

# BMJ Open Extended prediction rule to optimise early detection of heart failure in older persons with non-acute shortness of breath: a cross-sectional study

Evelien E S van Riet,<sup>1</sup> Arno W Hoes,<sup>1</sup> Alexander Limburg,<sup>2</sup> Marcel A J Landman,<sup>1</sup> Hans Kemperman,<sup>1</sup> Frans H Rutten<sup>1</sup>

**To cite:** van Riet EES, Hoes AW, Limburg A, *et al*. Extended prediction rule to optimise early detection of heart failure in older persons with non-acute shortness of breath: a cross-sectional study. *BMJ Open* 2016;**6**: e008225. doi:10.1136/bmjopen-2015-008225

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-008225>).

Received 17 March 2015

Accepted 23 November 2015



CrossMark

<sup>1</sup>Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

<sup>2</sup>Diakonessenhuis Zeist, Zeist, The Netherlands

#### Correspondence to

Dr Evelien ES van Riet; E.E.S.vanRiet@umcutrecht.nl

## ABSTRACT

**Objectives:** There is a need for a practical tool to aid general practitioners in early detection of heart failure in the elderly with shortness of breath. In this study, such a screening rule was developed based on an existing rule for detecting heart failure in older persons with a diagnosis of chronic obstructive pulmonary disease. The original rule included a history of ischaemic heart disease, body mass index, laterally displaced apex beat, heart rate, elevated N-terminal pro B-type natriuretic peptide and an abnormal ECG.

**Design:** Cross-sectional data were used to validate, update and extend the original prediction rule according to a standardised state-of-the-art stepwise approach.

**Setting:** Primary care with 30 participating general practices.

**Participants:** Community-dwelling people aged  $\geq 65$  years with shortness of breath on exertion.

**Methods and results:** Validation of the existing screening rule in our population showed satisfying discrimination with a concordance statistic of 0.84 (range 0.80–0.85), but poor calibration. Performance measures were most improved by adding the predictors age  $>75$  years, peripheral oedema and systolic murmur, resulting in a concordance statistic of 0.88 (range 0.85–0.90) and a net reclassification improvement of 31%. A risk score was computed, which showed high accuracy with a negative predictive value of 87% and a positive predictive value of 73%. Evaluating the improved rule in the derivation set and an independent set of patients with type 2 diabetes aged 60 years or older showed satisfying generalisability of the rule.

**Conclusions:** Our rule resulted in excellent prediction of heart failure in the large domain of the elderly with shortness of breath, and would help general practitioners to select those needing echocardiography.

**Trial registration number:** NCT01202006.

## INTRODUCTION

Heart failure is an emerging ‘epidemic’, causing high mortality rates, substantial loss

## Strengths and limitations of this study

- We did not perform echocardiography in participants who had a normal ECG and normal N-terminal pro B-type natriuretic peptide levels ( $<14.75$  pmol/L). This may have caused partial verification bias in the heart failure diagnosis. However, by state-of-the-art imputation techniques, we could provide an adequate prediction of the very low risk of heart failure for such persons.
- The use of an outcome panel to establish heart failure presence or absence may have resulted in incorporation bias, by knowing the results of the diagnostic tests under study. Importantly, there is general consensus that the resulting overestimation of the performance of some of the diagnostic items is outweighed by the gain in the accuracy of the outcome assessment by the panel.
- The use of tissue Doppler imaging in the assessment of diastolic function is a major strength of our study since it helps reduce misclassification of heart failure with preserved ejection fraction.
- Our prediction rule is able to detect both ‘phenotypes’ of heart failure in an early stage, and it can help general practitioners decide who needs referral for echocardiography.

in quality of life and high healthcare costs.<sup>1</sup> The majority of patients with heart failure are diagnosed and managed in primary care.<sup>2</sup> Diagnosing heart failure in an early phase is difficult because symptoms and signs such as fatigue, shortness of breath and oedema are non-specific. Other conditions that may even be present and already known as comorbidities can also cause a similar clinical picture. Echocardiography is the cornerstone investigation to diagnose heart failure, but there is limited access to this facility in primary care.

Non-acute shortness of breath is a very common symptom of older community-dwelling persons, with prevalence rates of

around 40%.<sup>3 4</sup> Many elderly experiencing such symptoms visit the general practitioner (GP), but this will not always result in an adequate diagnostic work up. Shortness of breath is frequently incorrectly labelled as pulmonary obstructive disease or as being caused by ageing itself or deconditioning. Heart failure is often not considered and high rates of unrecognised heart failure may be found in community-dwelling older persons labelled with chronic obstructive pulmonary disease (COPD).<sup>5</sup> The existing knowledge of underdiagnosis and misclassification underscores the need for a practical decision tool to aid GPs in targeted screening for heart failure of the large population of community-dwelling older persons with non-acute shortness of breath.

