16.3)	75.0 (12.6)	64.9 (17.5)
Age group, years:			
<50	1377 (13.3)	222 (4.9)	1155 (19.8)
50-75	4370 (42.1)	1674 (36.8)	2696 (46.3)
>75	4622 (44.6)	2653 (58.3)	1969 (33.8)
Ethnicity:	(n=5700)	(n=2532)	(n=3168)
Black	845 (14.8)	316 (12.5)	529 (16.7)
White	4112 (72.1)	2028 (80.1)	2084 (65.8)
Other	743 (13.0)	188 (7.4)	555 (17.5)
Medical history:			
Previous heart failure (n=9327)	3119 (33.4)	2286 (55.9)	833 (15.9)
Ischaemic heart disease (n=9136)	2953 (32.3)	1871 (46.8)	1082 (21.0)
Diabetes mellitus (n=8967)	2398 (26.7)	1382 (34.8)	1016 (20.3)
Hypertension (n=8548)	5071 (59.3)	2603 (71.0)	2468 (50.5)
Hyperlipidaemia (n=5501)	2269 (41.2)	1160 (50.8)	1109 (34.5)
Current smoker or ex-smoker (n=5946)	2458 (41.3)	918 (37.8)	1540 (43.8)
Asthma (n=4153)	770 (18.5)	98 (6.9)	672 (24.6)
Chronic obstructive pulmonary disease (n=7249)	2117 (29.2)	670 (22.9)	1447 (33.5)
Atrial fibrillation (n=3588)	1701 (20.9)	1243 (33.3)	458 (10.4)
Chronic kidney disease (n=6441)	1215 (18.9)	877 (33.6)	338 (8.8)
Mean (SD) body mass index	27.7 (7.2)	27.7 (6.8)	27.7 (7.6)
Body mass index category:	(n=7852)	(n=3528)	(n=4324)
<25	3062 (39.0)	1349 (38.2)	1713 (39.6)
25-29	2473 (31.5)	1172 (33.2)	1301 (30.1)
≥30	2317 (29.5)	1007 (28.5)	1310 (30.3)
Physiological parameters:			
Mean (SD) heart rate, bpm	91.7 (23.7)	91.5 (25.9)	91.9 (21.9)
Mean (SD) systolic blood pressure, mm Hg	140.0 (27.9)	140.2 (30.0)	139.9 (26.4)
Mean (SD) diastolic blood pressure, mm Hg	79.7 (17.0)	80.3 (18.3)	79.3 (15.9)
Clinical haematology and biochemistry:			
Mean (SD) haemoglobin, g/dL	13.1 (2.1)	12.7 (2.1)	13.4 (2.0)
Mean (SD) eGFR, mL/min/1.73 m2	68.2 (31.3)	56.8 (27.1)	77.2 (31.6)
Median (IQR) NT-proBNP, pg/mL	1182 (191-4737)	4362 (1883-9883)	279 (70-1054)

eGFR=estimated glomerular filtration rate; IQR=interquartile range; NT-proBNP=N-terminal pro-B-type natriuretic peptide; SD=standard deviation.

NT-proBNP for the overall population at the guideline recommended rule-out threshold of 300 pg/mL were 94.6% (95% confidence interval 91.9% to 96.4%), 96.8% (94.6% to 98.1%), 62.9% (51.3% to 73.3%), and 49.3% (35.4% to 63.4%), respectively (fig 1; supplementary table D). Overall, 30.4% (3148/10 369) of patients had NT-proBNP concentrations below 300 pg/mL. However, marked heterogeneity existed across patient subgroups and studies (fig 2; fig 3; supplementary figures C and D). Negative predictive value was lower in patients ≥75 years (88.2%, 83.5% to 91.8%) and in those with previous heart failure (79.4%, 68.4% to 87.3%) and obesity (90.4%, 84.5% to 94.2%).

Pooled meta-estimates of the positive predictive value for age specific NT-proBNP rule-in thresholds of 450, 900 and 1800 pg/mL were 61.0% (55.3% to 66.4%), 73.5% (62.3% to 82.3%), and 80.2% (70.9% to 87.1%), respectively (table 2). Corresponding specificities were 87.8% (79.5% to 93.0%), 81.1% (72.6% to 87.5%), and 73.1% (65.2% to 79.8%). Overall, 48.7% (5052/10 369) of patients with suspected acute heart failure had NT-proBNP above these age specific thresholds. Positive predictive values of the age specific rule-in thresholds were higher than the uniform 300 pg/mL threshold in subgroups,

although heterogeneity by age group, renal function, and prevalence of acute heart failure was present (supplementary figures E-I).

Overall, we identified seven studies as having a high risk of bias (supplementary table A). In sensitivity analyses restricted to studies in which the adjudication of acute heart failure was blinded to NT-proBNP concentrations and those with a low risk of bias, diagnostic performance of the guideline recommended and age specific NT-proBNP thresholds remained unchanged (supplementary tables E and F).

Optimised NT-proBNP thresholds
An NT-proBNP threshold of 100 pg/mL achieved our optimal rule-out criteria with a pooled negative predictive value of 97.8% (95.8% to 98.8%) and sensitivity of 99.3% (98.5% to 99.7%) (supplementary table D). However, only 17.9% (1851/10 369) of patients had NT-proBNP concentrations below 100 pg/mL and the negative predictive value remained lower in older patients and those with previous history of heart failure, ischaemic heart disease, and impaired renal function (supplementary figure J). Similarly, a NT-proBNP threshold of 1000 pg/mL achieved our optimal rule-in criteria with a positive predictive value of 74.9% (64.4% to 83.2%) and specificity of 76.1%

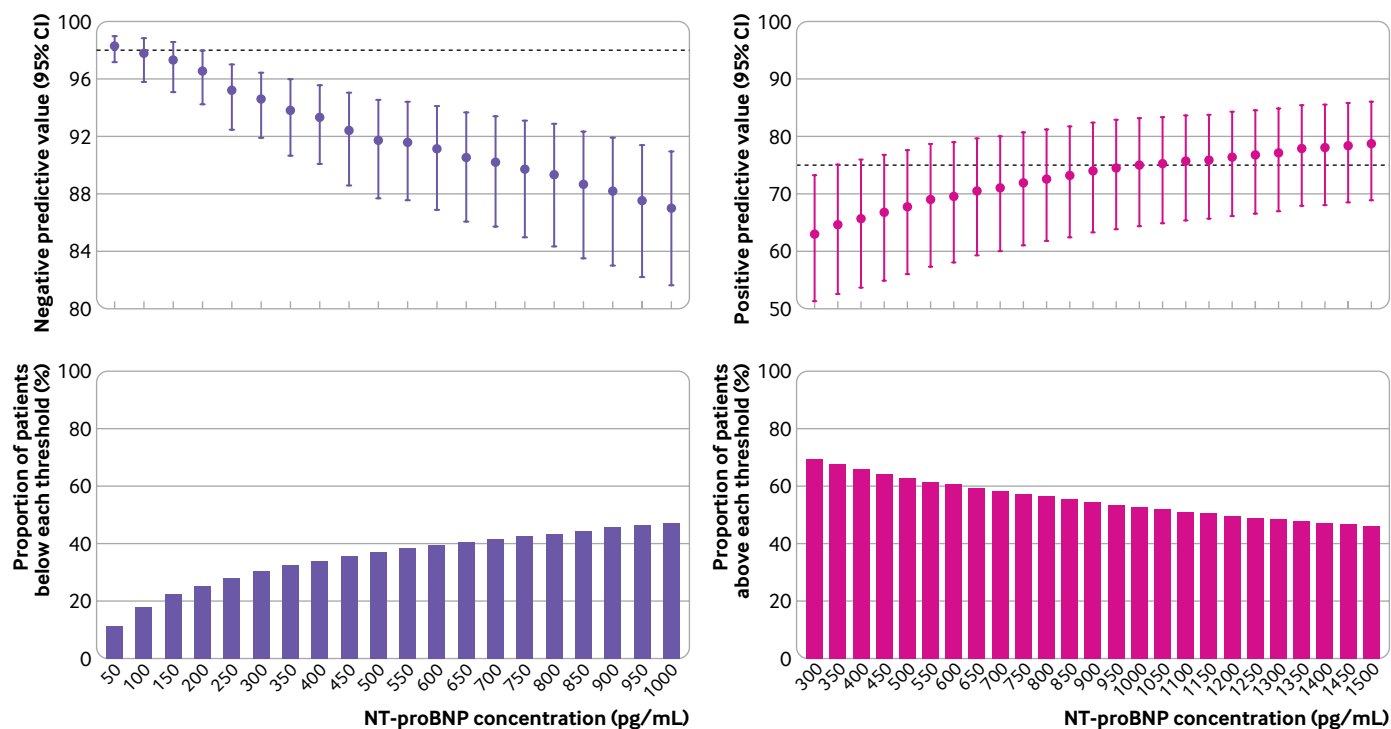doi: 10.1136/bmj-2021-068424 | BMJ 2022;377:e068424 | the bmj

Fig 1 | N-terminal pro-B-type natriuretic peptide (NT-proBNP) thresholds for acute heart failure. Top left: negative predictive values of NT-proBNP concentrations to rule out diagnosis of acute heart failure. Bottom left: cumulative proportion of patients presenting with suspected acute heart failure with NT-proBNP concentrations below each threshold. Top right: positive predictive values of NT-proBNP concentrations to rule in diagnosis of acute heart failure. Bottom right: cumulative proportion of patients presenting with suspected acute heart failure with NT-proBNP concentrations above each threshold

(65.6% to 84.2%), although performance was also lower within patient subgroups, particularly those without previous heart failure (positive predictive value 62%, 41% to 79%) (supplementary table D; supplementary figure K).

