Efficient selective screening for heart failure in elderly men and women from the community: A diagnostic individual participant data meta-analysis

Rogier F Kievit[1], Aisha Gohar[1,2], Arno W Hoes[1],
Michiel L Bots[1], Evelien ES van Riet[1], Yvonne van Mourik[1],
Loes CM Bertens[1], Leandra JM Boonman-de Winter[1,3],
Hester M den Ruijter[2] and Frans H Rutten[1];
on behalf of the Queen of Hearts and RECONNECT
consortium

Abstract

Background: Prevalence of undetected heart failure in older individuals is high in the community, with patients being at increased risk of morbidity and mortality due to the chronic and progressive nature of this complex syndrome. An essential, yet currently unavailable, strategy to pre-select candidates eligible for echocardiography to confirm or exclude heart failure would identify patients earlier, enable targeted interventions and prevent disease progression. The aim of this study was therefore to develop and validate such a model that can be implemented clinically.

Methods and results: Individual patient data from four primary care screening studies were analysed. From 1941 participants >60 years old, 462 were diagnosed with heart failure, according to criteria of the European Society of Cardiology heart failure guidelines. Prediction models were developed in each cohort followed by cross-validation, omitting each of the four cohorts in turn. The model consisted of five independent predictors; age, history of ischaemic heart disease, exercise-related shortness of breath, body mass index and a laterally displaced/broadened apex beat, with no significant interaction with sex. The c-statistic ranged from 0.70 (95% confidence interval (CI) 0.64–0.76) to 0.82 (95% CI 0.78–0.87) at cross-validation and the calibration was reasonable with Observed/Expected ratios ranging from 0.86 to 1.15. The clinical model improved with the addition of N-terminal pro B-type natriuretic peptide with the c-statistic increasing from 0.76 (95% CI 0.70–0.81) to 0.89 (95% CI 0.86–0.92) at cross-validation.

Conclusion: Easily obtainable patient characteristics can select older men and women from the community who are candidates for echocardiography to confirm or refute heart failure.

Introduction

Heart failure, a chronic and progressive syndrome, is highly prevalent amongst older people and is a leading cause of premature death and disability.[1] Early diagnosis of heart failure is crucial as prompt initiation of treatment can prevent or slow down further progression, improve quality of life and reduce mortality risk.[2,3] Studies, however, have shown that in older individuals in the community, especially those with

[1]Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht University, The Netherlands
[2]Department of Experimental Cardiology, University Medical Centre Utrecht, Utrecht University, The Netherlands
[3]Amphia Academy, Amphia, Breda, The Netherlands

Corresponding author:
Aisha Gohar, Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, PO Box 85500, 3508 AB Utrecht, The Netherlands.
Email: a.gohar-3@umcutrecht.nl

comorbidities, unrecognised heart failure is common.[4–6] Heart failure in the community is a challenge to diagnose.[7,8] Patients, and also physicians, often consider slowly developing and gradual worsening of shortness of breath and reduction in exercise tolerance in older patients to be part of ordinary aging ('deconditioning').[9] Moreover, shortness of breath is often considered to be of pulmonary origin and underlying cardiac problems such as evolving heart failure can be overlooked.[10]

To improve the ability of the general practitioner (GP) to diagnose heart failure in such patients, a focused screening approach should be at the GP's disposal in order to select the patients at high-risk of having heart failure who are candidates for echocardiography to confirm or exclude the diagnosis of heart failure, as recommended by current guidelines.[3]

Previous diagnostic studies and systematic reviews have mostly focused on diagnosing heart failure in community-dwelling people suspected of slow-onset heart failure,[11–14] that is, patients presenting with suggestive signs and symptoms in primary care. There is a scarcity of studies focusing on the development of useful decision tools to screen for heart failure in high-risk primary care populations. The few available tools typically focus on specific patients groups (e.g. older people with type 2 diabetes mellitus or chronic obstructive pulmonary disease (COPD)). The production of multiple, differing models (partly overlapping) and uncertainty about the applicability in everyday clinical practice hinders implementation of these models. The availability of a screening tool applicable to the much larger group of all-type community-dwelling older patients would greatly facilitate screening activities. Combining the screening studies into a large individual patient database (IPD), in which a model can be both developed and validated with state-of-the-art methodology, is an attractive method to produce such a tool.

We therefore combined four available primary care screening studies that have previously developed prediction models for detecting heart failure in older people from the community into one IPD. We examined whether one prediction model could be developed which was able to identify older individuals at high-risk of having heart failure and therefore who subsequently require echocardiography to confirm the diagnosis.

Methods

Study population

Four previously published studies (STRETCH, UHFO-DM, UHFO-COPD and TREE) performed in the primary care setting were combined into one large IPD file (for a description of the four cohorts see Supplementary Material Table 1 online).[4–6,15] In these studies, specific community-dwelling high-risk patient groups were screened for previously unknown heart failure. The studies consisted of patients with either symptoms of shortness of breath on exertion, type 2 diabetes mellitus, COPD or 'frail' elderly, the last definition based on multimorbidity or polypharmacy (defined as using five or more prescribed drugs daily in the past year). The data in all studies were collected cross-sectionally and participants received investigations, including echocardiography, during a one-day assessment.