Several diagnostic prediction rules have been developed for detecting or excluding non-acute heart failure in the primary care setting, mainly in patients suspected of heart failure.<sup>6–9</sup> Simply creating another prediction rule for people with shortness of breath would only contribute to a multiplication of rules. Therefore, we set out to validate and further develop the single existing *screening* rule for heart failure from primary care. This screening rule was created in a comparable group of participants, namely, community-dwelling persons aged 65 years or older with a GP's diagnosis of COPD.<sup>5</sup> We followed a state-of-the-art stepwise approach (table 1) to validate, update and extend this previously developed prediction rule.<sup>10</sup>

## METHODS

### Study population

Thirty primary care practices in the Netherlands participated in our study, executed between December 2010 and December 2012. Community-dwelling persons aged 65 years or older were eligible if they presented themselves with non-acute shortness of breath on exertion in the previous 12 months to one of the participating GPs, including those individuals with an already known pulmonary disease. Shortness of breath was not necessarily the main reason for contact. Eligibility was irrespective of whether patients were suspected of heart failure by their GP or not. We excluded patients with an already known established diagnosis of heart failure (confirmed by a cardiologist with echocardiography), those with a life expectancy shorter than 6 months and those not able or willing to give informed consent. The study was approved by the medical ethical committee of the University Medical Center Utrecht, the Netherlands.

### Data collection

After signing informed consent, participants underwent a standardised diagnostic work up. A standardised questionnaire was used to obtain information on symptoms, comorbidities, smoking status and medication use. Physical examination consisted of measurement of height and weight, blood pressure (two readings), pulse,

**Table 1** Steps in validating and improving prediction rules

Serial number	Step
<i>No updating</i>	
1	Applying the original prediction model
<i>Recalibration</i>	
2	Adjustment of the intercept using the calibration intercept
3	Adjustment of intercept and regression coefficients using the calibration intercept and the calibration slope
<i>Model revision</i>	
4	Step 3 + extra adjustment of regression coefficients for predictors with a different strength than in the development set
5	Re-estimation of the intercept and regression coefficients of all predictors
<i>Model extension</i>	
6	Step 4 + selective extension with additional (statistically significant) predictors
7	Step 5 + selective extension with additional (statistically significant) predictors
8	Fit a model with original and additional predictors

respiratory rate, pulmonary percussion and auscultation, heart auscultation, palpation of the apex beat, measurement of the jugular venous pressure (JVP), palpation of the liver, and inspection of the legs for oedema and signs of chronic venous insufficiency. An impalpable apex beat was classified as 'normal apex beat', and immeasurable JVP as 'non-elevated JVP'.

A standard 12-lead ECG was recorded and classified according to the Minnesota coding criteria,<sup>11</sup> without knowledge of the patients' clinical status. Blood samples were analysed for serum concentrations of N-terminal pro B-type natriuretic peptide (NTproBNP) with a non-competitive immunoradiometric assay (Roche, Mannheim, Germany). NTproBNP was considered elevated if the value exceeded 14.75 pmol/L (~125 pg/mL). This exclusionary cut-point for non-acute onset heart failure was chosen in line with the European Society for Cardiology (ESC) guidelines on heart failure.<sup>12</sup>

Only participants with an abnormal ECG or an NTproBNP level above 14.75 pmol/L underwent additional echocardiography within 2 weeks. This diagnostic strategy is advocated by the recent ESC guidelines on heart failure and largely resembles daily practice in the Netherlands.<sup>12</sup> Echocardiography, including Tissue Doppler techniques for measurement of diastolic function, was performed by a single trained and experienced cardiac sonographer, using a Philips iE33 imaging system (Andover, Massachusetts, USA), blinded to the patients' other test results and applying the recent recommendations of the American Society of Echocardiography.<sup>13 14</sup>

## Definition of heart failure

An expert panel, consisting of two cardiologists and one GP with special expertise in heart failure, determined presence or absence of heart failure based on all available diagnostic test results, including echocardiography. For classification of participants as having heart failure, the panel followed the criteria of the guidelines of the ESC.<sup>12</sup>

Heart failure was considered present when participants had suggestive symptoms and signs in combination with objective echocardiographic evidence of cardiac dysfunction at rest. Heart failure was further classified by the panel as heart failure with reduced ejection fraction (HF-REF) in case left ventricular ejection fraction (LVEF) was  $\leq 45\%$ , heart failure with preserved ejection fraction (HF-PEF) when there were structural or functional abnormalities compatible with diastolic dysfunction and LVEF was  $>45\%$ , and 'isolated' right-sided heart failure.

