

Diagnostic rules and algorithms for the diagnosis of non-acute heart failure in patients 80 years of age and older: a diagnostic accuracy and validation study

Miek Smeets,¹ Jan Degryse,^{1,2} Stefan Janssens,³ Catharina Mathei,¹ Pierre Wallemacq,⁴ Jean-Louis Vanoverschelde,⁵ Bert Aertgeerts,¹ Bert Vaes^{1,2}

To cite: Smeets M, Degryse J, Janssens S, et al. Diagnostic rules and algorithms for the diagnosis of non-acute heart failure in patients 80 years of age and older: a diagnostic accuracy and validation study. *BMJ Open* 2016;6:e012888. doi:10.1136/bmjopen-2016-012888

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2016-012888>).

Received 1 June 2016
Revised 7 September 2016
Accepted 9 September 2016



CrossMark

For numbered affiliations see end of article.

Correspondence to
Dr Miek Smeets; Miek.Smeets@kuleuven.be

ABSTRACT

Objectives: Different diagnostic algorithms for non-acute heart failure (HF) exist. Our aim was to compare the ability of these algorithms to identify HF in symptomatic patients aged 80 years and older and identify those patients at highest risk for mortality.

Design: Diagnostic accuracy and validation study.

Setting: General practice, Belgium.

Participants: 365 patients with HF symptoms aged 80 years and older (BELFRAIL cohort). Participants underwent a full clinical assessment, including a detailed echocardiographic examination at home.

Outcome measures: The diagnostic accuracy of 4 different algorithms was compared using an intention-to-diagnose analysis. The European Society of Cardiology (ESC) definition of HF was used as the reference standard for HF diagnosis. Kaplan-Meier curves for 5-year all-cause mortality were plotted and HRs and corresponding 95% CIs were calculated to compare the mortality risk predicting abilities of the different algorithms. Net reclassification improvement (NRI) was calculated.

Results: The prevalence of HF was 20% (n=74). The 2012 ESC algorithm yielded the highest sensitivity (92%, 95% CI 83% to 97%) as well as the highest referral rate (71%, n=259), whereas the Oudejans algorithm yielded the highest specificity (73%, 95% CI 68% to 78%) and the lowest referral rate (36%, n=133). These differences could be ascribed to differences in N-terminal probrain natriuretic peptide cut-off values (125 vs 400 pg/mL). The Kelder and Oudejans algorithms exhibited NRIs of 12% (95% CI 0.7% to 22%, p=0.04) and 22% (95% CI 9% to 32%, p<0.001), respectively, compared with the ESC algorithm. All algorithms detected patients at high risk for mortality (HR 1.9, 95% CI 1.4 to 2.5; Kelder) to 2.3 (95% CI 1.7 to 3.1; Oudejans). No significant differences were observed among the algorithms with respect to mortality risk predicting abilities.

Conclusions: Choosing a diagnostic algorithm for non-acute HF in elderly patients represents a trade-off between sensitivity and specificity, mainly depending on differences between cut-off values for natriuretic peptides.

Strengths and limitations of this study

- Community studies involving elderly patients are scarce, making this study particularly valuable.
- All variables needed to evaluate diagnostic performance of the different rules and algorithms were collected in the BELFRAIL study, including a detailed description of echocardiographic variables.
- No distinction was made between heart failure with reduced ejection fraction and heart failure with preserved ejection fraction as both are important to detect.
- All patients with heart failure symptoms were included, not just those suspected of heart failure by their general practitioner (GP), since suspicion is subjective and could lead to misclassification.
- Clinical assessment was performed by the personal GP of each patient, which may have increased heterogeneity, although GPs were trained to limit this risk.

INTRODUCTION

Early diagnosis of heart failure is important to initiate treatment in a timely fashion, as it may reduce mortality, hospitalisations and healthcare costs.¹ However, when access to echocardiography is limited, such as primary care, diagnosing heart failure in elderly individuals is challenging.^{1–2} Non-acute heart failure with a gradual onset of symptoms is especially often underdiagnosed or detected in a later stage.^{3–4} Primary care physicians are in need of a reliable diagnostic algorithm that helps them to decide which older symptomatic patients they must refer.

