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,1,7 Dimitrios Doudesis,1,2,7 Mohamed Anwar,1,7 Federica Asteng,1 Camille Chenevier-Gobeaux,3 Yann-Éric Claessens,4 Desiree Wussler,5,6 Nikola Kozhuharov,5,7 Ivo Streb,5 Zaid Sabti,5 Christopher deFilippis,8 Stephen Seliger,9 Gordon Moe,10 Carlos Fernando,10 Antoni Bayes-Genis,11 Roland R van Kimmenade,12 Yigal Pinto,13 Hanna K Gaggin,14,15 Jan C Wiemers,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 Bahrman,31 David E Newby,1 Alan G Japp,1 Athanasios Tsanas,2 Anoop S V Shah,31,32 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

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.
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 alâ 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.
(QUADAS-2) tool, 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) and age specific rule-in thresholds (450, 900, and 1800 pg/mL for patients aged <50, 50-75, and >75 years, respectively) 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. We further evaluated the performance of these thresholds in pre-specified subgroups stratified by sex, age, disease severity, 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 high 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). 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. 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. 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. 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 10369 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). 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% (4549/10369) 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
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% (318/10,369) 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 those with previous history of heart failure, ischaemic heart disease, and impaired renal function (supplementary figure J). Similarly, a lower negative predictive value was seen in patients with acute heart failure and obesity (90.4%, 84.5% to 91.8%) and in those with previous history of heart failure, ischaemic heart disease, and impaired renal function (supplementary figure J). 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).

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 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 pooled positive 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 pooled positive predictive value of 74.9% (64.4% to 83.2%) and specificity of 76.1%

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%
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%...
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

<table>
<thead>
<tr>
<th>Subgroups</th>
<th>True negative</th>
<th>False negative</th>
<th>Negative predictive value (95% CI)</th>
<th>Negative predictive value (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1522</td>
<td>87</td>
<td>94.5 (90.6 to 96.9)</td>
<td>95.2 (93.0 to 96.7)</td>
</tr>
<tr>
<td>Female</td>
<td>1465</td>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50 years</td>
<td>991</td>
<td>11</td>
<td>98.9 (97.0 to 99.5)</td>
<td></td>
</tr>
<tr>
<td>50-75 years</td>
<td>1518</td>
<td>84</td>
<td>95.1 (91.8 to 97.1)</td>
<td></td>
</tr>
<tr>
<td>&gt;75 years</td>
<td>478</td>
<td>66</td>
<td>88.2 (83.5 to 91.8)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>341</td>
<td>23</td>
<td>97.6 (75.4 to 99.8)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1041</td>
<td>67</td>
<td>93.1 (89.3 to 95.6)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>387</td>
<td>8</td>
<td>98.0 (96.0 to 99.0)</td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current or ex-smoker</td>
<td>996</td>
<td>49</td>
<td>96.1 (92.9 to 97.9)</td>
<td></td>
</tr>
<tr>
<td>Non-smoker</td>
<td>1133</td>
<td>53</td>
<td>94.5 (91.3 to 96.6)</td>
<td></td>
</tr>
<tr>
<td>Heart failure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>203</td>
<td>62</td>
<td>79.4 (68.4 to 87.3)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2694</td>
<td>92</td>
<td>96.6 (94.4 to 98.0)</td>
<td></td>
</tr>
<tr>
<td>Ischaemic heart disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>384</td>
<td>52</td>
<td>89.3 (82.7 to 93.6)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2459</td>
<td>97</td>
<td>96.0 (93.6 to 97.6)</td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>471</td>
<td>50</td>
<td>89.1 (81.2 to 94.0)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2352</td>
<td>101</td>
<td>95.7 (93.3 to 97.3)</td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1161</td>
<td>84</td>
<td>92.9 (88.6 to 95.6)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1624</td>
<td>40</td>
<td>97.6 (96.0 to 98.6)</td>
<td></td>
</tr>
<tr>
<td>Hyperlipidaemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>542</td>
<td>24</td>
<td>95.7 (93.3 to 97.3)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1251</td>
<td>31</td>
<td>97.6 (96.6 to 98.3)</td>
<td></td>
</tr>
<tr>
<td>Anaemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>433</td>
<td>27</td>
<td>94.2 (91.2 to 96.2)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2080</td>
<td>92</td>
<td>95.5 (93.4 to 96.9)</td>
<td></td>
</tr>
<tr>
<td>COPD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>706</td>
<td>36</td>
<td>95.7 (92.7 to 97.5)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1683</td>
<td>68</td>
<td>95.3 (92.4 to 97.2)</td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>82</td>
<td>21</td>
<td>83.2 (65.3 to 92.9)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2319</td>
<td>118</td>
<td>94.7 (91.7 to 96.7)</td>
<td></td>
</tr>
<tr>
<td>eGFR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;60 mL/min</td>
<td>348</td>
<td>48</td>
<td>90.0 (82.8 to 94.4)</td>
<td></td>
</tr>
<tr>
<td>60-90 mL/min</td>
<td>1078</td>
<td>68</td>
<td>95.0 (91.4 to 97.2)</td>
<td></td>
</tr>
<tr>
<td>≥90 mL/min</td>
<td>1429</td>
<td>44</td>
<td>96.6 (94.3 to 98.0)</td>
<td></td>
</tr>
<tr>
<td>Body mass index</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;25 (normal/underweight)</td>
<td>942</td>
<td>32</td>
<td>96.9 (95.1 to 98.1)</td>
<td></td>
</tr>
<tr>
<td>25-29.9 (overweight)</td>
<td>755</td>
<td>35</td>
<td>95.6 (92.1 to 97.6)</td>
<td></td>
</tr>
<tr>
<td>≥30 (obese)</td>
<td>777</td>
<td>72</td>
<td>90.4 (84.5 to 94.2)</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>2987</td>
<td>161</td>
<td>94.6 (91.9 to 96.4)</td>
<td></td>
</tr>
</tbody>
</table>