## Data analysis

### Imputation

The decision to not perform echocardiography was based on earlier test results, namely a normal ECG and NTproBNP value  $<14.75$  pmol/L. As a consequence, the outcome, presence or absence of heart failure, was partially missing 'at random' in our data set.<sup>15</sup> Methodological studies showed that if values are missing at random, multiple imputation of these variables results in rather unbiased estimates, even when the missing variable is the missing reference standard.<sup>16</sup>

We used logistic regression for single imputation of two missing ECGs and three missing NTproBNP values, and multiple logistic regression imputation for the outcome value.<sup>17</sup> Based on the percentage of missing information on the outcome (37.4%), we performed 40 repetitions of multiple imputation.<sup>18</sup>

Analyses were performed in each of the 40 imputed sets and, subsequently, combined estimates of regression coefficients and their variances were computed following Rubin's rules.<sup>19</sup>

### Validation of the original rule

The original prediction rule includes a history of ischaemic heart disease, body mass index, laterally displaced apex beat, heart rate, NTproBNP and ECG. The prediction rule has the following form:  $\alpha + \beta_1 \times \text{predictor 1} + \beta_2 \times \text{predictor 2} + \dots + \beta_6 \times \text{predictor 6}$ , with  $\alpha$  being the intercept and  $\beta_1 - \beta_6$  the regression coefficients. For validation (step 1) we applied the original intercept, regression coefficients and predictors before further updating.

### Model improvement

In step 2, we recalibrated the model by adjustment of the intercept, and in step 3, by adjustment of both the intercept and the regression coefficients.

Steps 4 and 5 are model revisions in which predictors are re-estimated in the validation set. In step 4, only

predictors with a different strength in the validation set after recalibration were adjusted. We tested whether deviations from the recalibrated regression coefficients had added predictive value by performing likelihood ratio tests in a forward stepwise manner. In step 5, all the predictors were newly fitted in the model.

Steps 6–8 are model extensions considering other variables for inclusion in the model. In step 6, predictors were included with additional value after performance of step 4 and in step 7, predictors were included with additional value after performance of step 5. Variables with a  $p < 0.15$  in univariable logistic regression were selected from the validation set and added in a forward stepwise manner, including the predictor with the strongest effect first. In step 8, a model with original and additional predictors was newly estimated. We included a maximum number of three additional predictors in the multiple regressions, to retain sufficient power of the model and prevent overfitting.<sup>20</sup>

### Shrinkage of the regression coefficients

In steps 4–8, the value of the individual regression coefficients was estimated in the validation set. Owing to overfitting, these estimations are often too optimistic, resulting in overly extreme predicted probabilities when the model is applied in a new data set.<sup>19</sup> We reduced overoptimism by shrinkage of the regression coefficients, following the formulas described by Janssen *et al.*<sup>21</sup>

### Model performances

To study the performance of the rules after the various updating and extension steps, we compared the predicted risks with the observed outcomes. We studied calibration with calibration plots and the Hosmer-Lemeshow goodness-of-fit test, and discrimination with the concordance statistic (C-statistic).<sup>21</sup>

Combining C-statistic estimates from 40 imputed sets into one estimate is not possible with Rubin's rules since C-statistics are bounded by zero and one.<sup>22</sup> Hence, we used the median and range to describe the distribution of this value over the imputed sets, as was carried out earlier by Clark and Altman.<sup>23</sup>

We used the  $\chi^2$  difference test to evaluate the added value of each step and thereby determine the best performing rule.<sup>24</sup>

We then cross-tabulated the probability classification of patients according to this improved rule with the original rule to assess reclassification. We calculated the net reclassification improvement (NRI) for cases upward and for non-cases downward the probability of heart failure scale.<sup>25</sup> We used a cut-point of 10% risk, considering heart failure to be absent below, and 'possibly present' above, that threshold (hence regarding it as an indication for referral for echocardiography).

Finally, we transformed the regression coefficients of the predictors from the improved rule to the nearest integers according to their relative contributions to the risk estimations to construct risk scores for practical use.

After calculating the score points for each patient, we estimated the absolute percentages of correctly diagnosed patients across score categories. We also calculated the sensitivity, specificity and predictive values at different score thresholds.

To test the improved rule on its generalisability, we assessed the performance of the rule in the derivation cohort of community-dwelling older persons with COPD, and in another, independent cohort of 581 community-dwelling persons aged 60 years and older with type 2 diabetes who had been assessed echocardiographically.<sup>26</sup>

Data were analysed with SPSS V.17.0 for Windows (SPSS Inc, Chicago, Illinois, USA) and R V.3.0.1.

## RESULTS

The mean age of the 585 participants was 74.1 (SD±6.3) years, and 54.5% were females. In total, 366 participants (62.6%) underwent echocardiography.

The panel diagnosed newly onset heart failure in 92 patients (15.7% of all participants); 17 (2.9%) with HF-REF, 70 (12.0%) with HF-PEF and 5 (0.9%) with

isolated right-sided heart failure. Characteristics of participants according to presence or absence of heart failure, considering those with a normal ECG and NTproBNP<14.75 pmol/L as not having heart failure, are shown in table 2.