Outcome, diagnostic predictors and model development

In all four studies, the outcome heart failure (all-type) was established by an expert panel as described previously,[4–6,15] according to the heart failure criteria in the European Society of Cardiology (ESC) guidelines.[16] The panel consisted of at least three experts: a GP was always present, a pulmonologist was present on the UHFO-COPD and TREE panels and at least two cardiologists were present on the panels of all cohorts except for the TREE cohort. All available diagnostic items from the assessment, including echocardiography, all performed similarly in the four studies with applying the same case record form, were taken into account by the panel when deciding on the presence or absence of heart failure. Natriuretic peptide measurements were used as an inclusion criterion for echocardiography in the STRETCH cohort, applying a cut-off point of N-terminal pro B-type natriuretic peptide (NTproBNP) level above 125 pg/m (≈ 14.75 pmol/l). The panel also assessed NTproBNP levels in the TREE cohort prior to diagnosis. The panels were not privy to the NTproBNP levels in the UHFO-DM and UHFO-COPD cohorts, therefore preventing incorporation bias in only these two cohorts.[17] Left ventricular diastolic dysfunction was assessed non-invasively using echocardiography according to the ESC heart failure guidelines.[3]

We started with 23 potential diagnostic predictors known from the literature of diagnostic studies evaluating those suspected of heart failure from primary care[11–14,18–21] and from the four primary care screening studies. The potential predictors were demographics (age, sex), medical history (ischaemic heart disease (IHD), atrial fibrillation, COPD or asthma, hypertension, peripheral vascular disease, diabetes mellitus), symptoms (dyspnoea leading at least to stop at a normal pace (Medical Respiratory Council (MRC) questionnaire (MRC ≥ 3)), orthopnoea, paroxysmal nocturnal dyspnoea), signs (systolic and diastolic blood pressure, heart rate, irregularity of pulse, body

mass index (BMI), ankle oedema, pulmonary crepitations, raised jugular venous pressure, laterally displaced or broadened/sustained apex beat, hepatomegaly), NTproBNP and electrocardiogram (ECG).

Two prediction models were defined for evaluation: (1) a clinical model including items from history taking, symptoms and signs; and (2) an extended model comprising the diagnostic predictors from the clinical model plus NTproBNP and ECG abnormalities. To assess which of the candidate predictors was of value when predicting the presence of heart failure, we first included all candidate predictors in the model, and then with the use of multivariable logistic regression analyses reduced the model one by one backwards. For model selection, we used Akaike Information Criteria (AIC), which is rather similar to the more widely accepted likelihood ratio test, but is considered superior for model selection[22] as it additionally includes a penalty for the number of candidate predictors, thereby discouraging over fitting.

In all analyses a linear relationship between the outcome heart failure and the continuous predictors age and BMI was assumed and checked. There was no collinearity between variables. All analyses were performed in R version 3.1.2.

Measurements

Data were gathered using a standardised case record form, including information on demographics, medical history and symptoms. Medical history was cross-checked with the GPs' electronic medical records. All participants underwent a systematic physical examination including examination of the heart and lungs and for signs of fluid retention. The apex beat was palpated in the supine and lateral decubital position. An impalpable apex beat was defined as an 'undisplaced apex beat' in all four studies.

A history of IHD was defined as a previous myocardial infarction, coronary bypass grafting or percutaneous coronary intervention. The ECGs were classified according to the Minnesota coding criteria.[23] An ECG was considered abnormal if one of the following was present: atrial fibrillation, tachycardia (heart rate >100 beats/min), left or right bundle branch block, left ventricle hypertrophy, and ST and/or T-waves abnormalities. NTproBNP was measured with a non-competitive immune-radiometric assay (Roche, Mannheim, Germany) in all cohorts.

Missing values

A summary of the missing values is displayed in Supplementary Material Table 2. Multiple imputation techniques were used to impute five sets of data of each individual study following the MICE algorithm for R

software.[24] For the imputation models we used all the variables that we considered as candidate diagnostic predictors as well as the outcome heart failure.

Cross validation

The Internal-External Cross Validation (IECV) method was used for model development and validation, a state-of-the-art method for use with an IPD from multiple prediction studies.[25] To explain the method briefly, the model is developed in all of the studies except one and the performance of this developed model is assessed in the omitted study; that is, the validation study. A model is then developed in a different combination of studies omitting a different study from before and so on and so forth, until all of the studies have been omitted and used as the validation study.[26] For the development of the final prediction model, the predictors that performed the strongest in all developmental datasets were used, according to the AIC criteria. The intercept used within the IECV was the estimated intercept from one of the development studies that was most similar in heart failure prevalence to the omitted study.[26]

The performance of the models was quantified with discrimination and calibration. Discrimination is the ability of a model to distinguish between patients with an outcome (i.e. heart failure) and without an outcome (i.e. without heart failure), quantified with the c-statistic. Calibration is the agreement between observed outcome frequencies and predicted probabilities, examined with the Observed/Expected (OE) ratio and visually with calibration plots.