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
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).
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 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. 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, on the basis of evidence from multiple previous studies, 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

---

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

<table>
<thead>
<tr>
<th>Age groups</th>
<th>NT-proBNP threshold (pg/mL)</th>
<th>True positive</th>
<th>False positive</th>
<th>False negative</th>
<th>True negative</th>
<th>Prevalence of acute heart failure (%)</th>
<th>NPV (95% CI)</th>
<th>PPV (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;50 years</td>
<td>450</td>
<td>203</td>
<td>130</td>
<td>1025</td>
<td>19</td>
<td>16.1</td>
<td>98.4 (96.2 to 99.3)</td>
<td>61.0 (55.3 to 66.6)</td>
<td>91.4 (87.0 to 95.5)</td>
<td>87.8 (79.5 to 93.0)</td>
</tr>
<tr>
<td>50-75 years</td>
<td>900</td>
<td>1407</td>
<td>575</td>
<td>2121</td>
<td>267</td>
<td>38.3</td>
<td>88.3 (82.9 to 92.2)</td>
<td>73.5 (62.3 to 82.3)</td>
<td>83.2 (76.0 to 88.6)</td>
<td>81.1 (72.6 to 87.5)</td>
</tr>
<tr>
<td>≥75 years</td>
<td>1800</td>
<td>2135</td>
<td>621</td>
<td>1348</td>
<td>518</td>
<td>57.4</td>
<td>72.2 (65.4 to 79.7)</td>
<td>80.2 (70.9 to 87.1)</td>
<td>79.3 (74.2 to 83.5)</td>
<td>73.1 (65.2 to 79.8)</td>
</tr>
<tr>
<td>All</td>
<td>300</td>
<td>4388</td>
<td>2833</td>
<td>2987</td>
<td>439</td>
<td>94.6</td>
<td>94.6 (91.9 to 96.4)</td>
<td>62.9 (51.3 to 73.3)</td>
<td>96.8 (94.6 to 98.1)</td>
<td>49.3 (35.4 to 63.6)</td>
</tr>
</tbody>
</table>

CI=confidence interval; NPV=negative predictive value; PPV=positive predictive value.
times as many patients as a previous study level meta-analysis, 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-
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.
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.

![Table](image)

**Table:** 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.
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 shown 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 care.

---

**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.

**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. 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, 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. 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. 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 population 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, Chanté – 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
Funding: This study was supported by the British Heart Foundation through a clinical research training fellowship (FS/18/25/33454), intermediate clinical research fellowship (FS/19/17/34172), chair award (CH/F/21/90010), programme grant (RG/20/10/34966), and research excellence award (RE/18/5/34216) and by the Medical Research Council (MR/N013166/1), Chief Scientist Office (CGA/19/01), and a joint grant from the Medical Research Council Confidence in Concept Scheme and the Data Driven Innovation programme (MRC/CiC/79). The funders had no role in the study design, the collection, analysis, and interpretation of data, in the writing of the report, or in the decision to submit the article for publication.