CoDE-HF score
The extreme gradient boosting (XGBoost) model and generalised linear mixed model were the best performing models (area under curve in the overall training cohort of 0.925 (95% confidence interval 0.919 to 0.932) and 0.931 (0.925 to 0.937), respectively) (supplementary text 2). Although the performance of XGBoost was similar to the generalised linear mixed model, a key advantage of XGBoost is its ability to compute a score despite missing values. This is an important functionality that we wished to build into the CoDE-HF decision support tool to facilitate its implementation in clinical practice, so we selected the XGBoost model as the final model for CoDE-HF.

CoDE-HF was well calibrated with excellent discrimination in patients with and without previous heart failure (area under the receiver operator curve 0.846 (0.830 to 0.862) and 0.925 (0.919 to 0.932) and Brier scores of 0.130 and 0.099, respectively) (fig 4; supplementary figure L). A CoDE-HF score of 4.7 achieved a negative predictive value of 98.6% (97.8% to 99.1%) and sensitivity of 98.1% (96.9% to 98.9%) (supplementary table G), and a score of 51.2 achieved a positive predictive value of 75.0% (65.7% to 82.5%)

and a specificity of 92.2% (87.5% to 95.2%) in patients without previous heart failure. These rule-in and rule-out scores had similar diagnostic performance across all subgroups (fig 5; fig 6; fig 7). If these scores were applied in patients with suspected acute heart failure, CoDE-HF would identify 40.3% (2502/6208) at low probability (<4.7) and 28.0% (1737/6208) at high probability (≥51.2) of acute heart failure. In patients with previous heart failure, no score achieved our target rule-out criteria in the training cohort. A CoDE-HF score of 84.5 achieved a positive predictive value of 92.7% (89.1% to 95.2%) and specificity of 90.2% (84.0% to 94.1%). This score would identify 45.5% (1420/3119) of patients as having a high probability of acute heart failure (fig 8). In a decision curve analysis, CoDE-HF had superior net benefit compared with the NT-proBNP alone across all threshold probabilities (supplementary figure M). The performance of CoDE-HF was marginally attenuated when trained without past medical history (area under the receiver operator curve of 0.922 (0.916 to 0.929) and 0.841 (0.825 to 0.857) in patients without and with previous heart failure). Internal-external cross validation showed good performance across cohorts for both models (supplementary figure N).

Patients who were identified as low probability by CoDE-HF had a substantially lower rate of all cause and cardiovascular mortality at 30 days and one year compared with those who were identified as intermediate and high probability (30 day all cause mortality: 1.0%

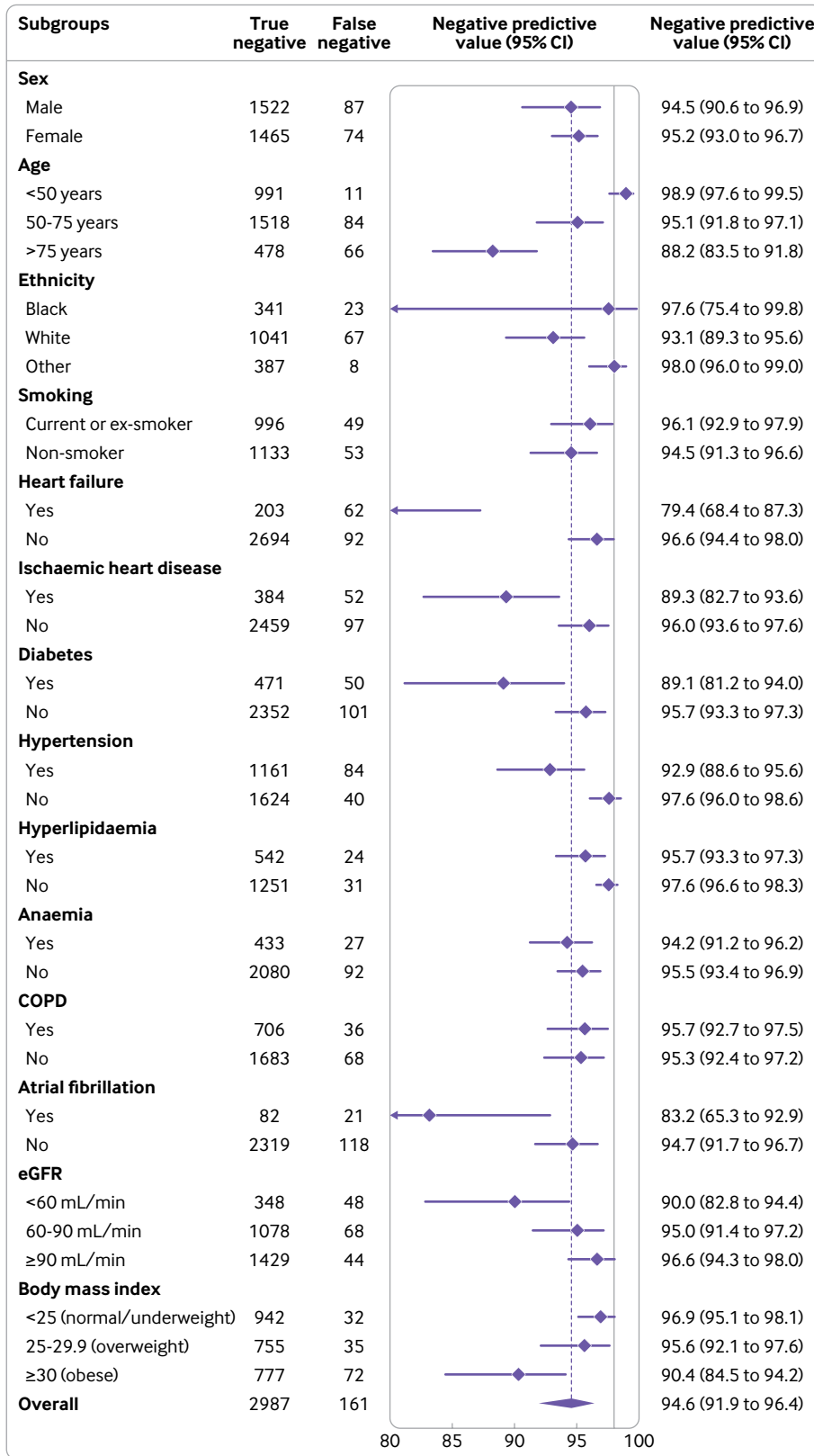| Subgroups	True negative	False negative	Negative predictive value (95% CI)	Negative predictive value (95% CI)
Sex				
Male	1522	87		94.5 (90.6 to 96.9)
Female	1465	74		95.2 (93.0 to 96.7)
Age				
<50 years	991	11		98.9 (97.6 to 99.5)
50-75 years	1518	84		95.1 (91.8 to 97.1)
>75 years	478	66		88.2 (83.5 to 91.8)
Ethnicity				
Black	341	23		97.6 (75.4 to 99.8)
White	1041	67		93.1 (89.3 to 95.6)
Other	387	8		98.0 (96.0 to 99.0)
Smoking				
Current or ex-smoker	996	49		96.1 (92.9 to 97.9)
Non-smoker	1133	53		94.5 (91.3 to 96.6)
Heart failure				
Yes	203	62		79.4 (68.4 to 87.3)
No	2694	92		96.6 (94.4 to 98.0)
Ischaemic heart disease				
Yes	384	52		89.3 (82.7 to 93.6)
No	2459	97		96.0 (93.6 to 97.6)
Diabetes				
Yes	471	50		89.1 (81.2 to 94.0)
No	2352	101		95.7 (93.3 to 97.3)
Hypertension				
Yes	1161	84		92.9 (88.6 to 95.6)
No	1624	40		97.6 (96.0 to 98.6)
Hyperlipidaemia				
Yes	542	24		95.7 (93.3 to 97.3)
No	1251	31		97.6 (96.6 to 98.3)
Anaemia				
Yes	433	27		94.2 (91.2 to 96.2)
No	2080	92		95.5 (93.4 to 96.9)
COPD				
Yes	706	36		95.7 (92.7 to 97.5)
No	1683	68		95.3 (92.4 to 97.2)
Atrial fibrillation				
Yes	82	21		83.2 (65.3 to 92.9)
No	2319	118		94.7 (91.7 to 96.7)
eGFR				
<60 mL/min	348	48		90.0 (82.8 to 94.4)
60-90 mL/min	1078	68		95.0 (91.4 to 97.2)
≥90 mL/min	1429	44		96.6 (94.3 to 98.0)
Body mass index				
<25 (normal/underweight)	942	32		96.9 (95.1 to 98.1)
25-29.9 (overweight)	755	35		95.6 (92.1 to 97.6)
≥30 (obese)	777	72		90.4 (84.5 to 94.2)
Overall	2987	161		94.6 (91.9 to 96.4)