Risk score

A risk score was constructed from the final model after finalising all IECV steps by multiplying the shrunken coefficients of the final model by two and then rounding up to the nearest integer. To reflect the difference in prevalence and therefore baseline risk between the UHFO-COPD and STRETCH studies on the one hand and the UHFO-DM and TREE studies on the other, a dummy variable was added representing whether a participant came from one of the higher baseline risk studies. Logistic regression was subsequently used to calibrate the risk of heart failure according to the scores, which resulted in a corresponding risk of heart failure for every score which was then graphically presented. Score thresholds with associated performance of the scoring rule were given for seven risk categories.

Results

The characteristics of the 1941 participants from the four primary studies who underwent screening for

heart failure are shown in Table 1. Although the mean age and distribution of sex were comparable between the four primary studies, there was a spread in the prevalence of previously undetected heart failure (16% to 34%), and in the proportion of heart failure with preserved ejection fraction (HFpEF) (50% to 82% of those with heart failure). The TREE study, including frail elderly, had a high prevalence of comorbidities.

The intercepts and therefore the baseline risks were comparable between the STRETCH and UHFO-COPD studies and between the UHFO-DM and TREE studies. From the 23 candidate predictors, the predictors age, a history of IHD, dyspnoea (MRC ≥ 3), BMI, a laterally displaced or broadened/sustained apex beat, NTproBNP and an abnormal ECG were important predictors for the presence of heart failure in all datasets (Table 2).

The c-statistic of the final clinical model consisting of the five predictors (age, a history of IHD, dyspnoea (MRC ≥ 3), BMI, and a laterally displaced or broadened/sustained apex beat) ranged at cross-validation from 0.70 to 0.82 (Supplementary Material Table 3). NTproBNP had independent added value, improving the discrimination considerably (c-statistic ranging from 0.76 to 0.89) (Table 2). Adding ECG on top of the clinical model plus NTproBNP did not have any independent added value, with the c-statistic not changing substantially (c-statistic ranging from 0.76 to 0.90). The calibration of the final clinical model was good with OE ratios ranging from 0.86 to 1.15 (Supplementary Table 3) and as visualised with the calibration plots (Supplementary Figure 2). Adding NTproBNP did not influence the calibration much, with OE ratios ranging from 0.85 to 1.18.

The corresponding bootstrap corrected c-statistic of the score was 0.78 (95% confidence interval (CI) 0.75–0.80) for the final clinical model and 0.84 (95% CI 0.82–0.86) for the extended model (clinical + NTproBNP) (Table 2). As cardiovascular medications can affect symptoms, especially shortness of breath, we adjusted the full final model for the use of diuretics and angiotensin converting enzyme/angiotensin receptor blockers. However, no effects of the coefficients of the predictors were seen. Using the constructed risk scores to categorise individuals into different risk groups, the highest number of patients was categorised as moderate or high risk of having heart failure. Of these, 28% and 52.2% respectively actually had heart failure present (Tables 3 and 4). The higher the summed score and therefore the higher the probability of having heart failure, the higher the sensitivity and negative predictive value of the model.

There were no sex interactions of any of the predictors in the final model. A table showing estimates of the predictors in the final model for men and women separately can be viewed in Supplementary Table 4.

Discussion

We developed and validated a prediction model that can identify, among community dwelling elderly men and women at high-risk of having heart failure, who are candidates for echocardiography to confirm/refute diagnosis. An easy to use clinical model with five predictors – age, a history of IHD, dyspnoea (MRC ≥ 3), BMI, and a laterally displaced or broadened/sustained apex beat – performed the strongest. By adding NTproBNP to an extended model the performance improved even more.

Comparison with previous studies

In agreement with previous studies, age, a history of IHD, BMI, dyspnoea on exertion and a laterally displaced or broadened/sustained apex beat were important in assessing the probability of having heart failure. Two of the predictors, however, are not yet commonly used in clinical practice; first BMI, with obese people having an increased risk of unrecognised heart failure. We found that in all development datasets in our study, BMI was a strong predictor, in line with previous diagnostic study findings.[27] Therefore, contrary to current practice, clinicians should consider taking BMI into account when assessing the probability of heart failure. Secondly, we found that the predictive value of a laterally displaced or broadened/sustained apex beat was high with a mean odds ratio of approximately 2.50 (95% CI 1.73–3.62). Most previous prediction studies on heart failure did not include this sign as it was previously considered not to be useful.[12] However, studies that did include it have already shown the predictive value of this physical exam variable.[12–14,21–28] In addition to it having excellent diagnostic predictive value, it also forms part of the recommendations in the ESC guidelines.[3,16] When interpreting this finding, it is important to take into account that in around 50% of older adults the apex cannot be palpated,[28] and in these studies such cases were considered to have a normal apical beat. Irrespective of this 'shortcoming' it still has a very good predictive value. Another aspect, often not mentioned, is that it can be assessed in two ways: in the decubital position, when an apical impulse is palpated outside the mid-clavicular line, and in the left decubital position, when the impulse is broadened (two or more fingers) or sustained.[28] Given our results and previous findings highlighting the predictive value, clinicians should be encouraged to perform this examination in general practice, especially as it is readily available and relatively easy to perform.

