Diagnosing heart failure in primary care

Roalfe, Andrea K.; Taylor, Clare J.; Kelder, Johannes C.; Hoes, Arno W.; Hobbs, F.D. Richard

DOI:
10.1002/ehf2.13311

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Roalfe, AK, Taylor, CJ, Kelder, JC, Hoes, AW & Hobbs, FDR 2021, 'Diagnosing heart failure in primary care: individual patient data meta-analysis of two European prospective studies', *ESC heart failure*, vol. 8, no. 3, pp. 2193-2201. https://doi.org/10.1002/ehf2.13311

ESC HEART FAILURE
ESC Heart Failure 2021; **8**: 2193–2201
Published online 23 March 2021 in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/ehf2.13311

ORIGINAL RESEARCH ARTICLE

Diagnosing heart failure in primary care: individual patient data meta-analysis of two European prospective studies

Andrea K. Roalfe[1], Clare J. Taylor[1], Johannes C. Kelder[2], Arno W. Hoes[3] and F.D. Richard Hobbs[1]*

[1]*Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, UK;* [2]*St Antonius Hospital, Nieuwegein, The Netherlands; and* [3]*Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht/Utrecht University, Utrecht, The Netherlands*

Abstract

Aims Natriuretic peptides are helpful in detecting chronic heart failure (HF) in primary care; however, there are a lack of data evaluating thresholds recommended by clinical guidelines. This study assesses the diagnostic accuracy of N-terminal pro-B-type natriuretic peptide (NT-proBNP) using combined individual patient data from two studies in the UK and the Netherlands.

Methods and results Random effects methods were used to estimate the performance characteristics of NT-proBNP thresholds recommended by the European Society of Cardiology (ESC) and the UK National Institute for Health and Care Excellence (NICE) guidelines. New onset HF was diagnosed in 313 of 1073 (29.2%) participants. Age, sex, and atrial fibrillation-adjusted NT-proBNP was a better predictor of HF with reduced ejection fraction (HFrEF) than HF preserved ejection fraction (HFpEF), with area under receiver operating characteristic curve of 0.82 95% CI (0.78 to 0.86) vs. 0.71 (0.66 to 0.75). In persons aged 70 years and over, the ESC threshold at 125 ng/L for detection of all-cause HF had summary negative predictive value (NPV) of 84.9% (81.6 to 88.2), positive predictive value (PPV) 68.1% (63.1 to 73.3), sensitivity 74.9% (69.5 to 80.3), and specificity 80.1% (76.9 to 83.4); the NICE threshold at 400 ng/L had summary NPV of 74.7% (72.1 to 77.2), PPV 81.8% (73.3 to 89.5), sensitivity 43.5% (37.2 to 49.8), and specificity 94.5% (92.3 to 96.7).

Conclusions N-terminal pro-B-type natriuretic peptide is better at detecting HFrEF than HFpEF in a primary care setting. In persons aged 70 and over, the ESC threshold of 125 ng/L is more accurate at detecting and excluding HF than the higher level suggested in NICE guidelines. More prospective data are required to establish the optimal NP threshold for detecting chronic HF in general practice.

Keywords Natriuretic peptide; NT-proBNP; Heart failure; Primary care; Diagnostic test accuracy

Received: 14 December 2020; Revised: 27 January 2021; Accepted: 2 March 2021
**Correspondence to: F.D. Richard Hobbs, Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK. Tel: +44(0)1865 289300. Email: richard.hobbs@phc.ox.ac.uk*

Introduction

Heart failure (HF) is an increasingly prevalent condition due to an ageing population[1] with only modest improvement observed in long-term survival in the UK over the last 20 years.[2,3] The importance of early diagnosis, when evidence-based treatments can be initiated, is therefore critical. There is evidence from routinely collected primary care data of delayed HF diagnosis despite most patients with HF consulting their GP prior to diagnosis.[4] However, making a diagnosis in primary care can be difficult as the symptoms of HF—typically breathlessness, ankle swelling, and fatigue—often overlap with other conditions.

Natriuretic peptide (NP) tests are a cost-effective aid to diagnosis.[5] These peptides are mainly produced in the myocardium of the left ventricle, increasing in response to stretch, commonly caused by pressure or fluid overload, and are raised in people with HF. B-type NP and the inactive

fragment NT-proBNP are available in primary care to assist HF diagnosis, but the threshold for ruling out HF differs between guidelines. The UK National Institute for Health and Care Excellence (NICE) guideline for chronic HF states that at a NT-proBNP level of <400 ng/L, a diagnosis of HF is unlikely.[6] The European Society of Cardiology (ESC) recommends a much lower cut-off of <125 ng/L to rule out HF.[7]

There is a paucity of prospective studies evaluating the use of NP tests in the primary care setting.[8] A recent systematic review by NICE[6] identified only five studies,[9–13] with low risk of bias, while even fewer studies reported on both HF with reduced (HFrEF) and preserved ejection fraction (HFpEF). The NICE guidelines considered evidence from the REFER study and its associated economic evaluation,[5] acknowledging the limited data available in primary care. By combining the REFER individual patient dataset with a similar prospective study carried out in primary care, we provide additional threshold evaluation in diagnosis of HF in primary care.

Methods

Data source and study population

Three hundred and fifty-two participants aged 55 and over were recruited to the primary care REFerral for EchocaRdiography (REFER) study from 28 general practices in the West Midlands region of the UK,[12] and 721 patients were recruited, with no age restriction, to the Utrecht HartFalen Onderzoek (UHFO) study from primary care practices in the catchment areas of eight hospitals in the Netherlands.[13] Each study recruited patients suspected of new onset HF in primary care with symptoms including shortness of breath, ankle oedema, or fatigue. The data from the two studies were combined for analysis.