80 85 90 95 100

Fig 2 | Diagnostic performance of guideline recommended N-terminal pro-B-type natriuretic peptide thresholds across patient subgroups: negative predictive value of threshold of 300 pg/mL. COPD=chronic obstructive pulmonary disease; eGFR=estimated glomerular filtration rate

versus 4.0% and 10.4%, respectively; one year all cause mortality: 5.9% versus 17.8% and 33.4%, respectively; 30 day cardiovascular mortality: 0.2% versus 0.8% and

4.1%; one year cardiovascular mortality: 1.4% versus 3.4% and 16.3%, respectively) (fig 9). In patients with NT-proBNP concentrations <300 pg/mL compared with
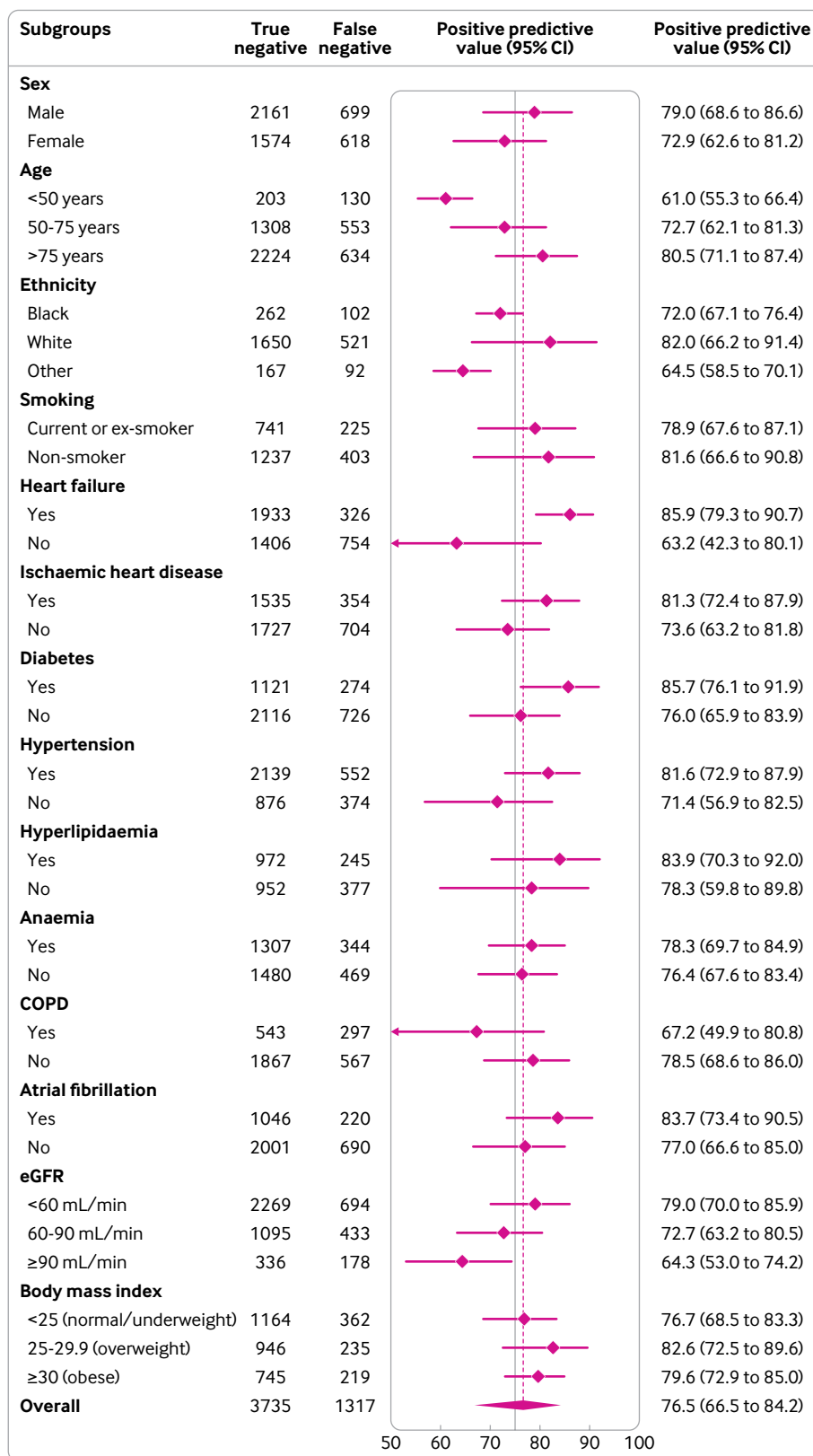
Subgroups	True negative	False negative	Positive predictive value (95% CI)	Positive predictive value (95% CI)
Sex				
Male	2161	699		79.0 (68.6 to 86.6)
Female	1574	618		72.9 (62.6 to 81.2)
Age				
<50 years	203	130		61.0 (55.3 to 66.4)
50-75 years	1308	553		72.7 (62.1 to 81.3)
>75 years	2224	634		80.5 (71.1 to 87.4)
Ethnicity				
Black	262	102		72.0 (67.1 to 76.4)
White	1650	521		82.0 (66.2 to 91.4)
Other	167	92		64.5 (58.5 to 70.1)
Smoking				
Current or ex-smoker	741	225		78.9 (67.6 to 87.1)
Non-smoker	1237	403		81.6 (66.6 to 90.8)
Heart failure				
Yes	1933	326		85.9 (79.3 to 90.7)
No	1406	754		63.2 (42.3 to 80.1)
Ischaemic heart disease				
Yes	1535	354		81.3 (72.4 to 87.9)
No	1727	704		73.6 (63.2 to 81.8)
Diabetes				
Yes	1121	274		85.7 (76.1 to 91.9)
No	2116	726		76.0 (65.9 to 83.9)
Hypertension				
Yes	2139	552		81.6 (72.9 to 87.9)
No	876	374		71.4 (56.9 to 82.5)
Hyperlipidaemia				
Yes	972	245		83.9 (70.3 to 92.0)
No	952	377		78.3 (59.8 to 89.8)
Anaemia				
Yes	1307	344		78.3 (69.7 to 84.9)
No	1480	469		76.4 (67.6 to 83.4)
COPD				
Yes	543	297		67.2 (49.9 to 80.8)
No	1867	567		78.5 (68.6 to 86.0)
Atrial fibrillation				
Yes	1046	220		83.7 (73.4 to 90.5)
No	2001	690		77.0 (66.6 to 85.0)
eGFR				
<60 mL/min	2269	694		79.0 (70.0 to 85.9)
60-90 mL/min	1095	433		72.7 (63.2 to 80.5)
≥90 mL/min	336	178		64.3 (53.0 to 74.2)
Body mass index				
<25 (normal/underweight)	1164	362		76.7 (68.5 to 83.3)
25-29.9 (overweight)	946	235		82.6 (72.5 to 89.6)
≥30 (obese)	745	219		79.6 (72.9 to 85.0)
Overall	3735	1317		76.5 (66.5 to 84.2)

(x-axis: 50, 60, 70, 80, 90, 100)

Fig 3 | Diagnostic performance of guideline recommended NT-proBNP thresholds across patient subgroups: positive predictive value of age specific thresholds across patient subgroups (450, 900, and 1800 pg/mL for <50, 50-75, and >75 years, respectively). COPD=chronic obstructive pulmonary disease; eGFR=estimated glomerular filtration rate

those ≥300 pg/mL, the all cause mortality rates were 0.8% versus 7.6% at 30 days and 5.9% versus 26.6% at one year, respectively, and the cardiovascular mortality rates were 0.1% versus 2.6% at 30 days and 1.3% versus 10.2% at one year, respectively (supplementary table H; supplementary figure O).