Table 1. Patient characteristics of the individual patient database dataset composed of four primary care studies.

	STRETCH (N = 585)	TREE (N = 370)	UHFO-COPD (N = 405)	UHFO-DM (N = 581)	All combined (N = 1941)
Mean age, years (SD)	74.1 (6.3)	75.4 (6.1)	72.9 (5.3)	71.6 (7.4)	73.4 (6.6)
Female sex, %	54.5	55.4	44.9	46.6	50.3
New diagnosis of heart failure, %	15.7	35.2	20.5	27.7	23.9
New HFpEF (EF > 45%), %	12.0	25.4	10.1	22.9	17.4
New HFrEF (EF ≤ 45%), %	2.9	9.7	10.4	4.8	6.3
Medical history					
Ischaemic heart disease, %	11.5	24.1	9.9	11.2	13.4
COPD or asthma, %	55.2	26.8	100.0	12.2	46.3
Hypertension, %	53.0	72.4	35.8	65.6	56.9
Peripheral arterial disease, %	6.0	10.0	2.5	6.7	6.2
Diabetes mellitus, %	13.5	32.2	10.4	100.0	42.3
Atrial fibrillation, %	7.2	13.5	8.4	7.2	8.7
Symptoms					
Moderate to severe dyspnoea (MRC ≥ 3), %	16.9	38.6	50.1	13.3	26.9
Orthopnoea and/or PND, %	15.4	9.5	23.0	10.5	14.4
Swollen ankles at the end of the day, %	30.9	27.3	23.5	27.7	27.7
Physical examination					
Systolic blood pressure in mmHg, mean (SD)	147.0 (18)	139.0 (18)	152.0 (18)	159.0 (20)	150.0 (20)
Diastolic blood pressure in mmHg, mean (SD)	78.0 (11)	76.0 (9)	84.0 (10)	89.0 (10)	82.0 (11)
Mean heart rate, beats/min (SD)	73.7 (12.6)	69.6 (11.2)	76.5 (14.1)	70.1 (11.6)	72.0 (13)
Irregular pulse, %	10.4	16.8	13.1	4.6	10.5
Mean BMI, kg/m^2 (SD)	27.6 (4.4)	28.1 (4.4)	26.3 (4.1)	27.9 (4.5)	27.5 (4.4)
Pulmonary crepitations, %	18.6	8.1	16.4	9.5	13.4
Laterally displaced or broadened/sustained apex beat, %[a]	3.6	10.6	17.3	12.7	10.5
Elevated jugular venous pressure, %	8.7	9.5	10.9	3.4	7.7
Additional testing					
NTproBNP in pg/ml, median (IQR)	118.4 (67.7, 219.9)	138.7 (74.8, 294.4)	127.5 (76.2, 244.5)	76.1 (42.3, 152.2)	112.0 (59.2, 218.6)
Abnormal ECG, %[b]	57.2	61.5	48.9	37.9	50.3

[a]Laterally displaced or broadened/sustained apex beat was defined as an apex beat palpable outside the mid-clavicular line in the decubital position and/or a broadened and sustained apex beat in the left decubital position.
[b]An abnormal ECG was defined as: atrial fibrillation, sinus tachycardia (heart rate > 100 beats/min), a left or right bundle branch block, left ventricle hypertrophy, P-wave abnormalities compatible with left atrial enlargement, pathological Q-waves suspected for previous myocardial infarction or any ST-segment/T-wave abnormalities.
COPD: chronic obstructive pulmonary disease; ECG: electrocardiogram; HFpEF: heart failure with preserved ejection fraction; HFrEF: heart failure with reduced ejection fraction; IQR: interquartile range; MRC: Medical Respiratory Council; NTproBNP: N-terminal pro B-type natriuretic peptide; PND: paroxysmal nocturnal dyspnoea

The item 'dyspnoea leading to stop at a normal pace (MRC ≥ 3 or more)' seems more typical for selective screening studies than for diagnostic studies, that is, in studies evaluating people suspected of having heart failure. This is most likely because it is a well-known symptom that should always trigger physicians to consider heart failure, certainly when it is present in combination with a reduced exercise tolerance/fatigue and ankle oedema.[16]

As one of the initial 23 potential predictors, and despite female sex being highly prevalent in HFpEF,[29] sex did not form part of the final clinical model. In addition, the predictors making up the final model did not behave differently in men or women, meaning that this model performs equally well in both sexes. This is in line with previous studies publishing diagnostic models for detecting heart failure.[14,28]

We found that the natriuretic peptide NTproBNP had an independent predictive value beyond our final clinical model. The natriuretic peptides BNP and NTproBNP are recommended in clinical practice to

Table 2. Selection from the 23 candidate variables, which predicted the presence of heart failure according to the Akaike Information Criteria in strata of cohort combinations plus the c-statistics for the final model (clinical and NTproBNP).