Definitions and measurements

REFER study

Full details of methods are provided elsewhere.[14] In summary, this was a prospective validation study of the MICE rule,[15] a decision tool based on clinical symptoms, natriuretic peptide, or a combination of both. Participants were referred to a research clinic where assessments were carried out by a research nurse and echocardiographer. Information collected included medical history, symptoms, physical examination measurements, ECG, and echocardiography results. Blood was sampled for NP testing and renal function following a study protocol at all sites. The NT-proBNP level was determined immediately using a point-of-care (POC) device, Roche CARDIAC proBNP (second generation), on a Cobas h 232 reader (Roche Diagnostics, UK). Analytical characteristics provided by the manufacturer include range 60 to 9000 ng/L, limit of detection (LoD) of 6 ng/L, and coefficient of variation (CV) <15%.[16]

UHFO study

This was a diagnostic accuracy study aimed at developing and validating a clinical prediction rule to detect HF cases in primary care.[13] Participants were referred by GPs to one of eight rapid access outpatient clinics for a standardized diagnostic work-up including patient medical history, ECG, chest X-ray, spirometry, echocardiography, and NT-proBNP test. The NT-proBNP level was measured at one central laboratory using an automated ELISA assay on the Elecsys system (Roche Diagnostics, Germany) with range 5–35 000, LoD of 5 ng/L, and CV of <4.6%. Plasma samples were stored in the laboratory at −80°C prior to the assay being performed.

Outcome definition

REFER study

Quantitative measurement of left ventricular ejection fraction (LVEF) was undertaken, with diastolic function classified as moderate or severe, based on E'/e, interventricular septum, sex, and left atrial volume. HF diagnosis was made by an expert consensus panel of three cardiologists, who reviewed each case blinded to assessment by other panel members. ESC 2012 guidelines were used to define HF.[7] The panel followed a three-stepped approach to diagnosis: at step 1, clinical variables (excluding sex, history of myocardial infarction (MI), basal crepitations and ankle oedema), ECG, and echocardiography results were available; at step 2, sex, MI, crepitations, and oedema were added; at step 3, the NT-proBNP result was revealed. HF diagnosis at step 2, where the panel were blinded to NT-proBNP, was defined as the outcome for this paper.

Heart failure cases identified by the panel, with ejection fraction ≤50%, were labelled as having HFrEF, and those with EF >50% as HFpEF. HF with EF between 40% and 50%, recently identified in the 'grey' mid-range area as HFmrEF, were categorized as HFrEF for the purpose of this study.

UHFO study

Left ventricular ejection fraction was assessed semi-quantitatively and classified as normal, mild, moderate, and severe dysfunction. Diastolic function was categorized as normal, impaired relaxation, or restrictive filling. All available diagnostic and 6 month follow-up information for each patient (excluding NT-proBNP) were used to determine the final diagnosis of HF following criteria outlined by the ESC 2005 guidelines.[17] Cardiac dysfunction in addition to signs and symptoms of heart failure was reviewed by a panel, comprised one of four cardiologists, one of four pulmonologists, and an outpatient HF clinic physician, with final diagnosis and probable cause made by discussion and consensus.

Heart failure is classified as HFrEF if left ventricular systolic function was mild, moderate, or severe and classified as HFpEF if normal left ventricular systolic function was observed.

Statistical analysis

Prior to combining the datasets, studies were assessed for methodological quality using QUADAS-2.[18] Variable definitions were checked for agreement and summary statistics compared with published papers. UHFO data queries were confirmed with the study's research team.

Test evaluation

Baseline characteristics of the pooled dataset were summarized using descriptive statistics. Participants were grouped by diagnosis into those with and without HF and also by type of HF. Individual patient data (IPD) analysis initially explored the potential influence of age, sex, atrial fibrillation (AF), and body mass index (BMI) on the peptide's discriminatory ability to detect HF using mixed-effects multinomial logistic regression. The dependent variable of the latter being the 3-level outcome (HFrEF: HFpEF: no HF). Independent variables included NT-proBNP ($\log_e$ transformed) with age (10 year age-bands), sex, AF, and BMI as covariates and study as a random effect. Two-way interaction terms between covariates and NP were also considered. To simplify the model, the optimal age cut-point was then identified from assessment of the age-specific coefficients and model re-fitting using age as a dichotomous variable. Model comparison was undertaken with likelihood ratio tests. Covariate adjusted NP discriminatory ability was calculated by summing the resultant model's predicted probabilities estimating area under the receiver operating characteristic curve (AUROC) and associated confidence intervals.

Following the results of the modelling, 2 by 2 tables of true positive, false positive, true negative, and false negative counts were constructed for each study, at selected thresholds of interest: 100, 125 (ESC), 150, 200, and 400 (NICE) ng/L (12, 15, 18, 24, and 47 pmol/L) and then combined using two-stage bivariate random effects modelling.[19,20] This method estimates the summary negative predictive value (NPV), positive predictive value (PPV), sensitivity, and specificity for each of the NT-proBNP thresholds for HF diagnosis. Due to data sparsity, univariate random-effects meta-analysis was performed to estimate these test characteristics for diagnosis of HFpEF and HFrEF and also for each age and sex subgroups,[21] with summary predictive values obtained by applying the median prevalence to the summary sensitivity and specificity using Bayes rule.[20] Subgroup analysis by AF was not undertaken due to an insufficient number of cases with AF in the REFER dataset.