Table 2 | Diagnostic performance of age specific thresholds of N-terminal pro-B-type natriuretic peptide (NT-proBNP) for acute heart failure

Age groups	NT-proBNP threshold (pg/mL)	True positive	False positive	True negative	False negative	Prevalence of acute heart failure (%)	NPV (95% CI)	PPV (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
‹50 years	450	203	130	1025	19	16.1	98.4 (96.2 to 99.3)	61.0 (55.3 to 66.4)	91.4 (87.0 to 94.5)	87.8 (79.5 to 93.0)
50-75 years	900	1407	575	2121	267	38.3	88.3 (82.9 to 92.2)	73.5 (62.3 to 82.3)	83.2 (76.0 to 88.6)	81.1 (72.6 to 87.5)
›75 years	1800	2135	621	1348	518	57.4	72.2 (63.4 to 79.7)	80.2 (70.9 to 87.1)	79.3 (74.2 to 83.5)	73.1 (65.2 to 79.8)
All	300	4388	2833	2987	161	43.9	94.6 (91.9 to 96.4)	62.9 (51.3 to 73.3)	96.8 (94.6 to 98.1)	49.3 (35.4 to 63.4)

CI=confidence interval; NPV=negative predictive value; PPV=positive predictive value.

Discussion

We did a meta-analysis of individual patient level data to evaluate the diagnostic performance of NT-proBNP thresholds in more than 10 000 patients
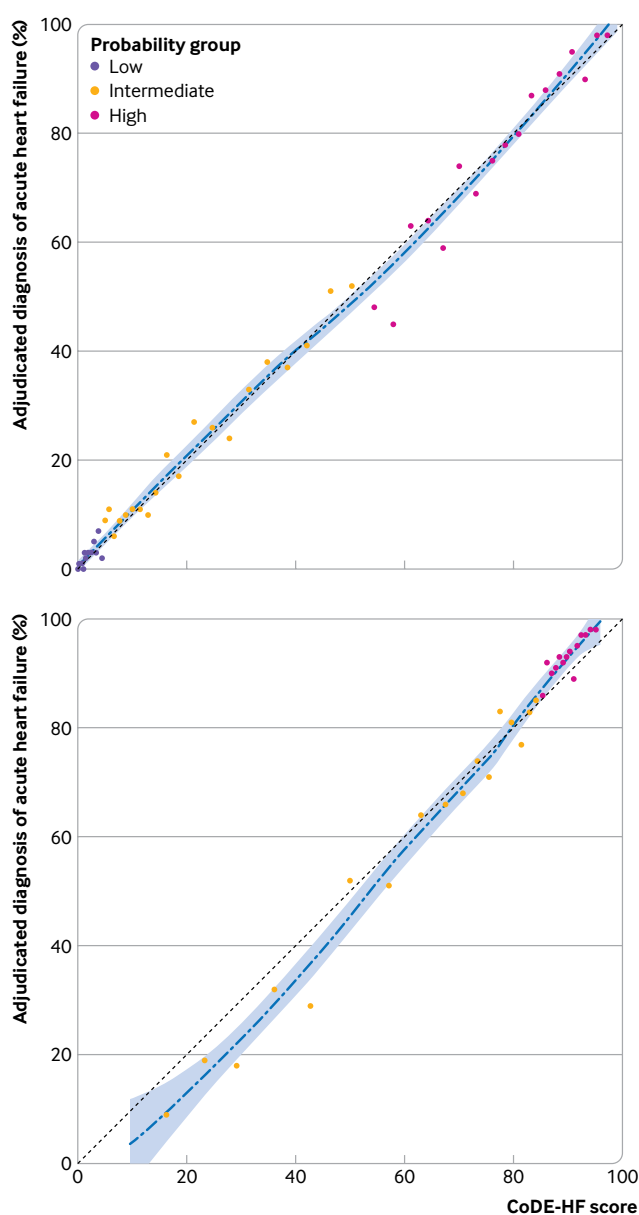

Fig 4 | Calibration of Collaboration for the Diagnosis and Evaluation of Heart Failure (CoDE-HF) score with observed proportion of patients with acute heart failure. Dashed line represents perfect calibration. Each point represents 100 patients. Top: calibration of CoDE-HF in patients with no previous heart failure. Bottom: calibration of CoDE-HF in patients with previous heart failure
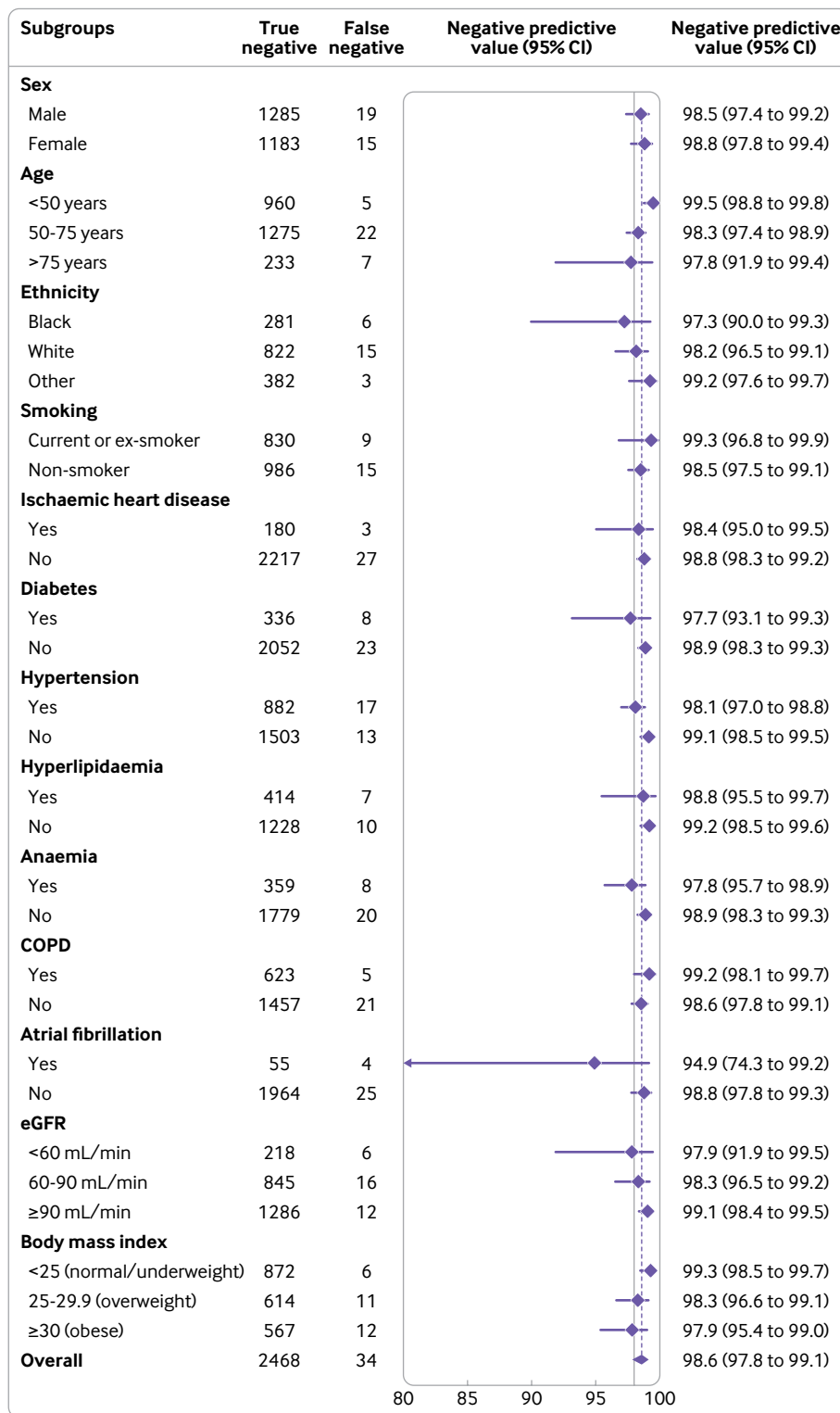
with suspected acute heart failure recruited in 14 prospective studies from 13 countries, and we developed and externally validated a decision support tool that uses NT-proBNP as a continuous measure. We report several important findings. Firstly, the guideline recommended threshold to rule out acute heart failure was heterogeneous across important patient subgroups.[3] Whereas performance was good in the overall population and in several subgroups such as younger patients and women, the negative predictive value was substantially lower in older patients and in those with obesity or previous heart failure, for whom the false negative rates were between one in 10 and one in five. Secondly, age stratified thresholds performed well to rule in the diagnosis of acute heart failure. However, the positive predictive value was lower in younger patients. Thirdly, although our optimised NT-proBNP thresholds of 100 pg/mL to rule out and 1000 pg/mL to rule in acute heart failure had excellent negative and positive predictive values in the overall population, performance was lower in older patients and in those with previous heart failure and obesity. Finally, we developed and validated a decision support tool, the CoDE-HF score, which had excellent diagnostic performance across all patient subgroups. This decision support tool ruled in and ruled out acute heart failure more accurately than any approach using NT-proBNP thresholds alone.