Developed in:	TREE, UHFO-COPD and UHFO-DM			STRETCH, UHFO-COPD and UHFO-DM			STRETCH, TREE and UHFO-DM			STRETCH, TREE and UHFO-COPD			STRETCH, TREE, UHFO-COPD and UHFO-DM		
	Bᵇ (SE)	OR (95% CI)	c-statistic	Bᵇ (SE)	OR (95% CI)	c-statistic	Bᵇ (SE)	OR (95% CI)	c-statistic	Bᵇ (SE)	OR (95% CI)	c-statistic	Bᵇ (SE)	OR (95% CI)	c-statistic
			0.89 (0.86–0.92)			0.80 (0.75–0.85)			0.76 (0.70–0.81)			0.80 (0.76–0.84)			0.84 (0.82–0.86)
Intercept STRETCH[a]	–	–		-10.29	–		-12.9	–		-10.18	–		-10.69	–	
Intercept TREE[a]	-8.24	–		–	–		-11.84	–		-9.44	–		-9.86	–	
Intercept UHFO-COPD[a]	-8.99	–		-10.31	–		–	–		-10.19	–		-10.67	–	
Intercept UHFO-DM[a]	-7.95			-9.2	–		-11.61	–		–	–		-9.47	–	
Age Q1 ≈ 68 years, Q3 ≈ 78 years	0.56 (0.10)	1.73 (1.37–2.18)		0.55 (0.13)	1.95 (1.55–2.44)		0.62 (0.14)	2.04 (1.60–2.61)		0.57 (0.15)	1.82 (1.43–2.30)		0.60 (0.15)	1.91 (1.56–2.34)	
Ischaemic heart disease	0.63 (0.19)	1.96 (1.35-–2.83)		0.92 (0.20)	2.64 (1.79–3.90)		0.74 (0.19)	2.16 (1.49–3.13)		0.67 (0.20)	2.04 (1.40–2.97)		0.69 (0.17)	2.06 (1.48–2.86)	
Peripheral arterial disease	0.77 (0.26)	2.26 (1.37–3.74)		–	–		–	–		–	–		0.67 (0.23)	2.01 (1.28–3.16)	
Moderate to severe dyspnoea, MRC ≥ 3	0.91 (0.163)	2.65 (1.93–3.65)		1.08 (0.17)	3.14 (2.24–4.39)		1.16 (0.17)	3.35 (2.40–4.66)		0.92 (0.16)	2.65 (1.93–3.66)		0.96 (0.15)	2.72 (2.05–3.62)	
Paroxysmal nocturnal dyspnoea	–	–		–	–		0.87 (0.26)	2.49 (1.51–4.10)		–	–		0.67 (0.21)	2.00 (1.34–3.00)	
Oedema at the end of the day	–	–		–	–		0.69 (0.16)	2.06 (1.51–2.81)		–	–		–	–	
BMI Q1 ≈ 25, Q3 ≈ 30	0.9 (0.10)	1.69 (1.41–2.03)		0.55 (0.10)	1.90 (1.58–2.28)		0.56 (0.10)	1.82 (1.50–2.20)		0.51 (0.10)	1.84 (1.51–2.24)		0.56 (0.10)	1.85 (1.58–2.17)	
SBP Q1, 135, Q3, 162	–			–	–		0.28 (0.11)	1.35 (1.10–1.66)		–	–		–	–	
Pulmonary crepitations	0.64 (0.21)	1.98 (1.31–3.00)		–	–		0.59 (0.21)	1.85 (1.22–2.80)		–	–		0.50 (0.18)	1.68 (1.19–2.38)	
Laterally displaced or broadened/sustained apex beat	0.80 (0.20)	2.35 (1.60–3.44)		0.82 (0.21)	2.39 (1.59–3.60)		0.98 (0.23)	2.79 (1.77–4.41)		0.98 (0.24)	2.83 (1.77–4.52)		0.88 (0.19)	2.50 (1.73–3.62)	
NTproBNP in pg/ml Q1 ≈ 59, Q3 ≈ 220	0.00 (0.00)	1.23 (1.16–1.31)		0.00 (0.00)	1.20 (1.15–1.26)		0.00 (0.00)	1.44 (1.30–1.51)		0.00 (0.00)	1.33 (1.24–1.43)		0.00 (0.00)	1.28 (1.21–1.35)	

The value of the diagnostic predictor is 1 when present and 0 when absent.

Probability of heart failure can be estimated with the following formula: P(heart failure) = 1 / 1 + exp(-linear predictor), and the linear predictor can be calculated with the intercept and regression coefficient as presented in the table. For example, from the development datasets consisting of the TREE, UHFO-COPD and UHFO-DM study and validated in the STRETCH study, linear predictor = −8.99 + 0.05 * Age + 0.63 * IHD + 0.77 * PAD + 0.91 * Dyspnoea (MRC ≥ 3) + 0.09 * BMI + 0.64 * Pulmonary crepitations + 0.80 * Laterally displaced or broadened/sustained apex beat + 0.002 * NTproBNP + 0.46 * Abnormal ECG.