Test evaluation was performed excluding missing NP data (complete cases). To test the robustness of results, sensitivity analyses examined the effects of missing data using best case (assigning true positive/true negative based on reference test result), worst case (assigning false positive/false negative), and multiple imputation (10 datasets) methods. The imputation assumption of missing at random was explored by comparison of patients' characteristics, with and without NT-proBNP recorded, using mixed logistic regression analysis.

Statistical analysis was undertaken using Stata 15.0 (gsem, ice, xtmelogit, and metaprop_one).

Results

The pooled dataset consisted of 1073 participants (*Table 1*). Overall, 37.1% were male, with mean age of 71.7 years (SD 11.1), 630 (58.7%) were hypertensive, 213 (19.8%) had diabetes, and 84 (7.8%) a history of myocardial infarction. At presentation, 610 (56.8%) had ankle oedema, 930 (86.7%) shortness of breath, and 759 (70.7%) fatigue. Some differences were observed between the individual studies with REFER participants being older, more symptomatic, and having higher prevalence of diabetes and hypertension (Supporting Information, *Table S1*).

New onset HF was confirmed in 313 (29.2%) of the pooled dataset, with a similar prevalence of HF recorded in each study [106 (30.1%) in REFER vs. 207 (28.7%) in UHFO]. In total, 48.2% of all HF cases had reduced ejection fraction (HFrEF). Within each study, 15.1% of HF cases in REFER and 65.2% of HF cases in UHFO were due to HFrEF. Atrial fibrillation was observed in a similar percentage of HF cases in each dataset (20.8% in REFER vs. 27.1% in UHFO). Direct

Table 1 Characteristics of the combined dataset

Characteristic	Combined studies *N* = 1073 *N* (%)
Age (years)	
Mean (SD)	71.7 (11.1)
Range	19 to 92
<70	408 (38.0)
≥70	665 (62.0)
Male	398 (37.1)
Ankle oedema	610 (56.8)
Breathlessness	930 (86.7)
Lethargy	759 (70.7)
Previous myocardial infarction	84 (7.8)
Basal crepitations	87 (8.1)
Hypertension	630 (58.7)
Diabetes	213 (19.8)
Respiratory disease	249 (23.2)
Medications	
ACE inhibitors	281 (26.2)
Beta-blockers	265 (24.7)
ARBs	145 (13.5)
Diuretics	524 (48.8)
NT-proBNP (ng/L)	
Mean (SD)	292.4 (729)
Median (IQR)	60 (13 to 230)
Missing	115 (10.7)

comparisons between other abnormalities found on ECG and echocardiography could not be made due to coding variation of these items (*Table 2*). Persons diagnosed with HFpEF were slightly older than those with HFrEF (76.3 vs. 74.9 years), and HFpEF was more common in women (55.4% vs. 47.1%).

N-terminal pro-B-type natriuretic peptide was log-normally distributed, and summary statistics are illustrated by study and type of HF in *Figure 1*. Median (IQR) NT-proBNP was 267.5 ng/L (63.5 to 842.5) for those with HFrEF, 206 ng/L (46 to 729) with HFpEF, and 30 ng/L (10 to 113) for those without a HF diagnosis.

Missing N-terminal pro-B-type natriuretic peptide

N-terminal pro-B-type natriuretic peptide was missing for 115 (10.7%) participants. No significant differences were observed between the characteristics of those participants with or without natriuretic peptide recorded (Supporting Information, *Table S2*).

Overall discriminatory ability of N-terminal pro-B-type natriuretic peptide

The mixed effects modelling suggests that the age, sex, and AF adjusted relative risk of HF doubles per unit increase in $\log_e$ NT-proBNP [HFpEF vs. no HF, relative risk ratio (RRR) = 1.81, 95% confidence interval (CI) (1.58 to 2.08); HFrEF vs. no HF, RRR = 2.11 (1.82 to 2.43)]. For example, the relative risk of HFrEF increases from 0.156 to 0.315 with a rise in NTproBNP from 150 to 400 ng/L (equivalent to one unit change in $\log_e$ NT-proBNP), when other covariates are held constant. (Supporting Information, *Table S3*). There was no evidence of BMI having a moderating effect on the association of NT-proBNP with HFpEF (P = 0.36) or with HFrEF (P = 0.68); therefore, BMI was excluded from the final model. The discriminatory power of the model was better for HFrEF than HFpEF [AUROC 0.82 (0.78 to 0.86) vs. 0.71 (0.66 to 0.75)]. Combining the probabilities of HFrEF and HFpEF gave an AUROC of 0.80 (0.77 to 0.83) for HF overall. The associated ROC curves are provided in Supporting Information, *Figure S1*. Similar conclusions were made from the sensitivity analysis when missing NT-proBNP were imputed (Supporting Information, *Table S3*).

Bivariate meta-analyses of heart failure

Applying the ESC guideline threshold of 125 ng/L for diagnosis of chronic HF, the combined dataset had summary NPV and PPV of 85.1% (80.5 to 88.9) and 56.2% (30.5 to 78.9), respectively, with summary sensitivity and specificity of 71.5% (46.0 to 88.1) and 74.7% (29.3 to 95.4). The NICE guideline threshold of 400 ng/L gave a summary NPV and PPV of 79.9% (75.6 to 83.7) and 68.3% (46.3 to 84.4), respectively; and sensitivity and specificity of 43.5% (24.7 to 64.4) and 92.0% (70.7 to 98.2).