Strengths of study

To our knowledge, this is the largest study evaluating the diagnostic performance of NT-proBNP for acute heart failure to date. All included studies were prospective, and the final diagnosis was adjudicated by a panel of clinicians using all available information. Importantly, the availability of individual patient level data across a large study population allowed a robust evaluation of diagnostic performance across a full range of possible NT-proBNP thresholds within patient subgroups and the development and validation of a novel diagnostic score.

Comparison with other studies

Most national and international guidelines recommend the use of an NT-proBNP threshold of 300 pg/mL to rule out acute heart failure,[5 8] on the basis of evidence from multiple previous studies,[3 4 41 42] which reported a negative predictive value of 98% at this threshold. However, these studies were unable to evaluate the diagnostic performance within important patient subgroups. Our study, which included three

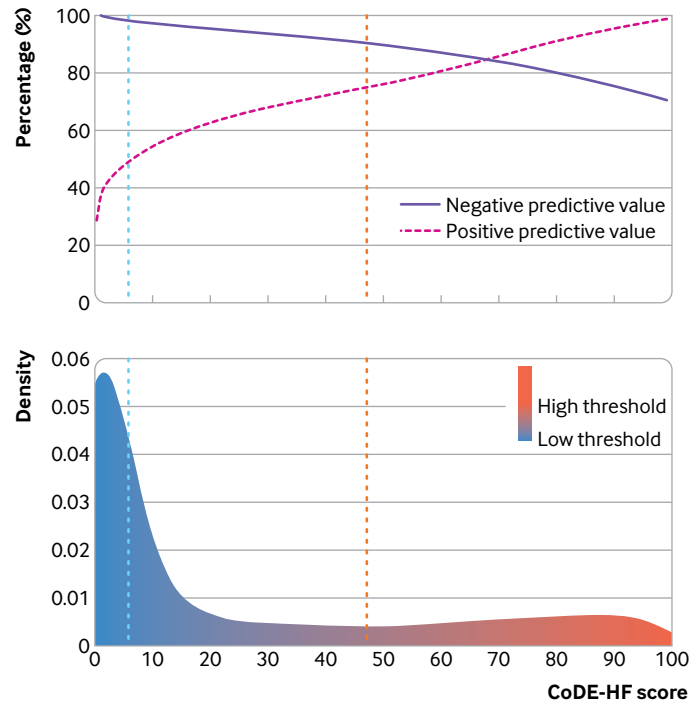| Subgroups	True negative	False negative	Negative predictive value (95% CI)	Negative predictive value (95% CI)
Sex				
Male	1285	19		98.5 (97.4 to 99.2)
Female	1183	15		98.8 (97.8 to 99.4)
Age				
<50 years	960	5		99.5 (98.8 to 99.8)
50-75 years	1275	22		98.3 (97.4 to 98.9)
>75 years	233	7		97.8 (91.9 to 99.4)
Ethnicity				
Black	281	6		97.3 (90.0 to 99.3)
White	822	15		98.2 (96.5 to 99.1)
Other	382	3		99.2 (97.6 to 99.7)
Smoking				
Current or ex-smoker	830	9		99.3 (96.8 to 99.9)
Non-smoker	986	15		98.5 (97.5 to 99.1)
Ischaemic heart disease				
Yes	180	3		98.4 (95.0 to 99.5)
No	2217	27		98.8 (98.3 to 99.2)
Diabetes				
Yes	336	8		97.7 (93.1 to 99.3)
No	2052	23		98.9 (98.3 to 99.3)
Hypertension				
Yes	882	17		98.1 (97.0 to 98.8)
No	1503	13		99.1 (98.5 to 99.5)
Hyperlipidaemia				
Yes	414	7		98.8 (95.5 to 99.7)
No	1228	10		99.2 (98.5 to 99.6)
Anaemia				
Yes	359	8		97.8 (95.7 to 98.9)
No	1779	20		98.9 (98.3 to 99.3)
COPD				
Yes	623	5		99.2 (98.1 to 99.7)
No	1457	21		98.6 (97.8 to 99.1)
Atrial fibrillation				
Yes	55	4		94.9 (74.3 to 99.2)
No	1964	25		98.8 (97.8 to 99.3)
eGFR				
<60 mL/min	218	6		97.9 (91.9 to 99.5)
60-90 mL/min	845	16		98.3 (96.5 to 99.2)
≥90 mL/min	1286	12		99.1 (98.4 to 99.5)
Body mass index				
<25 (normal/underweight)	872	6		99.3 (98.5 to 99.7)
25-29.9 (overweight)	614	11		98.3 (96.6 to 99.1)
≥30 (obese)	567	12		97.9 (95.4 to 99.0)
Overall	2468	34		98.6 (97.8 to 99.1)

(x-axis: 80 85 90 95 100)

Fig 5 | Diagnostic performance of Collaboration for the Diagnosis and Evaluation of Heart Failure (CoDE-HF) score across patient subgroups: negative predictive value of CoDE-HF rule-out score of 4.7 in patients without previous heart failure across patient subgroups. CoDE-HF incorporates N-terminal pro-B-type natriuretic peptide concentrations as continuous measure and predefined simple objective clinical variables (age, estimated glomerular filtration rate (eGFR), haemoglobin, body mass index, heart rate, blood pressure, peripheral oedema, chronic obstructive pulmonary disease (COPD), and ischaemic heart disease) to provide individualised assessment of likelihood of diagnosis of acute heart failure

times as many patients as a previous study level meta-analysis,[3] showed a lower overall negative predictive value at the 300 pg/mL threshold, with a pooled meta-estimate of 94.6%. More importantly, the negative predictive value was markedly lower in key subgroups, such as older patients and those with previous heart failure, ischaemic heart disease, and obesity. Moreover, nearly 70% of all patients had an NT-

Subgroups	True negative	False negative	Positive predictive value (95% CI)	Positive predictive value (95% CI)
Sex				
Male	748	223		77.5 (66.8 to 85.5)
Female	578	188		73.2 (63.7 to 81.0)
Age				
<50 years	63	24		72.4 (62.1 to 80.8)
50-75 years	466	166		76.7 (64.7 to 85.5)
>75 years	797	221		74.4 (65.4 to 81.7)
Ethnicity				
Black	52	26		66.7 (55.5 to 76.2)
White	723	216		79.1 (63.7 to 89.1)
Other	78	28		73.6 (64.4 to 81.1)
Smoking				
Current or ex-smoker	262	69		78.5 (72.0 to 83.8)
Non-smoker	507	133		80.1 (64.1 to 90.0)
Ischaemic heart disease				
Yes	457	123		77.1 (66.7 to 85.0)
No	824	273		74.2 (64.7 to 81.9)
Diabetes				
Yes	377	98		82.1 (70.4 to 89.8)
No	918	295		74.3 (65.9 to 81.2)
Hypertension				
Yes	841	210		78.6 (70.6 to 84.9)
No	379	120		73.9 (63.9 to 82.0)
Hyperlipidaemia				
Yes	294	73		81.8 (66.3 to 91.1)
No	426	132		78.7 (62.9 to 89.0)
Anaemia				
Yes	438	125		77.1 (70.0 to 82.9)
No	664	187		74.8 (67.7 to 80.8)
COPD				
Yes	152	65		69.3 (58.7 to 78.1)
No	757	224		75.9 (67.2 to 82.9)
Atrial fibrillation				
Yes	426	90		80.7 (68.0 to 89.1)
No	802	271		75.7 (67.3 to 82.5)
eGFR				
<60 mL/min	658	207		76.4 (67.3 to 83.6)
60-90 mL/min	501	155		74.4 (64.8 to 82.1)
≥90 mL/min	150	48		75.8 (69.3 to 81.2)
Body mass index				
<25 (normal/underweight)	460	123		78.6 (73.4 to 83.0)
25-29.9 (overweight)	335	109		74.2 (61.5 to 83.8)
≥30 (obese)	302	90		79.5 (69.2 to 87.0)
Overall	1326	411		75.0 (65.7 to 82.5)