[a]The intercept from one of the individual patient database studies that is most similar to the new study population was chosen as the intercept.

[b]Regression coefficient multiplied by the shrinkage factor. The shrinkage factor is obtained by the heuristic formula as proposed by Van Houwelingen.[32]

BMI: body mass index; CI: confidence interval; MRC: Medical Respiratory Council; NTproBNP: N-terminal pro B-type natriuretic peptide; OR: odds ratio; Q: quartile; SBP: systolic blood pressure

Table 3. Clinical scoring rule (a) without NTproBNP and (b) with NTproBNP.

Rule score: summation of points	Points
(a)	
Age per 10 years	2
History of ischaemic heart disease	2
Moderate to severe dyspnoea (MRC $\geq$ 3)	2
BMI per 5 kg/m^2	1
Laterally displaced or broadened/sustained apex beat	2
High-risk because of type 2 diabetes mellitus	2
High-risk because of multimorbidity and polypharmacy[a]	2

The probability of heart failure outcome is defined as 1 / (1 + (exp(-LP)), where LP refers to the linear predictor in a logistic regression model. The LP for the clinical score is defined as follows: LP = − 11.83 + 0.47 * total sum of the score.

Use of the clinical scoring rule: for example, a 70-year-old person (14 points), without a history of ischaemic heart disease, who stops for breath after walking a few minutes on level ground (MRC dyspnoea score 4) (two points), has a BMI of 30 kg/m^2 (six points), and no laterally displaced or broadened/sustained apex beat has a score of 22 points. According to Supplementary Figure 1(a) this score corresponds to a risk of heart failure of approximately 20%. According to Table 4, if a general practitioner decided that all individuals with a probability of 20% or less are not be referred for echocardiography, the negative predictive value is 79.5%.

Rule score: summation of points	Points
(b)	
Age per 10 years	1
History of ischaemic heart disease	1
Moderate to severe dyspnoea (MRC $\geq$ 3)	2
BMI per 5 kg/m^2	1
Laterally displaced or broadened/sustained apex beat	1
High-risk because of type 2 diabetes mellitus	2
High-risk because of multimorbidity and polypharmacy[a]	2
NTproBNP per 100 pg/ml	1

The probability of heart failure outcome is defined as 1 / (1 + (exp(-LP)), where LP refers to the linear predictor in a logistic regression model. The LP for the clinical score is defined as follows:

LP = − 10.40 + 0.54 * total sum of the score.

Use of the clinical scoring rule: for example, a 70-year-old person (seven points), without a history of ischaemic heart disease, who stops for breath after walking a few minutes on level ground (MRC dyspnoea score 4) (two points), a BMI of 25 kg/m^2 (five points), no laterally displaced or broadened apex beat, and a NTproBNP level of 130 pg/ml (130/100 $\approx$ 1) has a score of 15 points. According to Supplementary Figure 1(b) this score corresponds to a risk of heart failure of less than 9%. According to Table 4, if a general practitioner decided that all individuals with a probability of 9% or less will not be referred for echocardiography, the negative predictive value is 94.7%.

[a]Multimorbidity and polypharmacy is defined as having three or more chronic or vitality threatening diseases and/or using five or more prescribed drugs daily during the past year in people aged 65 years and over who filled out on a questionnaire that they experience symptoms of shortness of breath or reduced exercise tolerance.

BMI: body mass index; MRC: Medical Respiratory Council; NTproBNP: N-terminal pro B-type natriuretic peptide

exclude heart failure, considering heart failure unlikely if values are below the exclusion cut-off point (BNP < 35 pg/ml and NTproBNP < 125 pg/ml ($\approx$ 14.75 pmol/l) in those suspected of non-acute heart failure on a clinical basis.[16] Also, the higher the value, the more likely the diagnosis of heart failure is, making it a useful, easily applied predictor. Nevertheless, use in everyday practice is still rather low, particularly in primary care.[30]

Strengths and limitations of the study

A particular strength of our study was that we were able to combine four high quality screening studies for heart failure in community-dwelling older adults resulting in a large dataset consisting of 1941 people. As the different primary studies consisted of patients with a different background, our study consisted of participants representing various types of 'real life' patients who are in reality likely to seek the help of their GP. Therefore our results are generalisable to a broader patient population, more so than when compared with just a single study population, and can be applied to different types of patients with a few cardiovascular risk factors who are attending a GP's practice and may be suspected of having heart failure. On the other hand, as there were differences in baseline risk and study design, heterogeneity

Table 4. Application of the scoring rules with the diagnostic accuracy at different probability cut-off points.