Different diagnostic rules and algorithms to diagnose non-acute heart failure do exist and all of them incorporate natriuretic peptides.^{5–8} However, since natriuretic peptide levels increase with age and are influenced

by comorbidities, controversy remains regarding optimal cut-off values among elderly individuals.^{9–10} Additionally, Mant *et al*¹¹ incorporated clinical characteristics in their algorithm.⁶ This algorithm has not proven more beneficial than algorithms using only natriuretic peptides.^{12–13} Two new diagnostic algorithms have been published^{7–8} including an algorithm specifically designed for elderly patients.⁸ However, the ability of these algorithms to identify non-acute heart failure in symptomatic patients aged 80 years and older has never been compared. Furthermore, it is unknown whether these diagnostic algorithms differ with respect to the identification of participants at high risk for mortality.

Therefore, an ‘intention-to-diagnose’ analysis¹⁴ was performed within the BELFRAIL cohort and the ability of the different diagnostic rules and algorithms to, first, confirm the diagnosis of heart failure in very old symptomatic patients and, second, to identify participants at high risk for mortality.

METHODS

Study population

The BELFRAIL (BF_{C80+}) study is a prospective, observational, population-based cohort study of participants aged 80 years and older living in three well-circumscribed areas of Belgium. The study’s design, sampling methods and cohort characteristics have been described previously.¹⁵ In brief, a total of 29 general practice centres participated in the BELFRAIL study. The participating centres were asked to include patients aged 80 years and older in the cohort. Only three exclusion criteria were used: (1) severe dementia (known Mini-Mental State Examination (MMSE) <15/30), (2) undergoing palliative care or (3) a medical emergency like acute heart failure. Two sampling methods for the recruitment of patients were used. Two general practice centres were asked to include all eligible patients. The remaining 27 centres were asked to include a maximum of three consecutive patients during a 3-week interval. In these 3 weeks, the general practitioners (GPs) also planned their visit. This interval was repeated five times, so every participating centre included a maximum of 15 patients. Every interval of recruitment was started on a different day to avoid selection bias. Between 2 November 2008 and 15 September 2009, 567 participants were included in the BELFRAIL cohort. Three hundred participants were included using the first sampling method, and 267 using the second sampling method. Out of 567 patients, 365 (70%) presented with heart failure symptoms such as dyspnoea (Medical Research Council (MRC) ≥2), fatigue or ankle oedema. A detailed follow-up regarding cause-specific mortality was collected from the participants’ GPs until 5.2±0.25 years after baseline. The causes of death were divided into cardiovascular and non-cardiovascular. The BELFRAIL cohort is representative in gender and age of the very elderly living in Belgium.¹⁵ STARD recommendations for reporting of diagnostic accuracy studies were followed.¹⁶

Clinical assessment

Each GP recorded a full medical history and performed a detailed, standardised clinical examination at baseline. The analyses included questions regarding dyspnoea (according to the MRC dyspnoea scale),¹⁷ fatigue, orthopnoea, oedema of the lower extremities, wheezing and loss of appetite. The clinical examination consisted of a heart auscultation, a heart rate and breathing rate measurement. The apical beat was palpated and recorded when abnormal. Lung auscultation was performed to detect crepitus. Jugular venous pressure (JVP) was measured and noted if elevated, and the presence of hepatojugular reflux (HJR) and oedema of the lower extremities was checked. Body mass index (BMI) was calculated by a clinical research assistant based on a standardised measurement of height and weight. The patients’ GPs reported a history of myocardial infarction and important cardiac interventions such as percutaneous transluminal coronary angioplasty, stenting or coronary surgery. The Anatomic Therapeutic Chemistry classification system was used to register medication use.¹⁸ Data regarding loop diuretics were used for the analyses.

The serum N-terminal probrain natriuretic peptide (NT-proBNP) levels were measured using a Dade-Dimension Xpand (Siemens, Deerfield, Illinois, USA). The coefficient of variation ranged from 3.9% to 4.3%.

A 12-lead ECG was recorded by a clinical research assistant at baseline using a QRS Universal ECG device (QRS diagnostics, Plymouth, Minnesota, USA, <http://www.qrsdiagnostic.com>). A single cardiologist blinded to all other study results analysed each of the ECGs according to the Minnesota Code Classification System for Electrocardiographic Findings and was asked to report whether each patient’s ECG findings were completely normal (yes or no).