The 40 imputed sets resulted in a mean number of 6 (range 2–13) imputed heart failure cases in the 219 participants who did not have echocardiography because they had a normal ECG and an NTproBNP value <14.75 pmol/L.

Validation of the original rule (step 1) showed good discrimination with a C-statistic of 0.84 (range 0.80–0.85) (table 3). Calibration, however, was insufficient because of overly extreme predictions (figure 1).

After recalibration and revision, the predictors age >75 years, cardiovascular comorbidity, atrial fibrillation, nocturia, pulmonary crepitations, peripheral oedema and systolic murmur were tested for model extension. Age >75 years, peripheral oedema and a systolic murmur had the highest added value beyond the original six predictors, and were used in the extension steps. The best performing rule was created with step 8, resulting in a C-statistic of 0.88 (range 0.85–0.90). Regression

**Table 2** Characteristics of 585 participants according to presence or absence of HF, considering those with a normal ECG and NTproBNP<14.75 pmol/L as not having HF

Characteristics	HF present (n=92)	HF absent (n=493)	OR (95% CI)	p Value
<b>History</b>				
Mean age in years (SD)	78.1 (6.0)	73.3 (6.0)	1.13 (1.09 to 1.17)*	<0.001
Female sex	48 (52.2)	271 (55.0)	0.89 (0.57 to 1.40)	0.62
Median pack years of smoking (IQR)†	7.8 (0.0–30.0)	8.0 (0.0–30.0)	1.01 (1.00 to 1.02)*	0.99
Orthopnoea or paroxysmal nocturnal dyspnoea	18 (19.6)	72 (14.6)	1.42 (0.80 to 2.52)	0.23
Nocturia ≥ twice/night	31 (33.7)	110 (22.3)	1.77 (1.09 to 2.86)	0.02
<b>Comorbidities and risk factors‡</b>				
Ischaemic heart disease§	30 (32.6)	86 (17.4)	2.29 (1.40 to 3.75)	0.001
Cardiovascular comorbidity¶	72 (78.3)	279 (56.6)	2.76 (1.63 to 4.68)	<0.001
Atrial fibrillation	17 (18.5)	25 (5.1)	4.24 (2.19 to 8.23)	<0.001
<b>Physical examination</b>				
Mean BMI in kg/m <sup>2</sup> (SD)	29.9 (4.6)	27.2 (4.3)	1.13 (1.08 to 1.19)*	<0.001
Mean heart rate in bpm (SD)	71.9 (16.4)	74.0 (11.7)	0.99 (1.97 to 1.00)*	0.23
Pulmonary crepitations	26 (28.3)	83 (16.8)	1.95 (1.17 to 3.25)	0.01
Peripheral oedema	45 (48.9)	101 (20.5)	3.72 (2.34 to 5.91)	<0.001
Laterally displaced apex beat**	7 (7.6)	14 (2.8)	2.82 (1.11 to 7.19)	0.024
Systolic murmur	32 (34.8)	94 (19.1)	2.26 (1.40 to 3.67)	0.001
<b>Additional tests††</b>				
Median NTproBNP in pmol/L (IQR)	46 (24–89)	13 (7–20)	1.02 (1.02 to 1.03)*	<0.001
Abnormal ECG‡‡	60 (65.9)	160 (32.5)	4.06 (2.53 to 6.50)	<0.001

Values are numbers (percentage) unless stated otherwise.

\*OR per unit change.

†In current and former smokers.

‡Comorbidities mentioned by the patient during history taking.

§Including myocardial infarction, angina pectoris, percutaneous coronary intervention or coronary artery bypass grafting.

¶Diabetes mellitus, hypertension, stroke, or peripheral arterial disease.

\*\*Laterally displaced in supine position or broadened/sustained in left cubital position.

††There were two missing values for ECG measurements and three missing values for NTproBNP measurements.

‡‡Atrial fibrillation, sinus tachycardia (heart rate >100 bpm), left and right bundle branch block (complete or incomplete), left anterior and posterior fascicular block, left ventricular hypertrophy, Q-waves suspected for previous myocardial infarction, P-wave abnormalities compatible with left atrial enlargement or P-pulmonale, or any ST segment/T-wave abnormalities.

BMI, body mass index; HF, heart failure; NTproBNP, N-terminal pro B-type natriuretic peptide.

**Table 3** Performance measurements of the improved rule compared with the original rule

	Original rule	Improved rule
C-statistic (range)	0.84 (0.80–0.85)	0.88 (0.85–0.90)
Males	0.84 (0.82–0.86)	0.88 (0.85–0.90)
Females	0.84 (0.78–0.85)	0.89 (0.84–0.90)
HL-test	0.02	0.56
N cases correctly upgraded		0
N cases incorrectly downgraded		6
N non-cases correctly downgraded		180
N non-cases incorrectly upgraded		0
% NRI		31.0

The median C-statistic and range from the 40 imputed analyses are given.