Summed score from scoring rule	Probability of HF estimated by the scoring rule	Percentage of participants	Sensitivity	Specificity	Positive predictive value	Negative predictive value
(a) Application of the clinical prediction rule						
19	<5%	8.3%	0.99	0.11	26.1	98.9
20	<8%	19.2%	0.97	0.24	28.9	96.6
21	<12%	33.6%	0.90	0.41	32.7	89.8
22	<18%	49.4%	0.81	0.59	38.7	81
23	<27%	65.3%	0.65	0.75	45.2	64.8
24	<37%	77.7%	0.48	0.86	52.2	48.2
25	<48%	87.3%	0.32	0.93	60.7	31.2

Risk	Score range	Number of participants (%)	Number of patients with HF present (%)			
Very low	<20	372 (19.2)	16 (4.3)			
Low	20–22	586 (12.5)	73 (12.5)			
Moderate	22–24	550 (28.3)	154 (28.0)			
High	>24	433 (22.3)	226 (522)			

(b) Application of the clinical plus NTproBNP scoring rule						
13	<3%	9.00%	0.99	0.12	26.3	99.2
14	<5%	19.60%	0.98	0.25	29.5	98.1
15	<9%	34.50%	0.95	0.44	35	94.7
16	<15%	49.50%	0.86	0.61	41.2	86.1
17	<23%	63.30%	0.75	0.76	49.6	75.3
19	<33%	82.00%	0.48	0.92	64.6	48.2
21	<46%	93.40%	0.23	0.98	82.2	22.6

Risk	Score range	Number of participants (%)	Number of patients with HF present (%)			
Very low	<14	380 (19.6)	9 (2.4)			
Low	14–16	581 (29.9)	56 (9.6)			
Moderate	16–18	470 (24.2)	112 (28.8)			
High	>18	510 (26.3)	292 (57.3)			

HF: heart failure; NTproBNP: N-terminal pro B-type natriuretic peptide

between cohorts is present. However, the IECV approach takes this heterogeneity into account and adjusts for it with stratified estimation of the model's intercept.[26]

Another strength of our study is that, given the IECV methodology used, our model has already been externally validated in, again, a cohort that is representative of the real world clinical situation in the general population. Each primary study was used as an 'external' validation cohort and also, given the heterogeneity of the cohorts, this method is an accurate and effective way of validating our results.

In diagnostic studies the outcome should be measured as accurately as possible.[31] This presents itself as a limitation in two ways. First, there is no 'gold' standard to diagnose heart failure but the fact that the final diagnosis of heart failure was made by an expert panel is, on the other hand, a strength of this study. The expert panel based the diagnosis on all available diagnostic information and applied the criteria of the ESC. A disadvantage and therefore a limitation of such an expert panel is the risk of incorporation bias, as the reference standard (panel diagnosis) is not independent of the predictors studied. However, the extent

of incorporation bias in studies on the diagnosis of heart failure is limited because of the overriding importance of echocardiography for making the diagnosis, and that information from echocardiography was not used as a predictor when creating the prediction models.

In one of the primary studies, the STRETCH study, there was selectively incomplete diagnostic work-up. In said study only those individuals with a combination of an abnormal ECG and/or NTproBNP value >125 pg/ml (≈ 14.75 pmol/l) underwent echocardiography, and thus a small number of heart failure patients could have been missed, especially those in very early stages of HFpEF.

In conclusion, our study population is representative of, and our study results are generalisable to, the large population of older men and women from the community with considerable comorbidities, such as type 2 diabetes mellitus or COPD, and therefore at high risk of suffering undetected heart failure. With this study we offer tools for GPs to select those in need of echocardiography. We show which patient characteristics independently contribute to the estimation of the probability that a patient suffers from heart failure and to what extent the presence of one of these patient characteristics changes this probability. By use of a prediction-scoring rule, we have determined which cut-off scores should be used to determine who is at high enough risk to require echocardiography. Furthermore, use of the proposed rule in a high-risk population to select patients who could undergo echocardiography will reduce the number of under-diagnosed heart failure patients in this population and reduce healthcare costs involved in unnecessary referrals and echocardiography.

Author contribution

RK, AH, EVR, LB-W, MB, HdR and FR contributed to the conception and design of the work. EVR, YvM, LB, LB-W and FR contributed to the acquisition of data. RK, AG, AH MB and FR contributed to the analysis and interpretation of the data. RK and AG drafted the manuscript. EVR, YvM, LB, LB-W, AH, MB, HdR and FR critically revised the manuscript. All authors gave final approval and agree to be accountable for all aspects of work ensuring integrity and accuracy.

Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: AH chairs a large research and teaching institute within the University Medical Center. Both investigator- and industry-driven research projects are performed with a number of pharmaceutical and diagnostic companies. In addition, some members of staff receive unrestricted grants for research projects from a number of companies. It is policy to work with several companies and not to focus on one or two industrial partners. AH receives no personal payment from any industrial partner.

Funding

References

1. Heidenreich PA, Albert NM, Allen LA, et al. Forecasting the Impact of Heart Failure in the United States. *Circ Hear Fail* 2013; 6: 606–619.

2. Piepoli MF, Hoes AW, Agewall S, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice. *Eur J Prev Cardiol* 2016; 23: NP1–NP96.

3. Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J* 2016; 37: 2129–2200.

4. Rutten FH, Cramer MJM, Grobbee DE, et al. Unrecognized heart failure in elderly patients with stable chronic obstructive pulmonary disease. *Eur Heart J* 2005; 26: 1887–1894.