Univariate meta-analyses of heart failure by age group

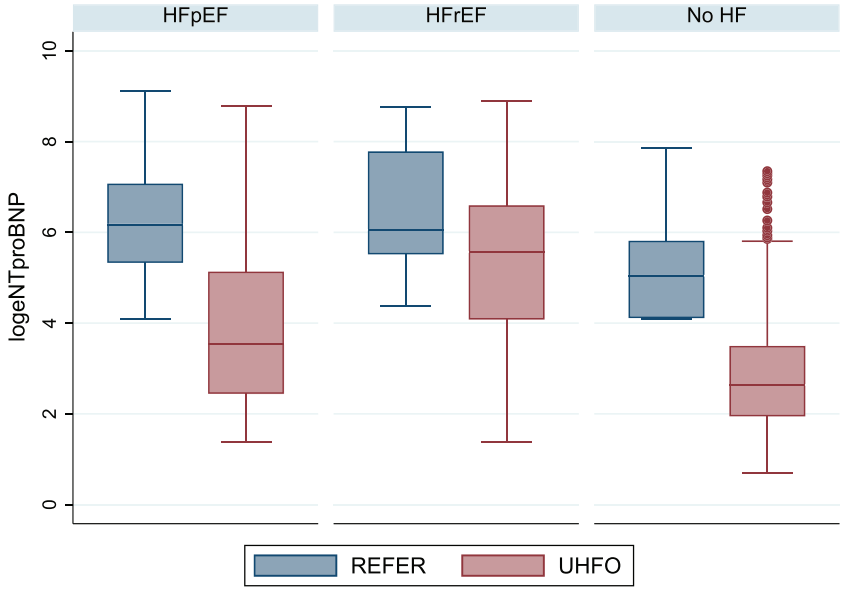
Test accuracy estimates for participants aged 70 years and over are more reliable (Supporting Information, *Table S4*), where the diagnostic accuracy of the ESC threshold has summary NPV of 84.9% (81.6 to 88.2), summary PPV of 68.1% (63.1 to 73.3), summary sensitivity of 74.9% (69.5 to 80.3), and summary specificity of 80.1% (76.9 to 83.4); and NICE threshold has summary NPV of 74.7% (72.1 to 77.2), summary PPV of 81.8% (73.3 to 89.5), summary sensitivity of 43.5% (37.2 to 49.8), and summary specificity of 94.5% (92.3 to 96.7). Summary statistics for PPV and sensitivity in persons aged below 70 have wide confidence intervals due to the heterogeneity observed and low sample size with summary NPV of 89.8% (87.1 to 92.5), summary PPV of 58.0% (44.3 to 72.0), summary sensitivity of 54.0% (41.9 to 66.0), and summary specificity of 91.2% (88.1 to 94.2); and NICE threshold

Table 2 Objective abnormalities found on ECG and echo in participants with and without heart failure

	REFER		UHFO	
Abnormality	Heart failure[a] N = 106 N (%)	No heart failure N = 246 N (%)	Heart failure N = 207 N (%)	No heart failure N = 514 N (%)
Moderate to severe left ventricular systolic dysfunction (LVSD)—ejection fraction <40%	4 (3.8)	0 (0)	81/203 (39.9)[b]	3/503 (0.6)[b]
Borderline LVSD—ejection fraction 41–50%	12 (11.3)	1 (0.4)	50/203 (24.6)[b]	40/503 (7.9)[b]
Diastolic dysfunction	22 (20.8)	4 (1.6)	96/152 (63.2)	86/462 (18.6)
Significant valve disease	40 (37.7)	34 (13.8)	108 (52.2)[c]	321 (62.5)[c]
Atrial fibrillation	22 (20.8)	11(4.5)	56 (27.1)	30 (5.8)
Unknown			18	

Some patients had >1 abnormality.
[a]Diagnosis at step 2.
[b]LVEF assessed semi-quantitatively and classified as normal/mild/moderate/severe dysfunction.
[c]Tricuspid/mitral insufficiency.

Figure 1 Distribution of $\log_e$ NT-proBNP by diagnosis and study.



with summary NPV of 85.1% (83.1 to 87.1), summary PPV of 68.6% (42.2 to 90.7), summary sensitivity of 24.2% (13.9 to 34.6), and summary specificity of 97.5% (95.7 to 99.2).

Univariate meta-analyses of heart failure by sex

The magnitude of differences between performance characteristics across sexes are smaller than observed previously for age (Supporting Information, *Table S5*). The ESC threshold has summary NPV and PPV of 84.7% (80.5 to 88.9) and 72.3% (65.0 to 79.7), respectively, for men and summary NPV and PPV of 88.2% (85.7 to 90.6) and 63.6% (56.2 to 71.2) for women. The NICE threshold has summary NPV and PPV of 76.6% (73.3 to 80.1) and 84.4% (74.4 to 93.3), respectively, for men and summary NPV and PPV of 80.5% (78.7 to 82.5) and 74.4% (60.8 to 86.9) for women.

Univariate meta-analyses of heart failure with preserved ejection fraction and heart failure with reduced ejection fraction

Forest plots of sensitivity and specificity (*Figures 2* and *3*) illustrate the differences between the performance characteristics of each NP threshold for diagnosing HFpEF and HFrEF. The estimates of specificity are consistent across the studies in both HF types with summary NPVs ranging from 93% at 100 ng/L to 86% at 400 ng/L for HFpEF and 95% to 92% for HFrEF. Sensitivity is more consistent across the studies in the HFrEF plot with lower thresholds providing increased detection—with summary sensitivity ranging from 74% to 38% and summary specificity from 74% to 93%.