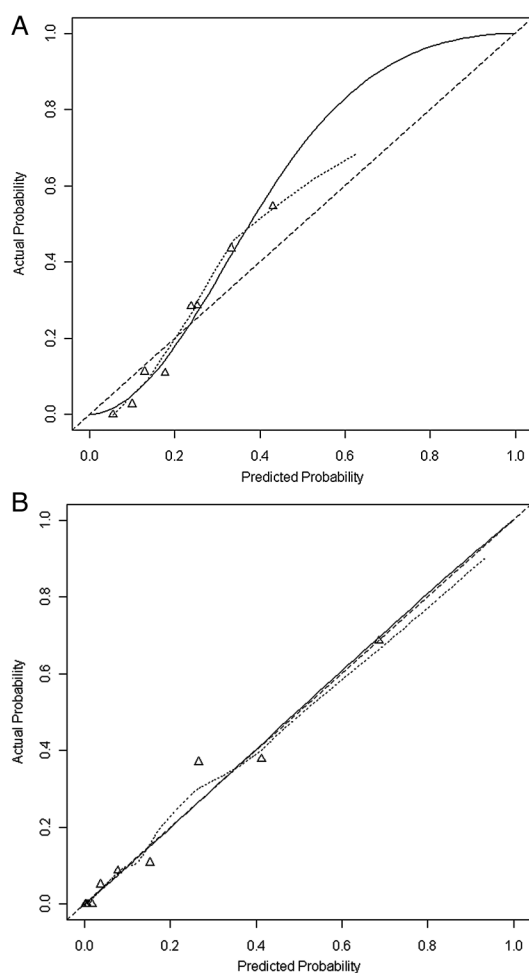
NRI is defined as the difference in proportions of patients moving up and down for cases and non-cases separately, and is interpreted as the percentage reclassified, adjusted for reclassification direction.

C-statistic, concordance statistic; HL-test, Hosmer-Lemeshow test; NRI, net reclassification improvement.

**Table 4** Risk score for estimating the probability of heart failure with the improved rule

Predictor	Points
Ischaemic heart disease	2
Body mass index >30 kg/m <sup>2</sup>	4
Laterally displaced apex beat	4
Heart rate >90 bpm	1
NTproBNP >14.75 pmol/L	9
Abnormal ECG	5
Peripheral oedema	4
Age >75 years	3
Systolic murmur	3
Median C-statistic	0.88
Range from 40 analyses	0.85–0.89

C-statistic, concordance statistic; NTproBNP, N-terminal pro B-type natriuretic peptide.



**Figure 1** Calibration plots of the original (A) and improved (B) prediction rule. Agreement between the predicted risks of heart failure according to the different prediction rules and the observed proportions in the validation set. The broken line indicates ideal calibration (line of identity), the dotted line is the non-parametric calibration line and the smooth line the parametric calibration line.

coefficients of this improved rule can be found in online supplementary table S1.

The NRI for the improved rule was 31.0%, mainly caused by the correct down-classification of non-cases (table 3). The calibration plot of the improved rule is depicted in figure 1. Neither the original nor the improved rule performed differently between sexes (table 3).

Tables 4 and 5 show the construction of the risk score with the improved rule and the result of calculating individual risks and dividing participants over score categories, respectively. Dichotomising the scale by considering heart failure absent under the score of 21 points and considering heart failure possibly present (regarding it as an indication for echocardiography) above this score, yielded a positive predictive value (PPV) of 73% and a negative predictive value (NPV) of 87%.

The improved rule showed good performance with a C-statistic of 0.76 (95% CI 0.70 to 0.81) in the derivation cohort of community-dwelling older persons with COPD, exactly matching with the C-statistic of the original rule in that data set. In the independent cohort of community-dwelling persons aged 60 years and older with type 2 diabetes, the discrimination was even better, with a C-statistic of 0.80 (95% CI 0.76 to 0.84).

## DISCUSSION

We used a stepwise approach to further develop an existing prediction rule that may help GPs identify which community-dwelling older persons presenting with shortness of breath on exertion might have non-acute heart failure. By extending the original six predictors with three additional predictors, performance measures were most improved, resulting in a good-to-excellent C-statistic of 0.88 (range 0.85–0.90) and an NRI of 31.0%. A risk score was computed, which showed high accuracy with a NPV of 87% and a PPV of 73%. The improved rule was tested on its generalisability with satisfying results in two sets (C-statistic 0.76 and 0.80).