50 60 70 80 90 100

Fig 6 | Diagnostic performance of Collaboration for the Diagnosis and Evaluation of Heart Failure (CoDE-HF) score across patient subgroups: positive predictive value of CoDE-HF rule-in score of 51.2 in patients without previous heart failure across patient subgroups. CoDE-HF incorporates NT-proBNP concentrations as continuous measure and predefined simple objective clinical variables (age, estimated glomerular filtration rate (eGFR), haemoglobin, body mass index, heart rate, blood pressure, peripheral oedema, chronic obstructive pulmonary disease (COPD), and ischaemic heart disease) to provide individualised assessment of likelihood of diagnosis of acute heart failure

proBNP concentration above the 300 pg/mL threshold, highlighting the limitation of using a single threshold in practice. Although a lower threshold of 100 pg/mL achieved an overall negative predictive value of 98%, performance was poorer within important patient subgroups. Furthermore, the age specific and optimised thresholds to rule in acute heart failure had heterogeneous performance within patient subgroups,

Subgroups	True negative	False negative	Positive predictive value (95% CI)	Positive predictive value (95% CI)
Sex				
Male	822	59		93.5 (89.5 to 96.0)
Female	503	36		92.4 (87.4 to 95.5)
Age				
<50 years	62	3		95.4 (86.6 to 98.5)
50-75 years	410	34		92.1 (87.5 to 95.2)
>75 years	853	58		93.0 (89.2 to 95.5)
Ethnicity				
Black	103	8		92.8 (86.2 to 96.4)
White	595	49		92.5 (86.5 to 95.9)
Other	39	5		88.6 (75.5 to 95.2)
Smoking				
Current or ex-smoker	302	18		94.1 (89.6 to 96.7)
Non-smoker	467	37		93.0 (87.9 to 96.0)
Ischaemic heart disease				
Yes	797	62		92.3 (86.9 to 95.5)
No	501	31		94.1 (89.9 to 96.6)
Diabetes				
Yes	491	28		95.0 (88.9 to 97.9)
No	823	65		92.3 (88.7 to 94.8)
Hypertension				
Yes	916	64		93.3 (89.4 to 95.8)
No	294	13		95.4 (89.0 to 98.2)
Hyperlipidaemia				
Yes	494	33		93.7 (91.3 to 95.5)
No	343	21		95.4 (86.8 to 98.5)
Anaemia				
Yes	646	39		93.4 (89.4 to 96.0)
No	517	37		93.0 (89.0 to 95.6)
COPD				
Yes	171	8		95.3 (89.4 to 98.0)
No	757	59		92.3 (88.9 to 94.7)
Atrial fibrillation				
Yes	455	32		92.9 (86.9 to 96.2)
No	814	55		93.5 (90.4 to 95.6)
eGFR				
<60 mL/min	942	72		92.5 (89.2 to 94.8)
60-90 mL/min	293	21		93.1 (86.1 to 96.7)
≥90 mL/min	75	1		98.7 (91.2 to 99.8)
Body mass index				
<25 (normal/underweight)	488	37		92.6 (89.0 to 95.1)
25-29.9 (overweight)	431	26		94.5 (89.0 to 97.3)
≥30 (obese)	264	16		94.3 (90.9 to 96.5)
Overall	1325	95		92.7 (89.1 to 95.2)

50 60 70 80 90 100

Fig 7 | Diagnostic performance of Collaboration for the Diagnosis and Evaluation of Heart Failure (CoDE-HF) score across patient subgroups: positive predictive value of CoDE-HF rule-in score of 84.5 in patients with previous heart failure across patient subgroups. CoDE-HF incorporates NT-proBNP concentrations as continuous measure and predefined simple objective clinical variables (age, estimated glomerular filtration rate (eGFR), haemoglobin, body mass index, heart rate, blood pressure, peripheral oedema, chronic obstructive pulmonary disease (COPD), and ischaemic heart disease) to provide individualised assessment of likelihood of diagnosis of acute heart failure

particularly in those without previous heart failure. This heterogeneity in diagnostic performance is particularly concerning as our patient populations are ageing and living with more comorbidities. This raises the question as to whether clinical guidelines should continue to recommend use of uniform thresholds when NT-proBNP is influenced by many risk factors and comorbidities.

Fig 8 | Diagnostic performance of Collaboration for the Diagnosis and Evaluation of Heart Failure (CoDE-HF) score in patients without previous heart failure. Top: negative and positive predictive values of CoDE-HF scores. Blue vertical dashed line represents target rule-out score of 4.7. Red vertical dashed line represents target rule-in score of 51.2. Bottom: density plot of CoDE-HF score in patients without previous heart failure. Target rule-out and rule-in scores identify 40.3% of patients as low probability and 28.0% as high probability respectively

To improve the clinical utility of NT-proBNP, we developed and externally validated a clinical decision support tool, the CoDE-HF score. This score incorporates NT-proBNP as a continuous measure together with simple, objective clinical variables to provide an individualised assessment of the likelihood of the diagnosis of acute heart failure. We showed that the diagnostic performance of the CoDE-HF score was robust across patient subgroups. CoDE-HF was able to rule out and rule in the diagnosis of acute heart failure in a larger proportion of patients compared with using optimised NT-proBNP thresholds alone. Furthermore, in our decision curve analysis, we found that CoDE-HF had a higher net benefit than NT-proBNP alone across the full range of threshold probabilities. We believe that this finding is intuitive given that NT-proBNP is

a continuous marker of risk and concentrations are influenced by other patient related factors, such as body mass index, age, and renal function.[43-45] Although these proportions were calculated on the basis of pre-specified performance criteria, we acknowledge that these targets may not be universally supported and different healthcare settings may have different tolerances for risk. The advantage of using a decision support tool, such as CoDE-HF, is that clinicians or institutions have the option to select the diagnostic performance criteria used to guide local decisions depending on their priorities and the availability of echocardiography or heart failure specialists.

Implications for practice and future research
We anticipate that CoDE-HF, our novel decision support tool, has the potential to improve the triage of patients with suspected acute heart failure presenting to multiple medical specialties and to transform their care by facilitating more accurate diagnosis. Previous studies have showed that prompt and accurate delivery of evidence based treatments for patients with acute heart failure can lead to a substantial reduction in mortality and length of hospital stay, whereas delays are associated with worse outcomes.[46] Furthermore, CoDE-HF uses routinely collected variables and therefore can be embedded within the clinical workflow as part of the triage pathway in the emergency department to facilitate more efficient assessment. Currently, the vast majority of patients with suspected acute heart failure undergo echocardiography during the course of their hospital admission to guide their

Fig 9 | Cumulative incidence of all cause mortality stratified by Collaboration for the Diagnosis and Evaluation of Heart Failure (CoDE-HF) probability group

care, but only a proportion of these patients ultimately have the diagnosis.[2] Echocardiography is a relatively time consuming and resource intensive specialist investigation. We anticipate that use of CoDE-HF to guide more accurate and judicious use of specialist services such as echocardiography could lead to significant cost and efficiency savings for healthcare systems. Additionally, cost savings could also be made by triaging patients at low risk to outpatient care. A prospective study is now needed to evaluate the clinical and cost effectiveness of different CoDE-HF decision thresholds in clinical practice.

Limitations of study

We acknowledge several limitations. Firstly, we were able to obtain individual patient level data for 14 of 30 studies that met our eligibility criteria, so the introduction of selection bias is a possibility. Nevertheless, the eligible studies that were not included had similar prevalence of acute heart failure, publication dates, and geographical coverage, and the population had similar demographic and clinical characteristics to those that were included. Secondly, in combining information from multiple studies, data were missing for some variables in several studies. To maximise the use of information, we have used a hierarchical multiple imputation method. Thirdly, we did not have consistently recorded data from the electrocardiogram and chest radiograph to enable inclusion of these data in our model. The interpretation of NT-proBNP in patients with suspected acute heart failure should be made in conjunction with these investigations,[47] and future studies are needed to determine whether approaches that integrate these investigations can improve the performance of CoDE-HF. Fourthly, not all studies have adjudicated the diagnosis blinded to the results of NT-proBNP testing. In our sensitivity analysis, diagnostic performance was unchanged when we excluded the two studies in which adjudication was unblinded. Fifthly, the adjudicated diagnosis of acute heart failure did not differentiate between heart failure with reduced ejection fraction and heart failure with preserved ejection fraction.[48] The increasing prevalence of heart failure with preserved ejection fraction in older patients may explain some of the heterogeneity observed across age groups, but current guidelines recommend the same NT-ProBNP threshold for both heart failure with reduced ejection fraction and heart failure with preserved ejection fraction.[5] [8] Sixthly, although most of the studies included consecutive patients with acute dyspnoea, the prevalence of acute heart failure was high and some selection bias may have existed. However, the performance of guideline recommended and age specific NT-proBNP thresholds were unchanged in a sensitivity analysis excluding studies at high risk of bias. Finally, acute heart failure is a clinical syndrome and the diagnostic adjudication itself has inherent uncertainty and variability across studies. This uncertainty may be greater among older adults,

which may, in part, explain some of the observed heterogeneity in diagnostic performance.