Univariate meta-analyses of heart failure with preserved ejection fraction and heart failure with reduced ejection fraction by age

Evidence of heterogeneity was also observed in meta-analyses of HFpEF and HFrEF by age group. Older persons with HFpEF were better detected (Supporting Information, *Figures S2* and *S3*) with summary sensitivity ranging from 83% at 100 ng/L to 42% at 400 ng/L and summary specificity from 57% to 84%. Test characteristics for those aged under 70 were based on small numbers resulting in wide confidence intervals and should be viewed with caution. Threshold evaluation for HF type was not examined by sex due to sparsity of data in the 2 by 2 tables.

Discussion

Summary of findings

We found that NT-proBNP is better at detecting HFrEF than HFpEF in a primary care setting. Furthermore, in persons aged 70 and over, the ESC threshold of 125 ng/L is more sensitive at detecting HF than the higher level suggested in NICE guidelines. For example, based on our results, in a population of 1000 adults aged over 70 years in which 36% have HF, the lower ESC threshold (NPV = 84.9%) would rule out HF in 92

Figure 2 Sensitivity and specificity plots at various thresholds of NT-proBNP to detect HFpEF.

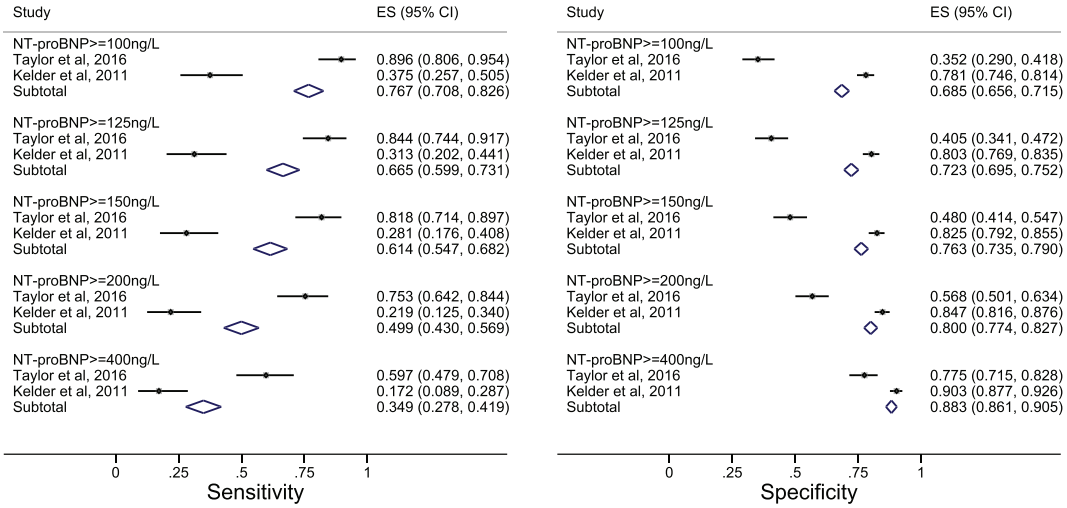
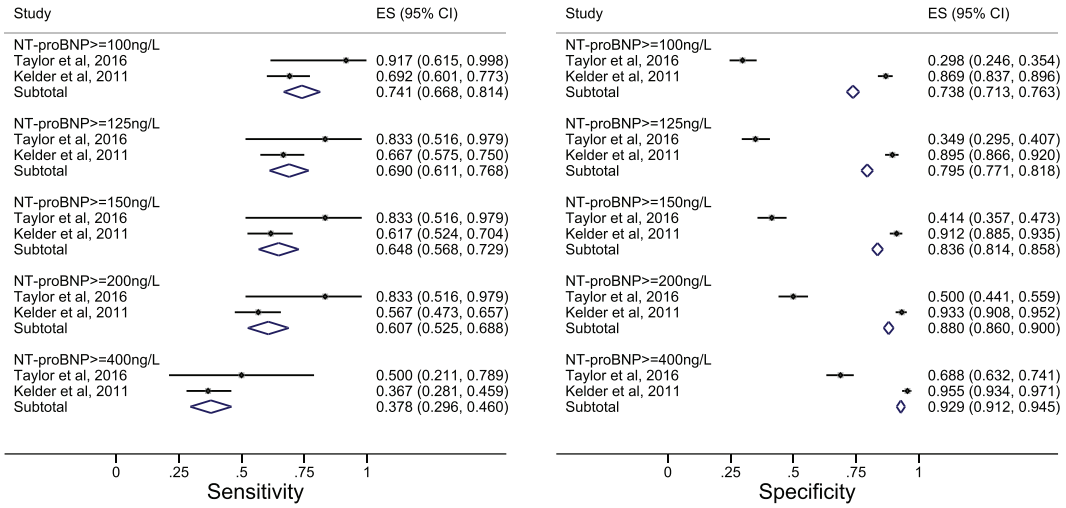


Figure 3 Sensitivity and specificity plots at various thresholds of NT-proBNP to detect HFrEF.



more cases than the upper NICE threshold (NPV = 74.7%) but, more importantly, miss 113 fewer HF cases. Identification of the optimal NP threshold is, in general, a careful balance of sensitivity and specificity. NPV and sensitivity of the test are the most important indicators of test evaluation to ensure cases are not missed; however, in resource-limited healthcare systems, the cost of increased unnecessary referrals and echocardiography due to poorer specificity is also a significant consideration.