Diagnostic algorithms

Four diagnostic algorithms were chosen because they applied to patients aged 80 years or older with non-acute heart failure in primary care. The 2012 European Society of Cardiology (ESC) diagnostic algorithm⁵ and the diagnostic rules proposed by Mant *et al*,⁶ Kelder *et al*⁷ and Oudejans *et al*⁸ were studied (see online supplementary appendix 1). The ESC algorithm (NT-proBNP≥125 ng/mL or positive ECG) was studied both with and without the inclusion of patients’ ECG results.⁵ The diagnostic rule by Mant was evaluated by considering direct referral for echocardiography of patients with a history of myocardial infarction or basal crepitus or men with ankle oedema. In the remaining patients, referral for echocardiography was considered for women without ankle oedema and NT-proBNP levels above 620 pg/mL, women with ankle oedema and NT-proBNP levels above 190 pg/mL and men with NT-proBNP levels above 390 pg/mL.⁶ Using the diagnostic rule of Kelder, heart failure was ruled out if the summed score was <24 (<20% probability of heart

failure).⁷ A score between 24 and 54 was considered uncertain (between 20% and 70% probability of heart failure). For patients with a score above 54, referral for echocardiography was considered (>70% probability of heart failure). The diagnostic rule of Oudejans considered heart failure unlikely if the summed score was 16 or less. Heart failure was considered likely with a score of 32 or more and uncertain with a score between 16 and 32.⁸

Echocardiography

Echocardiograms were performed at baseline using a commercially available portable system (CX50, Philips, Andover, Massachusetts, USA) with M-mode, two-dimensional and pulsed, continuous-wave and colour-flow Doppler capabilities. The echocardiograms were performed at the participants' homes by a single cardiologist blinded to the patients' clinical characteristics and laboratory test results. A complete examination was performed in accordance with the recommendations of the American Society of Echocardiography and the European Association of Echocardiography (ASE-EAE).¹⁹ The methods, prevalences of the echocardiographic abnormalities and quality of the echocardiographic images were described previously.²⁰ Briefly, left ventricular (LV) systolic function was calculated using Simpson's biplane method.¹⁹ Systolic dysfunction was defined as an LV ejection fraction (LVEF) $\leq 50\%$. The functions of the mitral valve and the aortic valve were evaluated using colour Doppler echocardiography after optimising the gain and Nyquist limit. Stenotic and regurgitant valve diseases were evaluated using semi-quantitative and quantitative methods recommended by the ASE.^{21 22} Participants with prosthetic valves were evaluated separately.²¹ Clinically relevant valve disease was defined as mitral stenosis of any severity, severe aortic stenosis (aortic valve area $< 1 \text{ cm}^2$), moderate or severe mitral regurgitation, and moderate or severe aortic regurgitation. Diastolic function was assessed using mitral flow velocities obtained via pulsed Doppler and pulsed tissue Doppler at the level of the mitral annulus.²³ Additional apical and parasternal views were also recorded to assess tissue velocity (colour tissue Doppler). The ASA-EAE guidelines were used to assess diastolic dysfunction.²³

Reference standard

Heart failure defined according to the ESC guideline based on a combination of heart failure symptoms and signs (**table 1**) and objective cardiac dysfunction was used as the reference standard.⁵ Objective cardiac dysfunction included a LVEF $\leq 50\%$; clinically relevant valvular heart disease and severe diastolic dysfunction, as used in previous BELFRAIL publications.^{14 20}

Data analysis

Continuous variables are presented either as means \pm SDs or as medians and IQRs. Categorical data are presented

as frequencies and proportions. Baseline variables were compared using the χ^2 test or the independent Student's t-test to compare means, and the Mann-Whitney U test to compare medians. p Values below 0.05 were considered statistically significant. The diagnostic accuracy (the sum of true negatives and true positives divided by the total number of participants) of the different algorithms and rules for heart failure was calculated using an intention-to-diagnose analysis. This method, analogous to an 'intention-to-treat' analysis, starts from a diagnostic intent, at any given moment, in a population at risk. This means that all patients who are at risk of a specific target condition should be involved in the analyses, regardless of all previously known diagnoses.¹⁴ Kaplan-Meier curves for all-cause mortality after 5 years were plotted for the different diagnostic algorithms, with log-rank tests used for comparisons. HRs and corresponding 95% CIs were calculated using Cox proportional hazard models. The net reclassification improvement (NRI) was calculated for Oudejans', Kelder's and Mant's diagnostic rules and compared with the ESC algorithm for the diagnosis of heart failure and the prediction of 5-year all-cause and cardiovascular mortality in order to measure and compare the ability of the rules to detect participants at high risk for mortality. All data analyses were performed using SPSS V.22.0 for Windows (SPSS, Chicago, Illinois, USA).