5. Boonman-de Winter LJM, Rutten FH, Cramer MJM, et al. Efficiently screening heart failure in patients with type 2 diabetes. *Eur J Heart Fail* 2015; 17: 187–195.

6. van Mourik Y, Bertens LCM, Cramer MJM, et al. Unrecognized heart failure and chronic obstructive pulmonary disease (COPD) in frail elderly detected through a near-home targeted screening strategy. *J Am Board Fam Med* 2014; 27: 811–821.

7. Fuat A, Hungin APS and Murphy JJ. Barriers to accurate diagnosis and effective management of heart failure in primary care: qualitative study. *BMJ* 2003; 326: 196.

8. Cowie MR, Anker SD, Cleland JGF, et al. Improving care for patients with acute heart failure: Before, during and after hospitalization. *ESC Hear Fail* 2014; 1: 110–145.

9. Rutten FH, Clark AL and Hoes AW. How big a problem is heart failure with a normal ejection fraction? *BMJ* 2016; 1706: i1706.

10. Caroci A de S and Lareau SC. Descriptors of dyspnea by patients with chronic obstructive pulmonary disease versus congestive heart failure. *Hear Lung J Acute Crit Care* 2004; 33: 102–110.

11. Mant J, Doust J, Roalfe A, et al. Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care. *Health Technol Assess* 2009; 13(32).

12. Madhok V, Falk G, Rogers A, et al. The accuracy of symptoms, signs and diagnostic tests in the diagnosis of

left ventricular dysfunction in primary care: a diagnostic accuracy systematic review. *BMC Fam Pract* 2008; 9: 56.

13. Fahey T, Jeyaseelan S, McCowan C, et al. Diagnosis of left ventricular systolic dysfunction (LVSD): Development and validation of a clinical prediction rule in primary care. *Fam Pract* 2007; 24: 628–635.

14. Kelder JC, Cramer MJM, Van Wijngaarden J, et al. The diagnostic value of physical examination and additional testing in primary care patients with suspected heart failure. *Circulation* 2011; 124: 2865–2873.

15. Van Riet EES, Hoes AW, Limburg A, et al. Prevalence of unrecognized heart failure in older persons with shortness of breath on exertion. *Eur J Heart Fail* 2014; 16: 772–777.

16. McMurray JJV, Adamopoulos S, Anker SD, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. *Eur Heart J* 2012; 33: 1787–1847.

17. Moons KGM and Grobbee DE. When should we remain blind and when should our eyes remain open in diagnostic studies? *J Clin Epidemiol* 2002; 55: 633–636.

18. Severo M, Gaio AR, Lourenço P, et al. Diagnostic value of patterns of symptoms and signs of heart failure: application of latent class analysis with concomitant variables in a cross-sectional study. *BMJ Open* 2012; 2: 1–9.

19. Srivastava PM, Calafiore P, Macisaac RJ, et al. Prevalence and predictors of cardiac hypertrophy and dysfunction in patients with Type 2 diabetes. *Clin Sci* 2008; 114: 313–320.

20. Aspromonte N, Ceci V, Chiera A, et al. Rapid brain natriuretic peptide test and doppler echocardiography for early diagnosis of mild heart failure. *Clin Chem* 2006; 52: 1802–1808.

21. Davie AP, Francis CM, Caruana L, et al. Assessing diagnosis in heart failure: which features are any use? *QJM* 1997; 90: 335–339.

22. Harrel Jr. FE. Regression modeling strategies – with applications to linear models, logistic and ordinal regression, and survival analysis. *R Softw* 2015; 70: 598.

23. Rose GA, Blackburn H, Gillum RF, et al. Cardiovascular survey methods. *World Heal Organ – Monogr Ser*, 1982.

24. Van Buuren S and Groothuis-Oudshoorn K. Multivariate Imputation by Chained Equations. *J Stat Softw* 2011; 45: 1–67.

25. Steyerberg EW and Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol* 2016; 69: 245–247.

26. Debray TPA, Moons KGM, Ahmed I, et al. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. *Stat Med* 2013; 32: 3158–3180.

27. Lavie CJ, Alpert MA, Arena R, et al. Impact of obesity and the obesity paradox on prevalence and prognosis in heart failure. *JACC Hear Fail* 2013; 1: 93–102.

28. Rutten FH, Moons KGM, Cramer MJM, et al. Recognising heart failure in elderly patients with stable chronic obstructive pulmonary disease in primary care: cross sectional diagnostic study. *BMJ* 2005; 331: 1379.

29. Scantlebury DC and Borlaug BA. Why are women more likely than men to develop heart failure with preserved ejection fraction? *Curr Opin Cardiol* 2011; 26: 562–568.

30. Hancock HC, Close H, Fuat A, et al. Barriers to accurate diagnosis and effective management of heart failure have not changed in the past 10 years: a qualitative study and national survey. *BMJ Open* 2014; 4: e003866.

31. Grobbee DE and Hoes AW. *Clinical Epidemiology. Principles, Methods, and Applications.* Boston: Jones and Bartlett, 2009.

32. Van Houwellingen JC and Le Cessie S. Predictive value of statistical models. *Stat Med* 1990; 9: 1303–1325.