Strengths and limitations

The main strength of this study is the ability to combine the two largest prospective contemporary primary care studies of suspected heart failure in two similar health care systems and perform an IPD meta-analysis—enabling us to explore the relationship between NP and different types of HF, adjusting for differences in age, sex, and presence of AF and also accounting for clustering of patients within each study. The main limitation of combining these studies is the heterogeneity observed between them, with the REFER cohort having more symptoms and co-morbidities and more HFpEF phenotype. This may be partially explained by REFER participants being older, with 70% aged 70 years and over compared with 58% in the UHFO cohort. Indeed, the REFER investigators have suggested the lower than expected rates of HFrEF were due to some differential recruitment due to the new availability of natriuretic peptide assays to the NHS for suspected HFrEF shortly before the study started.

In contrast, in the Dutch study, the proportion of HFpEF may have been somewhat underestimated because echocardiographic assessments (and definitions of abnormalities) may have differed between the eight hospitals. Albeit that this study population roughly included 50% of each heart failure phenotype, which is close to that observed in routine practice.

Age also explained some of the differences observed in the NT-proBNP distribution between the groups. We have addressed this by adjustment for age in the mixed modelling and also by exploring diagnostic test accuracy by age group. Statistical methods of diagnostic test accuracy within age groups were however restricted by the smaller sample size of those aged under 70. This resulted in convergence problems with the bivariate method; hence, age-specific performance estimates were based on a univariate method and therefore unadjusted for the correlation between sensitivity and specificity, resulting in narrower confidence intervals. We were also unable to implement the usual meta-regression methods suitable for subgroup analyses or a method more applicable to multiple thresholds of multiple studies[22] as there were an insufficient number of studies combined. We have however applied appropriate pooling methods recommended for sparse data[21] where possible, but acknowledge results should be interpreted with caution.

N-terminal pro-B-type natriuretic peptide was also measured differently in the two studies, with the REFER study using a POC test and UHFO a laboratory assay. We were unable to directly compare the two methods, but agreement between these methods has been reported previously,[23,24] with mean bias ranging from −10% to +17%.[23] In addition, diagnostic equivalence of the methods in the diagnosis of HF, using cut-off 125 ng/L, has been reported with sensitivity of 89% for both POC and NT-proBNP and AUCs of 0.88 and 0.89, respectively.[24]

The demographic characteristics of the combined cohort reflect those of the population of interest; however, we were unable to confirm its applicability with respect to ethnic mix because ethnicity was not recorded in the UHFO study.

Despite these limitations, we chose to pool the data as the studies were similar in design; recruited participants using comparable methods; used similar standardized assessment and included the same, universally accepted, reference standard of consensus panel for final diagnosis, blinded to results of NT-proBNP. The size of the study is a major strength and also its representativeness of typical patients in primary care presenting to their GP with symptoms suggestive of HF, differing from the mainly acute dyspnoea symptoms observed in an emergency care setting. The studies also had similar overall prevalence of new HF detected which is comparable to previously reported estimates in primary care.[9,11] The majority of variables collected had a high completion rate and, although 11% of NT-proBNP values were missing, the characteristics of participants with NP recorded were found to be similar to those with missing NP. In addition, the imputation analyses confirmed the robustness of our results. Given the paucity of test evaluation data available to identify optimal cut-offs for NP in a primary care setting, this study provides some of the best limited evidence from IPD meta-analysis to date.

Comparison with previous studies

There are currently no established thresholds discriminating between HFpEF and HFrEF. Current guidelines are based on detection of HF regardless of type. Our results show that NT-proBNP is better at detecting HFrEF than HFpEF. This finding is not surprising given previous studies have found that around a third of patients with HFpEF have BNP values <100 ng/L.[25]

Age-dependent cut-offs for detection of systolic dysfunction have been suggested by others.[26] More recently, a systematic review identified an optimal threshold of 250 ng/L for persons aged over 75, with NPV of 92%, sensitivity of 87.9%, and specificity 53.7%.[27] This review also concluded that the 125 ng/L threshold demonstrated higher sensitivity than the 400 ng/L threshold across all age categories evaluated (<50, 50–75, and >75 years).

N-terminal pro-B-type natriuretic peptide concentrations have previously been shown to be higher in women,[26,28] in all types of HF,[29] although the magnitude of this difference may not be large enough to warrant sex-specific thresholds.[27] Our results similarly indicated smaller differences in performance between sex-specific thresholds compared with differences in age-specific thresholds for HF. We were, however, unable to explore these effects by HF type, due to insufficient data. We were also unable to explore AF specific cut-offs for this reason.

Body mass index has been shown to lower NT-proBNP[30]; however, this moderating effect was not observed in our IPD analysis.

Implications for clinical practice

Current NICE guidelines warn that high levels of serum NPs can have causes other than HF, including being aged over 70. Our research suggests that the lower ESC threshold of 125 ng/L is better at detecting and excluding this condition in the elderly than NICE recommended threshold of 400 ng/L. However, this will have implications because its modest specificity and positive predictive value will increase the number of patients sent for echocardiography and could overwhelm an already stretched system.

Implications for future research

A prospective diagnostic accuracy study with sufficient number of participants to enable calculation of optimal NP thresholds, with adequate precision, is required to improve identification of all types of chronic HF in primary care.

Conclusions

There is evidence after adjustment for age, sex, and presence of AF that NT-proBNP is better at detecting HFrEF than HFpEF in primary care. In persons aged 70 and over, the lower ESC threshold of 125 ng/L provides greater sensitivity to detect and rule out HF than higher level of 400 ng/L suggested in current NICE guidelines for chronic HF. More prospective data are required to establish the optimal NP threshold in detecting HF in general practice.