Conclusions

We have shown that the diagnostic performance of guideline recommended NT-proBNP thresholds for acute heart failure varies across important patient subgroups. We developed and validated the CoDE-HF score, which combines NT-pro-BNP as a continuous measure with clinical variables by using statistical modelling to determine the probability of acute heart failure for individual patients. This decision support tool accurately ruled in and ruled out acute heart failure and performed consistently across all subgroups. Prospective studies are now needed to evaluate the effect of implementing this decision support tool on healthcare resource utilisation and patients' outcomes.

AUTHOR AFFILIATIONS

[1]British Heart Foundation Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, UK

[2]Usher Institute, University of Edinburgh, Edinburgh, UK

[3]Department of Biochemistry, Cochin Hospital, Assistance Publique-Hopitaux de Paris, Paris, France

[4]Department of Emergency Medicine, Princess Grace Hospital Center, Monaco, Principality of Monaco

[5]Cardiovascular Research Institute of Basel, Department of Cardiology, University Hospital Basel, Basel, Switzerland

[6]Department of Internal Medicine, University Hospital Basel, University of Basel, Switzerland

[7]Liverpool Heart and Chest Hospital, Liverpool, UK

[8]Inova Heart and Vascular Institute, Falls Church, VA, USA

[9]Division of Nephrology, University of Maryland School of Medicine, Baltimore, MD, USA

[10]University of Toronto, St Michael's Hospital, Toronto, ON, Canada

[11]Heart Institute, Hospital Universitari Germans Trias i Pujol, Badalona, CIBERCV, Spain

[12]Department of Cardiology, Radboud University Medical Center, Nijmegen, Netherlands

[13]University of Amsterdam, Amsterdam, Netherlands

[14]Harvard Medical School, Boston, MA, USA

[15]Division of Cardiology, Massachusetts General Hospital, Boston, MA, USA

[16]BRAHMS, Thermo Fisher Scientific, Hennigsdorf, Germany

[17]Department of Emergency and Acute Medicine with Chest Pain Units, Charité – Universitätsmedizin Berlin, Campus Mitte and Virchow, Berlin, Germany

[18]Department of Internal Medicine, Radboud University Medical Center, Nijmegen, Netherlands

[19]Department of Internal Medicine, Division of Pharmacology and Vascular Medicine, Erasmus Medical Center, Rotterdam, Netherlands

[20]Department of Surgical, Medical and Molecular Pathology and Critical Care Medicine, University of Pisa, Pisa, Italy

[21]Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy

[22]Christchurch Heart Institute, University of Otago, Christchurch, New Zealand

[23]Emergency Medicine Department, National University Hospital, Singapore

[24]Department of Internal Medicine, Krankenhaus Bad Ischl, Bad Ischl, Austria

[25]Department of Laboratory Medicine, Hospital Voecklabruck, Voecklabruck, Austria

[26]Institute for Clinical Chemistry, University Medical Centre Mannheim, Faculty of Medicine Mannheim, University of Heidelberg, Mannheim, Germany

[27] First Department of Medicine, University Medical Centre Mannheim, Faculty of Medicine Mannheim, University of Heidelberg, Mannheim, Germany

[28] University of Milan Bicocca, ASST-Brianza, Pio XI Hospital of Desio, Internal Medicine, Desio, Italy

[29] Clinica Medica, University Milan Bicocca, Milan, Italy

[30] Department of Emergency Medicine, Azienda Ospedaliero-Universitaria Careggi, Florence, Italy

[31] Department of Internal Medicine III, Division of Cardiology, University Hospital of Heidelberg, Ruprecht-Karls University Heidelberg, Heidelberg, Germany

[32] London School of Hygiene and Tropical Medicine, London, UK

[33] Cardiovascular Research Institute, National University Heart Centre Singapore, Singapore

[34] British Heart Foundation Cardiovascular Research Centre, University of Glasgow, Glasgow, UK

1 Conrad N, Judge A, Tran J, et al. Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. *Lancet* 2018;391:572-80. doi:10.1016/S0140-6736(17)32520-5

2 National Institute for Cardiovascular Outcomes Research. National Heart Failure Audit 2019 Summary Report. 2019. https://www.nicor.org.uk/wp-content/uploads/2019/09/Heart-Failure-2019-Report-final.pdf.

3 Roberts E, Ludman AJ, Dworzynski K, et al, NICE Guideline Development Group for Acute Heart Failure. The diagnostic accuracy of the natriuretic peptides in heart failure: systematic review and diagnostic meta-analysis in the acute care setting. *BMJ* 2015;350:h910. doi:10.1136/bmj.h910

4 Januzzi JLJr, Chen-Tournoux AA, Christenson RH, et al, ICON-RELOADED Investigators. N-Terminal Pro-B-Type Natriuretic Peptide in the Emergency Department: The ICON-RELOADED Study. *J Am Coll Cardiol* 2018;71:1191-200. doi:10.1016/j.jacc.2018.01.021

5 National Institute for Health and Care Excellence. Acute heart failure: diagnosis and management. Clinical Guideline 187. 2014. https://www.nice.org.uk/guidance/cg187.

6 Yancy CW, Jessup M, Bozkurt B, et al. 2017 ACC/AHA/HFSA Focused Update of the 2013 ACCF/AHA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. *Circulation* 2017;136:e137-61. doi:10.1161/CIR.0000000000000509

7 Mueller C, McDonald K, de Boer RA, et al, Heart Failure Association of the European Society of Cardiology. Heart Failure Association of the European Society of Cardiology practical guidance on the use of natriuretic peptide concentrations. *Eur J Heart Fail* 2019;21:715-31. doi:10.1002/ejhf.1494

8 McDonagh TA, Metra M, Adamo M, et al, Authors/Task Force Members, ESC Scientific Document Group. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). With the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur J Heart Fail* 2022;24:4-131. doi:10.1002/ejhf.2333

9 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. doi:10.1161/CIRCULATIONAHA.116.025941

10 Baggish AL, van Kimmenade RR, Januzzi JLJr. The differential diagnosis of an elevated amino-terminal pro-B-type natriuretic peptide level. *Am J Cardiol* 2008;101(3A):43-8. doi:10.1016/j.amjcard.2007.11.019

11   Kozhuharov N, Sabti Z, Wussler D, et al, BASEL V Investigators. Prospective validation of N-terminal pro B-type natriuretic peptide cut-off concentrations for the diagnosis of acute heart failure. *Eur J Heart Fail* 2019;21:813-5. doi:10.1002/ejhf.1471

12   Than MP, Pickering JW, Sandoval Y, et al, MI3 collaborative. Machine Learning to Predict the Likelihood of Acute Myocardial Infarction. *Circulation* 2019;140:899-909. doi:10.1161/CIRCULATIONAHA.119.041980

13   Rahimi K, Bennett D, Conrad N, et al. Risk prediction in patients with heart failure: a systematic review and analysis. *JACC Heart Fail* 2014;2:440-6. doi:10.1016/j.jchf.2014.04.008

14   Pocock SJ, Ariti CA, McMurray JJ, et al, Meta-Analysis Global Group in Chronic Heart Failure. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. *Eur Heart J* 2013;34:1404-13. doi:10.1093/eurheartj/ehs337

15   Wussler D, Kozhuharov N, Sabti Z, et al. External Validation of the MEESSI Acute Heart Failure Risk Score: A Cohort Study. *Ann Intern Med* 2019;170:248-56. doi:10.7326/M18-1967

16   Simpson J, Jhund PS, Lund LH, et al. Prognostic Models Derived in PARADIGM-HF and Validated in ATMOSPHERE and the Swedish Heart Failure Registry to Predict Mortality and Morbidity in Chronic Heart Failure. *JAMA Cardiol* 2020;5:432-41. doi:10.1001/jamacardio.2019.5850

17   Baggish AL, Siebert U, Lainchbury JG, et al. A validated clinical and biochemical score for the diagnosis of acute heart failure: the ProBNP Investigation of Dyspnea in the Emergency Department (PRIDE) Acute Heart Failure Score. *Am Heart J* 2006;151:48-54. doi:10.1016/j.ahj.2005.02.031

18   Steinhart B, Thorpe KE, Bayoumi AM, Moe G, Januzzi JLJr, Mazer CD. Improving the diagnosis of acute heart failure using a validated prediction model. *J Am Coll Cardiol* 2009;54:1515-21. doi:10.1016/j.jacc.2009.05.065

19   Pivetta E, Goffi A, Lupia E, et al, SIMEU Group for Lung Ultrasound in the Emergency Department in Piedmont. Lung Ultrasound-Implemented Diagnosis of Acute Decompensated Heart Failure in the ED: A SIMEU Multicenter Study. *Chest* 2015;148:202-10. doi:10.1378/chest.14-2608

20   Whiting PF, Rutjes AW, Westwood ME, et al, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-36. doi:10.7326/0003-4819-155-8-201110180-00009

21   DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88. doi:10.1016/0197-2456(86)90046-2

22   Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *arXiv:1603.02754v3*2016. doi:10.1145/2939672.2939785.