Conflicting interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare: support from the British Heart Foundation, Medical Research Council, and Chief Scientist Office; V-EC has received honorariums for lectures and publications from Biomérieux, Roche Diagnostics, and Thermo Fisher; Cjd has received consulting fees from Fuji Rebio, Ortho Diagnostics, Quidel, and Roche Diagnostics and a patent entitled “Methods for assessing differential risk for developing heart failure” (patent number: PCT/US2015/029938); SS has received a grant from Roche Diagnostics and a patent entitled “Methods for assessing differential risk for developing heart failure” (patent number: PCT/US2015/029938); ABG has received personal fees and non-financial support from Roche Diagnostics during the conduct of the study and personal fees from Abbott and AstraZeneca, grants, personal fees, and non-financial support from Boehringer-Ingelheim, and personal fees and non-financial support from Novartis and Vifor outside the submitted work. YP has received consulting fees from Roche Diagnostics, Pfizer, and Forbion and honorariums from CVOI and Daiichi Sankyo; HKG has received grants from Roche Diagnostics, Jana Care, Ortho Clinical, Novartis, Pfizer, Alynlam, and Accea (IONIS), consulting fees from Amgen, Eko, Merck, and Pfizer, and stock in Eko; JCW works as a biostatistician at the biotech company BRAHMS GmbH, part of Thermo Fisher Scientific; MM has received grants from Health Care Research Projects and Biomarker Research and personal fees from Consulting outside the submitted work; ASVS has received speaker fees from Abbott Diagnostics outside the submitted work; AMR has received grants, personal fees, and non-financial support from Roche Diagnostics outside the submitted work; JHV has received consulting fees from Alynlam, AstraZeneca, Bayer, Boehringer Ingelheim, BMS, Card unin, Dai- C, GSK, Ionis, KBP Biosciences, Novartis, Pfizer, and Therac cs, payments for advisory boards, symposiums, or lectures from Abbott, Alkem Metabolics, Canadian Medical and Surgical Knowledge Translation Research Group, Eric Lifesciences, Hikma, Lupin, Sun Pharmaceuticals, Medscape/Heart.Org, ProAdWise Communications, Radcliffe Cardiology, Servier, and the Corpus, has participated on a data safety monitoring board or advisory board for Cardi saly (MONITOR study) and Merck (VICTORIA trial), and works as company director for Global Clinical Trial Partners Ltd (GCTP) outside the submitted work; CM has received grants and non-financial support from several diagnostic companies during the conduct of the study and grants, personal fees, and non-financial support from several diagnostic companies outside the submitted work; LJ has received grants from Abbott Diagnostics, Innolife, and Novartis and consulting fees from Abbott, Jana Care, Novartis, Roche Diagnostics, Bristol-Myers Squibb, Janssen, and Prevenc o, NLM has received grants from Siemens Healthineers, consulting fees from Roche Diagnostics and LumiraDx, and speaker fees from Abbott Diagnostics and Siemens Healthineers outside the submitted work; KKL, DD, and NLM are employed by the University of Edinburgh, which has filed a patent on the CoDE-HF score (patent reference: PCT/GB2021/051470), no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: 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 analysis.

Data sharing: The R code and anonymised data used to develop and validate the CoDE-HF score can be made available to researchers on request to the corresponding author.

The lead author (KKJ) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Dissemination to participants and related patient and public communities: Findings of this study will be shared with clinicians and patients through national and international cardiology conferences and through social media platforms. A plain language summary will also be disseminated through a press release.

Provenance and peer review: Not commissioned; externally peer reviewed.

This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited: see: http://creativecommons.org/licenses/by/4.0/.
doi:10.1016/j.ijcard.2008.03.045

acute dyspnea or peripheral edema. N-terminal pro brain natriuretic peptide in patients presenting with performance and cost effectiveness of measurements of plasma


doi:10.1016/j.ejheart.2008.07.014


markers for diagnosis and prognosis in acute dyspnea: results from the BACH (Biomarkers in Acute Heart Failure) trial. Am Coll Cardiol 2010;55:2062-76. doi:10.1016/j.jacc.2010.02.025


jamcardio.2019.5850


15