Acknowledgements

This research was supported by the NIHR Oxford BRC. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care. This work uses data provided by patients and collected by the NHS as part of their care and support and would not have been possible without access to these data. The NIHR recognizes and values the role of patient data, securely accessed and stored, both in underpinning and leading to improvements in research and care.

Conflict of interest

C.J.T. reports speaker fees from Vifor and Novartis and non-financial support from Roche outside the submitted work. F.D.R.H. reports speaker fees and sponsorship from Novartis, Bayer, and Boehringer Ingelheim, and grants from Pfizer outside the submitted work. A.K.R., A.W.H., and J.C.K. have no conflicts of interest.

Funding

A.K.R. is supported by the NIHR Oxford Biomedical Research Centre (BRC). C.J.T. is funded by a National Institute for Health Research (NIHR) Academic Clinical Lectureship. F.D.R.H. acknowledges support from the NIHR School for Primary Care Research, NIHR ARC Oxford Thames Valley, and NIHR Oxford BRC.

Authors' contributions

A.K.R. drafted the manuscript and undertook the statistical analysis. F.D.R.H. conceived the idea and planned the study with A.W.H. C.J.T. and J.C.K. provided the data of the two individual studies and enabled the meta-analysis of the data. All authors contributed to the writing and approved the final manuscript.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Baseline characteristics of REFER and UHFO studies.
Table S2. Comparison of participants with and without NT-proBNP recorded.
Table S3. Mixed multinomial logistic regression models for complete case and imputation of NT-proBNP.
Table S4. Performance characteristics of selected thresholds of NT-proBNP in detecting chronic heart failure by age group.
Table S5. Performance characteristics of selected thresholds of NT-proBNP in detecting chronic heart failure by sex.
Figure S1. ROC curves of the multinomial logistic model for diagnosis of Heart Failure.
Figure S2. Paired sensitivity and specificity plots at selected thresholds of NT-proBNP for diagnosis of HFpEF in participants aged under 70 years.
Figure S3. Paired sensitivity and specificity plots at selected thresholds of NT-proBNP for diagnosis of HFpEF in participants aged 70 years and over.

References

1. Conrad N, Judge A, Tran J, Mohseni H, Hedgecott D, Crespillo AP, Allison M, Hemingway H, Cleland JG, McMurray JJV, Rahimi K. Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. *Lancet* 2018; **391**: 572–580.
2. Taylor CJ, Ryan R, Nichols L, Gale N, Hobbs FR, Marshall T. Survival following a diagnosis of heart failure in primary care. *Fam Pract* 2017; **34**: 161–168.
3. Taylor CJ, Ordóñez-Mena JM, Roalfe AK, Lay-Flurrie S, Jones NR, Marshall T, Hobbs FDR. Trends in survival after a diagnosis of heart failure in the United Kingdom 2000-2017: population based cohort study. *BMJ* 2019; **364**: l223.
4. Bottle A, Kim D, Aylin P, Cowie MR, Majeed A, Hayhoe B. Routes to diagnosis of heart failure: observational study