**Table 5** Presence and absence of heart failure (HF) per score category with the improved rule, and corresponding sensitivity, specificity and predictive values when dichotomised at different thresholds

Risk category (points)	HF (n=98)	No HF (n=487)	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Very low ( $\leq 9$ ), n=268	5	263				
Low (10–15), n=128	11	117	0.95	0.54	0.29	0.98
Medium (16–21), n=149	53	96	0.84	0.78	0.43	0.96
High (22–30), n=40	29	11	0.30	0.98	0.73	0.87

Our study has several limitations. Not performing echocardiography in those with a normal ECG in combination with NTproBNP levels below 14.75 pmol/L may have caused partial verification bias. Importantly, heart failure in such patients is very unlikely.<sup>12 27</sup> Moreover, by state-of-the-art imputation techniques, we could provide adequate predictions of the very low risk of heart failure for these patients who did not undergo echocardiography. A study by de Groot *et al*<sup>16</sup> demonstrated that multiple imputation is the preferred correction method in case of partial verification bias. The authors performed a series of simulations in a data set with complete verification by setting a varying number of outcome values to missing and comparing the performance of various correction methods. In case the mechanism of missing data is known, such as in our study, multiple imputation showed reliable estimates. Even introducing up to 30% missings resulted in estimates of performance fluctuating around the ‘true’ values, only at the costs of wider CIs with increasing numbers of missing outcomes.

After imputation, six cases of heart failure were added: 2.7% of those who did not undergo echocardiography. This percentage corresponds very well with a previous diagnostic study evaluating suspected heart failure patients from primary care showing a prevalence of heart failure in 2.9% of patients with a normal ECG and NTproBNP levels <14.75 pmol/L.<sup>27</sup> In the cohort of older COPD patients from primary care in which the original prediction rule was developed, this percentage was somewhat higher: in total, 5% of the patients had heart failure in the presence of a normal ECG and NTproBNP<14.75 pmol/L; all these cases had HF-PEF.<sup>5</sup> It is reassuring that by multiple imputation of missing outcomes comparable percentages of heart failure cases were imputed in our study.

To evaluate the effect imputation might have had on our results, we executed a kind of sensitivity analysis. We also performed validation and model improvement steps in two other data compositions; a data set in which patients with a normal ECG and NTproBNP<14.75 pmol/L were straightforwardly considered not to have heart failure (set 2), and a data set comprised of those who had undergone echocardiography, selectively excluding 219 persons with a normal ECG and NTproBNP<14.75 pmol/L (set 3). The results are presented in the online supplementary material. In short, the overall performance of the prediction rule in

set 2 was close to our imputed data set (C-statistic 0.89, 95% CI 0.86 to 0.92), while set 3, consisting of the selection of 366 persons who underwent echocardiography, was lower (C-statistic 0.82, 95% CI 0.77 to 0.87). Set 3 represents another domain, that is, older patients with dyspnoea on exertion who have an abnormal ECG and/or NTproBNP>14.75 pmol/L; a more selective patient category with a higher probability of heart failure. As could be expected, especially, the ability to rule out heart failure was reduced, resulting in a lower NPV than the rule in our imputed data set (81% vs 87%).

As in all diagnostic heart failure studies, appliance of the outcome panel may have resulted in incorporation bias, by knowing the results of the diagnostic tests under study. Importantly, however, there is general consensus that the resulting overestimation of the performance of some of the diagnostic items (eg, signs and symptoms) is outweighed by the gain in the accuracy of the outcome (heart failure presence or absence) assessment of the panel.<sup>28</sup>

A major strength of our study is the use of tissue Doppler imaging (TDI) in the assessment of diastolic function. Previously developed prediction rules vary in their definition of heart failure and, especially, the definition of HF-PEF has changed over time and by study. Nowadays, HF-PEF is considered present when heart failure symptoms and signs concur with structural or functional abnormalities of the heart, preferably visualised with echocardiography including TDI of ventricular wall movement, to reduce the risk of misclassification.<sup>12</sup>

We chose not to make separate rules for prediction of HF-REF and HF-PEF. Our aim was to construct a rule to detect both ‘phenotypes’ of heart failure in an early stage. Implications for therapy differ between both ‘phenotypes’ and therefore it is important to make an adequate distinction with echocardiography. The main advantage of the rule and resulting risk score is that it is helpful to GPs in dealing with older persons with shortness of breath by easing the shift between those who need and those who do not need referral for echocardiography. For this utility one rule can be used.