RESULTS

In order to validate the existing rules, only participants with heart failure symptoms such as dyspnoea (MRC ≥ 2), fatigue or ankle oedema were studied (n=365, 70%). The mean age of the study participants was 85 ± 3.8 years, 123 of whom were men (34%). A total of 40 patients were institutionalised (11%). Heart failure was present in 74 patients (20%).

Table 1 includes the distribution of the different variables used in the diagnostic algorithms and rules. Patients with heart failure were older and presented more often with orthopnoea, ankle oedema in men, an irregular pulse, systolic heart murmur, elevated JVP or HJR. Loss of appetite and a lower BMI, proposed as clinical markers in the elderly (see online supplementary appendix 1), were not statistically different across the two patient groups.

Prediction of heart failure

All rules and algorithms exhibited a strong ability to exclude heart failure. However, a high number of unnecessary referrals were noted.

The 2012 ESC diagnostic algorithm

The 2012 ESC diagnostic algorithm proposes to refer patients suspected of heart failure based on an abnormal ECG or elevated natriuretic peptide levels (NT-proBNP $\geq 125 \text{ pg/mL}$; see online supplementary appendix 1).⁵ However, the proportion of patients with

Table 1 Variables used in the diagnostic algorithms and rules according to the presence of heart failure

	All n=365	No heart failure n=291	Heart failure n=74	p Value*
<i>Sociodemographic characteristics</i>				
Male gender, n (%)	123 (34)	93 (32)	30 (41)	0.16
Age, mean±SD	85±3.8	85±3.7	86±4.0	0.004†
80 years, n (%)	50 (14)	42 (14)	8 (11)	0.42
81–85 years, n (%)	178 (49)	151 (52)	27 (37)	0.018
86–90 years, n (%)	108 (30)	78 (27)	30 (41)	0.021
≥91 years, n (%)	29 (8)	20 (6.9)	9 (12)	0.13
<i>Clinical characteristics</i>				
Symptoms				
Nocturnal dyspnoea or orthopnoea, n (%)	29 (7.9)	18 (6.2)	11 (15)	0.014
Wheezing, n (%)	49 (13)	41 (14)	8 (11)	0.45
Loss of appetite, n (%)	51 (14)	42 (14)	9 (12)	0.62
Ankle oedema, n (%)	176 (48)	137 (47)	39 (53)	0.39
Men, n (%)	49 (13)	34 (12)	17 (23)	0.012
Women, n (%)	127 (35)	104 (36)	23 (31)	0.45
Signs				
Displaced apical beat, n (%)	22 (6.0)	15 (5.2)	7 (9.5)	0.17
Basal crepitus lung, n (%)	56 (15)	46 (16)	10 (18)	0.63
Irregular pulse, n (%)	51 (14)	31 (11)	20 (27)	<0.001
Systolic heart murmur, n (%)	131 (36)	81 (28)	50 (68)	<0.001
Heart frequency, mean±SD	69±10	69±10	67±10	0.32†
Elevated jugular venous pressure, n (%)	46 (13)	26 (8.9)	20 (27)	<0.001
Tachypnoea‡, n (%)	59 (16)	44 (15)	15 (20)	0.28
Hepatojugular reflux, n (%)	47 (13)	24 (8.2)	23 (31)	<0.001
Peripheral oedema on clinical examination, n (%)	158 (43)	118 (41)	40 (54)	0.036
BMI (kg/m^2), mean±SD	28±4.9	28±4.8	28±5.3	0.63†
<20, n (%)	15 (4.1)	12 (4.1)	3 (4.1)	0.98
20.0–24.9, n (%)	88 (24)	69 (24)	19 (26)	0.72
25.0–29.9, n (%)	137 (38)	111 (38)	26 (35)	0.63
30.0–34.9, n (%)	99 (27)	80 (28)	19 (26)	0.75
≥35.0, n (%)	26 (7.1)	19 (6.5)	7 (9.5)	0.38
<i>Cardiovascular antecedents</i>				
MI, n (%)	46 (13)	34 (12)	12 (16)	0.29
MI, CABG or PCI, n (%)	70 (19)	49 (17)	21 (28)	0.024
<i>ECG</i>				
Positive ECG§, n (%)	238 (65)	177 (63)	61 (85)	0.001
<i>Medication</i>				
Loop diuretics, n (%)	80 (22)	54 (19)	26 (35)	0.002
NT-proBNP (pg/mL), median (IQR)	220 (110–593)	181 (100–394)	786 (350–1586)	<0.001¶
<400, n (%)	240 (66)	219 (75)	21 (28)	<0.001
400–2000, n (%)	102 (28)	63 (22)	39 (53)	<0.001
>2000, n (%)	23 (6.3)	9 (3.1)	14 (19)	<0.001