using linked data in England. *Heart* 2018; **104**: 600–605.
5. Monahan M, Barton P, Taylor CJ, Hobbs FDR, Roalfe AK. MICE or NICE? An economic evaluation of clinical decision rules in the diagnosis of heart failure in primary care. *Int J Cardiol* 2017; **241**: 255–261.
6. National Institute for Health and Care Excellence. Chronic heart failure in adults: diagnosis and management (NICE guideline NG106) 2018. https://www.nice.org.uk/guidance/ng106 (accessed 14 March 2021).
7. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, Falk V, González-Juanatey JR, Harjola VP, Jankowska EA, Jessup M, Linde C, Nihoyannopoulos P, Parissis JT, Pieske B, Riley JP, Rosano GM, Ruilope LM, Ruschitzka F. Rutten FH van der MP. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur J Heart Fail* 2016; **18**: 891–975.
8. Taylor KS, Verbakel JY, Feakins BG, Price CP, Perera R, Bankhead C, Pluddemann A. Diagnostic accuracy of point-of-care natriuretic peptide testing for chronic heart failure in ambulatory care: systematic review and meta-analysis. *BMJ* 2018; **361**: k1450.
9. Zaphiriou A, Robb S, Murray-Thomas T, Mendez G, Fox K, McDonagh T, Hardman SMC, Dargie HJ, Cowie MR. The diagnostic accuracy of plasma BNP and NTproBNP in patients referred from primary care with suspected heart failure: results of the UK natriuretic peptide study. *Eur J Heart Fail* 2005; **7**: 537–534.
10. Verdú JM, Comin-Colet J, Domingo M, Lupon J, Gomez M, Molina L, Casacuberta JM, Munoz MA, Mena A, Bruguera-Cortada J. Rapid point-of-care NT-proBNP optimal cut-off point for heart failure diagnosis in primary care. *Rev Esp Cardiol* 2012; **65**: 613–619.
11. Cowie MR, Struthers AD, Wood DA, Coats AJ, Thompson SG, Poole-Wilson PA, Sutton GC. Value of natriuretic peptides in assessment of patients with possible new heart failure in primary care. *Lancet* 1997; **350**: 1349–1353.
12. Taylor CJ, Roalfe AK, Iles R, Hobbs FR, REFER investigators, Barton P, Deeks J, McCahon D, Cowie MR, Sutton G, Davis RC, Mant J, McDonagh T, Tait L. Primary care REFerral for EchocaRdiogram (REFER) in heart failure: a diagnostic accuracy study. *Br J Gen Pract* 2017; **67**: e94–e102.
13. Kelder JC, Cramer MJ, van Wijngaarden J, van Tooren R, Mosterd A, Moons KGM, Lammers JW, Cowie MR, Grobbee DE, Hoes AW. The diagnostic value of physical examination and additional testing in primary care patients with suspected heart failure. *Circulation* 2011; **124**: 2865–2873.
14. Tait L, Roalfe AK, Mant J, Cowie M, Deeks JJ, Iles R, Taylor C, Derit M, Hobbs FDR. The REFER (REFer for EchocaRdiogram) protocol: a prospective validation of a clinical decision rule, NT-proBNP, or their combination, in the diagnosis of heart failure in primary care. *Rational and design BMC Cardiovasc Disord* 2012; **12**: 97.
15. Roalfe AK, Mant J, Doust J, Barton P, Cowie MR, Glasziou P, Mant D, McManus R, Holder R, Deeks JJ, Doughty RN, Hoes AW, Fletcher K, Hobbs FDR. Development and initial validation of a simple clinical decision tool to predict the presence of heart failure in primary care: the MICE (male, infarction, crepitations, edema, MICE) rule. *Eur J Heart Fail* 2014; **14**: 1000–1008.
16. IFCC Committee on Clinical Applications of Cardiac Bio-Markers (C-CB). BNP, NT-proBNP and MR-proANP assays: analytical characteristics designated by manufacturer. http://bnp-nt-probnp-and-mr-proanp-assays-analytical-characteristics-designated-by-manufacturer-v08232018.pdf (ifcc.org) (6 December 2020)
17. Swedberg K, Cleland J, Dargie H, Drexler H, Follath F, Komajda M, Tavazzi L, Smiseth OA, Gavazzi A, Haverich A, Hoes A, Jaarsma T, Korewicki J, Levy S, Linde C, Lopez-Sendon JL, Nieminen MS, Pierard L, Remme WJ. Guidelines for the diagnosis and treatment of chronic heart failure: executive summary (update 2005): The Task Force for the Diagnosis and Treatment of Chronic Heart Failure of the European Society of Cardiology. *Eur Heart J* 2005; **26**: 1115–1140.
18. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; **155**: 529–536.
19. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clinical Epidemiol* 2005; **58**: 982–990.
20. Leeflang MG, Deeks JJ, Rutjes AWS, Reitsma JB, Bossuyt PMM. Bivariate meta-analysis of predictive values of diagnostic tests can be an alternative to bivariate meta-analysis of sensitivity and specificity. *J Clinical Epidemiol* 2012; **65**: 1088–1097.
21. Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Stat Methods Med Res* 2015; **26**: 1896–1911.
22. Steinhauser S, Schumacher M, Rucker G. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. *BMC Med Res Methodol* 2016; **16**: 97.
23. Jorgensen B, Bertsch T, Bröker H-J, Schafer M, Chapelle J-P, Gadisseur R, Cowie M, Dikkeschei LD, Gurr E, Hayen W, Hafner G, Stiegler Y, Schonherr H-R, Strasser RH, Weidtimann B, Folkerts H, Zugck C, Hofmann K, Zerback R. Multicentre evaluation of a second generation point-of-care assay with an extended range for the determination of N-terminal pro-brain natriuretic peptide. *Clin Lab* 2012; **58**: 515–525.
24. Schafer M, Broker H-J, Luchner A, Jungbauer C, Zugck C, Mitrovic V, Willenbrock R, Robert R, Flieger RR, Wittmer BA, Graves MW, Fluder K, Zerback R. Diagnostic equivalence of an n-terminal pro-brain natriuretic peptide point-of-care test to the laboratory method in patients with heart failure and in reference populations. *Point of Care* 2010; **9**: 91–97.
25. Anjan VJ, Loftus TM, Burke MA, Akhter N, Fonarow GC, Gheorghiade M, Shah SJ. Prevalence, clinical phenotype, and outcomes associated with normal B-type natriuretic peptide levels in heart failure with preserved ejection fraction. *Am J Cardiol* 2012; **110**: 870–876.
26. Raymond I, Groenning BA, Hildebrandt PR, Nilsson JC, Baumann M, Trawinski J, Pedersen F. The influence of age, sex and other variables on the plasma level of N-terminal pro brain natriuretic peptide in a large sample of the general population. *Heart* 2003; **89**: 745–751.
27. Hildebrandt P, Collinson P, Doughty R, Fuat A, Gaze DC, Gustafsson F, Januzzi J, Rosenberg J, Senior R, Richards M. Age-dependent values of N-terminal pro-B-type natriuretic peptide are superior to a single cutpoint for ruling out suspected systolic dysfunction in primary care. *Eur Heart J* 2010; **31**: 1881–1889.
28. Hamada M, Shigematsu Y, Takezaki M, Ikeda S, Ogimoto A. Plasma levels of atrial and brain natriuretic peptides in apparently healthy subjects: effects of sex, age and haemoglobin concentration. *Int J Cardiol* 2017; **228**: 599–604.
29. Faxen UL, Lund LH, Orsini N, Stromberg A, Andersson DC, Linde C, Dahlstrom U, Savarese G. N-Terminal pro-B-type natriuretic peptide in chronic heart failure: the impact of sex across the ejection fraction spectrum. *Int J Cardiol* 2019; **287**: 66–72.
30. Bayes-Genis A, Lloyd-Jones DM, van Kimmenade RRJ, Lainchbury JG, Richards AM, Ordonez-Llanos J, Santalo M, Pinto YM, Januzzi JL. Effect of body mass index on diagnostic and prognostic usefulness of amino-terminal pro-brain natriuretic peptide in patients with acute dyspnea. *Arch Intern Med* 2007; **167**: 400–407.