Earlier research on diagnosing heart failure focused on the development of prediction rules in people *suspected* of heart failure. Our rule is of unique value because it is applicable to all elderly individuals who visit the GP with shortness of breath, irrespective of being

suspected of heart failure; a much larger population that may even be as large as 40% of those aged 65 years and over.<sup>3,4</sup> Importantly, suspicion of heart failure is subjective, and physicians may be affected by ‘cognitive bias’ because of prior misclassification of shortness of breath and fatigue as caused by a respiratory disease. Misclassification of symptoms as being caused by a pulmonary disease may have unfortunate consequences, as treatment with bronchodilators potentially may cause adverse cardiac effects.<sup>29</sup>

The nine predictors included in our clinical prediction rule are known for being associated with heart failure, and are comparable with those used in rules for people suspected of heart failure.<sup>6–9</sup> In contrast to most other rules, however, in our rule, male gender was not independently related to heart failure. This could at least partly be explained by the relatively large group of patients with HF-PEF in our study, with HF-PEF being known to be more prevalent in females.

The number of predictors in our rule is fairly high, but this is not unique.<sup>9</sup> Most importantly, each predictor had independent added value. It is generally accepted, from a methodological point of view, that the final rule should consist of determinants gaining independent added value with multivariable logistic regression analysis. With modern technology, the number of items should not form a barrier to implementation because clinical prediction rules can be built into medical electronic systems or in apps to be downloaded on mobile devices. Implementation in primary care would largely increase the yield of case findings of heart failure without being inefficient, costly and time-consuming. Starting with the large number of persons with shortness of breath to be considered for screening, our rule can help GPs select the sample of persons who need echocardiography, with a low risk of incorrectly excluding heart failure.

By using the threshold of 21 points to select those needing referral for echocardiography, especially early stages of HF-PEF could be missed (13% false negatives) in the low-risk and medium-risk groups. Given the chronic progressive character of heart failure, these patients may have another chance of being detected in case of persistence or progression of symptoms. Such a strategy would best fit in a setting with open access to echocardiography. Open access, however, is still limited in many countries, including the Netherlands, mainly because of the belief that it may lead to unnecessary referrals. Given the relatively high PPV of 73% in our study, the number of false-positives is limited when considering the screening setting. By targeting resources at the high-risk group found with our selective screening, this approach is most likely to be efficient.

Early detection of heart failure could lead to timely management. For HF-REF the effect of renin-angiotensin system inhibitors and  $\beta$ -blockers on mortality and heart failure related hospitalisations is well established.

Disappointingly, these medications have, at best, shown a tendency to improve the prognosis in HF-PEF. However, focusing only on prognostically beneficial treatment would be too nihilistic. Establishing the diagnosis of HF-PEF is also important to help explain its symptoms to patients, to prevent misclassification (ie, such as a pulmonary disease) and to help predict prognosis. Most importantly, diuretics are known to relieve shortness of breath and other symptoms related to fluid and salt retention, and the blood pressure should be adequately managed to reduce the afterload for the heart. Moreover, there are multiple ongoing studies investigating novel compounds aimed at targets such as inflammation, coronary microvascular dysfunction and early stages of myocardial fibrosis. One or more of these compounds may eventually show to have clear prognostically beneficial effects.<sup>30</sup>

Whenever evidence-based prognostically beneficial treatment for HF-PEF becomes available, selective screening for heart failure with our improved rule could prove to be cost-effective, given the high number of discovered cases of previously undetected HF-PEF.

We used the original rule of Rutten *et al*<sup>5</sup> for the development of our improved rule. We re-estimated the intercept and regression coefficients of all predictors (old and new) on the data of our own set. Therefore, we had to test our improved rule again on its generalisability. A real cohort similar to ours is lacking in the literature and we therefore had to use validation cohorts with a somewhat different domain. We first validated the rule in the derivation cohort consisting of elderly community-dwelling persons with a diagnosis of COPD, with good discrimination as a result (C-statistic 0.76, 95% CI 0.70 to 0.81). Our improved rule also showed good discrimination in a completely independent cohort of community-dwelling elderly individuals with type 2 diabetes (C-statistic 0.80, 95% CI 0.76 to 0.84). Although these validation results are reassuring, we recommend first to perform an impact study before considering large-scale implementation in daily practice.<sup>31</sup>

In conclusion, further development of an existing rule to selectively screen for heart failure in a large population of community-dwelling older persons with non-acute shortness of breath resulted in an excellent performing prediction rule that is readily applicable, and has large potential impact on daily general practice and population health.

**Contributors** AWH and FHR designed the study. FHR coordinated the study. EESvR managed the study and data collection. HK participated in data collection and analysis. AL, MAJL and FHR were members of the expert panel. EESvR conducted the statistical analyses and wrote the first draft of the manuscript. All the authors read and approved the final draft of the manuscript, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

**Funding** The study was conducted with a research grant from the Dutch Heart Foundation (‘Nederlandse Hartstichting’ grant number 2009B048).