23   Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. *Ann Stat* 2001;29:1189-232. doi:10.1214/aos/1013203451.

24   Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting. *Ann Stat* 2000;28:337-407. doi:10.1214/aos/1016218223.

25   Quartagno M, Carpenter JR. Multiple imputation for IPD meta-analysis: allowing for heterogeneity and studies with missing covariates. *Stat Med* 2016;35:2938-54. doi:10.1002/sim.6837

26   Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128-38. doi:10.1097/EDE.0b013e3181c30fb2

27   Debray TP, Riley RD, Rovers MM, Reitsma JB, Moons KG, Cochrane IPD Meta-analysis Methods group. Individual participant data (IPD) meta-analyses of diagnostic and prognostic modeling studies: guidance on their use. *PLoS Med* 2015;12:e1001886. doi:10.1371/journal.pmed.1001886

28   Bahrmann P, Bahrmann A, Hofner B, et al. Multiple biomarker strategy for improved diagnosis of acute heart failure in older patients presenting to the emergency department. *Eur Heart J Acute Cardiovasc Care* 2015;4:137-47. doi:10.1177/2048872614541904

29   Behnes M, Brueckmann M, Ahmad-Nejad P, et al. Diagnostic performance and cost effectiveness of measurements of plasma N-terminal pro brain natriuretic peptide in patients presenting with acute dyspnea or peripheral edema. *Int J Cardiol* 2009;135:165-74. doi:10.1016/j.ijcard.2008.03.045

30   Bombelli M, Maloberti A, Rossi S, et al. Clinical value of NT-proBNP assay in the emergency department for the diagnosis of heart failure (HF) in very elderly people. *Arch Gerontol Geriatr* 2015;61:296-300. doi:10.1016/j.archger.2015.05.001

31   Chenevier-Gobeaux C, Claessens YE, Voyer S, Desmoulins D, Ekindjian OG. Influence of renal function on N-terminal pro-brain natriuretic peptide (NT-proBNP) in patients admitted for dyspnoea in the Emergency Department: comparison with brain natriuretic peptide (BNP). *Clin Chim Acta* 2005;361:167-75. doi:10.1016/j.cccn.2005.05.021

32   deFilippi CR, Seliger SL, Maynard S, Christenson RH. Impact of renal disease on natriuretic peptide testing for diagnosing decompensated heart failure and predicting mortality. *Clin Chem* 2007;53:1511-9. doi:10.1373/clinchem.2006.084533

33   Gargani L, Frassi F, Soldati G, Tesorio P, Gheorghiade M, Picano E. Ultrasound lung comets for the differential diagnosis of acute cardiogenic dyspnoea: a comparison with natriuretic peptides. *Eur J Heart Fail* 2008;10:70-7. doi:10.1016/j.ejheart.2007.10.009

34   Ibrahim I, Kuan WS, Frampton C, et al. Superior performance of N-terminal pro brain natriuretic peptide for diagnosis of acute decompensated heart failure in an Asian compared with a Western setting. *Eur J Heart Fail* 2017;19:209-17. doi:10.1002/ejhf.612

35   Januzzi JL, van Kimmenade R, Lainchbury J, et al. NT-proBNP testing for diagnosis and short-term prognosis in acute destabilized heart failure: an international pooled analysis of 1256 patients: the International Collaborative of NT-proBNP Study. *Eur Heart J* 2006;27:330-7. doi:10.1093/eurheartj/ehi631

36   Maisel A, Mueller C, Nowak R, et al. Mid-region pro-hormone markers for diagnosis and prognosis in acute dyspnea: results from the BACH (Biomarkers in Acute Heart Failure) trial. *J Am Coll Cardiol* 2010;55:2062-76. doi:10.1016/j.jacc.2010.02.025

37   Moe GW, Howlett J, Januzzi JL, Zowall H, Canadian Multicenter Improved Management of Patients With Congestive Heart Failure (IMPROVE-CHF) Study Investigators. N-terminal pro-B-type natriuretic peptide testing improves the management of patients with suspected acute heart failure: primary results of the Canadian prospective randomized multicenter IMPROVE-CHF study. *Circulation* 2007;115:3103-10. doi:10.1161/CIRCULATIONAHA.106.666255

38   Mueller T, Gegenhuber A, Poelz W, Haltmayer M. Diagnostic accuracy of B type natriuretic peptide and amino terminal proBNP in the emergency diagnosis of heart failure. *Heart* 2005;91:606-12. doi:10.1136/hrt.2004.037762

39   Nazerian P, Vanni S, Zanobetti M, et al. Diagnostic accuracy of emergency Doppler echocardiography for identification of acute left ventricular heart failure in patients with acute dyspnea: comparison with Boston criteria and N-terminal prohormone brain natriuretic peptide. *Acad Emerg Med* 2010;17:18-26. doi:10.1111/j.1553-2712.2009.00630.x

40   Rutten JH, Steyerberg EW, Boomsma F, et al. N-terminal pro-brain natriuretic peptide testing in the emergency department: beneficial effects on hospitalization, costs, and outcome. *Am Heart J* 2008;156:71-7. doi:10.1016/j.ahj.2008.02.021

41   Sanz MP, Borque L, Rus A, Vicente B, Ramírez Y, Lasa L. Comparison of BNP and NT-proBNP assays in the approach to the emergency diagnosis of acute dyspnea. *J Clin Lab Anal* 2006;20:227-32. doi:10.1002/jcla.20146

42   Klemen P, Golub M, Grmec S. Combination of quantitative capnometry, N-terminal pro-brain natriuretic peptide, and clinical assessment in differentiating acute heart failure from pulmonary disease as cause of acute dyspnea in pre-hospital emergency setting: study of diagnostic accuracy. *Croat Med J* 2009;50:133-42. doi:10.3325/cmj.2009.50.133

43   Santaguida PL, Don-Wauchope AC, Ali U, et al. Incremental value of natriuretic peptide measurement in acute decompensated heart failure (ADHF): a systematic review. *Heart Fail Rev* 2014;19:507-19. doi:10.1007/s10741-014-9444-9

44   Maisel A, Mueller C, Adams KJr, et al. State of the art: using natriuretic peptide levels in clinical practice. *Eur J Heart Fail* 2008;10:824-39. doi:10.1016/j.ejheart.2008.07.014

45   Madamanchi C, Alhosaini H, Sumida A, Runge MS. Obesity and natriuretic peptides, BNP and NT-proBNP: mechanisms and diagnostic implications for heart failure. *Int J Cardiol* 2014;176:611-7. doi:10.1016/j.ijcard.2014.08.007

46   Maisel AS, Peacock WF, McMullin N, et al. Timing of immunoreactive B-type natriuretic peptide levels and treatment delay in acute decompensated heart failure: an ADHERE (Acute Decompensated Heart Failure National Registry) analysis. *J Am Coll Cardiol* 2008;52:534-40. doi:10.1016/j.jacc.2008.05.010

47   Galasko GIW, Barnes SC, Collinson P, Lahiri A, Senior R. What is the most cost-effective strategy to screen for left ventricular systolic dysfunction: natriuretic peptides, the electrocardiogram, hand-held echocardiography, traditional echocardiography, or their combination?*Eur Heart J* 2006;27:193-200. doi:10.1093/eurheartj/ehi559

48   Tschöpe C, Kasner M, Westermann D, Gaub R, Poller WC, Schultheiss HP. The role of NT-proBNP in the diagnostics of isolated diastolic dysfunction: correlation with echocardiographic and invasive measurements. *Eur Heart J* 2005;26:2277-84. doi:10.1093/eurheartj/ehi406

Web appendix: Supplementary materials