**Competing interests** None declared.

**Ethics approval** The medical ethical committee (METC) of the University Medical Center Utrecht, the Netherlands.

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

**Data sharing statement** Statistical code can be obtained from the corresponding author.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

## REFERENCES

- Schocken DD, Benjamin EJ, Fonarow GC, *et al.*, American Heart Association Council on Epidemiology and Prevention; American Heart Association Council on Clinical Cardiology; American Heart Association Council on Cardiovascular Nursing, *et al.* Prevention of heart failure: a scientific statement from the American Heart Association Councils on Epidemiology and Prevention, Clinical Cardiology, Cardiovascular Nursing, and High Blood Pressure Research; Quality of Care and Outcomes Research Interdisciplinary Working Group; and Functional Genomics and Translational Biology Interdisciplinary Working Group. *Circulation* 2008;117:2544–65.
- Mosterd A, Hoes AW. Clinical epidemiology of heart failure. *Heart* 2007;93:1137–46.
- Mulrow CD, Lucey CR, Farnett LE. Discriminating causes of dyspnea through clinical examination. *J Gen Intern Med* 1993;8:383–92.
- van Mourik Y, Rutten FH, Moons KG, *et al.* Prevalence and underlying causes of dyspnoea in older people: a systematic review. *Age Ageing* 2014;43:319–26.
- Rutten FH, Moons KG, Cramer MJ, *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.
- 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;24:628–35.
- Oudejans I, Mosterd A, Bloemen JA, *et al.* Clinical evaluation of geriatric outpatients with suspected heart failure: value of symptoms, signs, and additional tests. *Eur J Heart Fail* 2011;13:518–27.
- Kelder JC, Cramer MJ, van Wijngaarden J, *et al.* The diagnostic value of physical examination and additional testing in primary care patients with suspected heart failure. *Circulation* 2011;124:2865–73.
- Roalfe AK, Mant J, Doust JA, *et al.* 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) rule. *Eur J Heart Fail* 2012;14:1000–8.
- Steyerberg EW. Chapter 20 Updating for a new setting. In: *Clinical prediction models*. Springer, 2009:361–89.
- Rose GA, Blackburn H. *Cardiovascular survey methods*. Geneva: World Health Organisation, 1982.
- McMurray JJ, 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. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *Eur J Heart Fail* 2012;14:803–69.
- Lang RM, Bierig M, Devereux RB, *et al.* Recommendations for chamber quantification: a report from the American Society of Echocardiography's Guidelines and Standards Committee and the Chamber Quantification Writing Group, developed in conjunction with the European Association of Echocardiography, a branch of the European Society of Cardiology. *J Am Soc Echocardiogr* 2005;18:1440–63.
- Nagueh SF, Appleton CP, Gillebert TC, *et al.* Recommendations for the evaluation of left ventricular diastolic function by echocardiography. *J Am Soc Echocardiogr* 2009;22:107–33.
- Donders AR, van der Heijden GJ, Stijnen T, *et al.* Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006;59:1087–91.
- de Groot JA, Janssen KJ, Zwiderman AH, *et al.* Correcting for partial verification bias: a comparison of methods. *Ann Epidemiol* 2011;21:139–48.
- Kenward MG, Carpenter J. Multiple imputation: current perspectives. *Stat Methods Med Res* 2007;16:199–218.
- Bodner TE. What improves with increased missing data imputations? *Struct Equation Model Multidisciplinary J* 2008;15:651–75.
- Rubin DB. *Multiple imputation for non response in surveys*. Wiley, 1987.
- Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
- Janssen KJ, Moons KG, Kalkman CJ, *et al.* Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol* 2008;61:76–86.
- Marshall A, Altman DG, Holder RL, *et al.* Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med Res Methodol* 2009;9:57.
- Clark TG, Altman DG. Developing a prognostic model in the presence of missing data. *J Clin Epidemiol* 2003;56:28–37.
- Bollen KA. *Structural equations with latent variables*. Wiley, 1989.
- Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 2011;30:11–21.
- Boonman-de Winter LJ, Rutten FH, Cramer MJ, *et al.* High prevalence of previously unknown heart failure and left ventricular dysfunction in patients with type 2 diabetes. *Diabetologia* 2012;55:2154–62.
- Zaphiriou A, Robb S, Murray-Thomas T, *et al.* 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;7:537–41.
- Grobbbee DE, Hoes AW. Chapter 3 diagnostic research. In: *Clinical epidemiology*. Jones and Bartlett Publishers, 2009:28–102.
- Rutten FH, Cramer MJ, Lammers JW, *et al.* Heart failure and chronic obstructive pulmonary disease: an ignored combination? *Eur J Heart Fail* 2006;8:706–11.
- Paulus WJ, van Ballegoij JJ. Treatment of heart failure with normal ejection fraction: an inconvenient truth! *J Am Coll Cardiol* 2010;55:526–37.
- Toll DB, Janssen KJ, Vergouwe Y, *et al.* Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol* 2008;61:1085–94.