* χ^2 test.

†Independent samples t-test.

¶Mann-Whitney test.

‡Breathing frequency >20/min.

§Abnormal ECG as read by a cardiologist.

BMI, body mass index; CABG, coronary artery bypass graft surgery; MI, myocardial infarction; NT-proBNP, N-terminal probrain natriuretic peptide; PCI, percutaneous coronary intervention.

an abnormal ECG was high in this population of very elderly people (n=238, 65%). As a consequence, the ESC algorithm with ECG results resulted in a higher number of false positives (n=246, 77% of total referrals, specificity 15%) and a similar sensitivity compared with the ESC algorithm without ECG results. Therefore, the ESC algorithm without ECG results was chosen for subsequent analyses and considered an index test for the

calculation of the NRI (see online supplementary appendix 1).

By applying the ESC algorithm without ECG results ($\text{NT-proBNP} \geq 125 \text{ pg/mL}$) to our study population, only six patients with heart failure were missed, corresponding to a sensitivity of 92%. However, 191 patients (74% of total referrals) were referred despite the absence of heart failure (specificity 34%; tables 2 and 3).

Table 2 The test characteristics of the different rules and algorithms for the detection of heart failure

	Non-referral						Referral					
	Number of non-referrals n (%)	Heart failure absent n	Sensitivity (%), 95% CI	NPV (%), 95% CI	LR- 95% CI	n (%)	Number of referrals n (%)	Heart failure present n	Specificity (%), 95% CI	PPV (%), 95% CI	LR+ 95% CI	
ESC guideline, NT-proBNP ≥ 125 pg/mL ⁵	106 (29)	100	92 (83 to 97)	94 (88 to 98)	0.21 (0.18 to 0.52)	259 (71)	68	34 (29 to 40)	26 (21 to 32)	1.4 (1.3 to 1.6)		
Kelder* diagnostic rule ⁷	181 (50)	167	81 (70 to 89)	92 (87 to 96)	0.33 (0.20 to 0.53)	184 (50)	60	57 (52 to 63)	33 (26 to 40)	1.9 (1.6 to 2.3)		
Oudejans† diagnostic rule ⁸	232 (64)	213	74 (63 to 84)	92 (88 to 95)	0.35 (0.24 to 0.52)	133 (36)	55	73 (68 to 78)	41 (33 to 50)	2.8 (2.2 to 3.5)		
Mant et al/ clinical decision rule ⁶	158 (43)	143	80 (69 to 88)	91 (85 to 95)	0.41 (0.26 to 0.66)	207 (57)	59	49 (43 to 55)	29 (23 to 35)	1.6 (1.3 to 1.8)		

*Dichotomous variable with a cut-off >54.
†Dichotomous variable with a cut-off >16.

ESC, European Society of Cardiology; LR-, negative likelihood ratio; LR+, positive likelihood ratio; NPV, negative predictive value; NT-proBNP, N-terminal probrain natriuretic peptide; PPV, positive predictive value.

The diagnostic rule of Oudejans *et al*

By applying the lower score of the diagnostic rule by Oudejans *et al* (summed score >16), the highest number of patients with HF was missed (n=19, a sensitivity of 74%), but the rule also corresponded to the lowest number of unnecessary referrals (n=78, 59% of total referrals) and consequently the highest specificity (73%; tables 2 and 3).

Using the higher score of Oudejans' rule (summed score >32) resulted in only 29 false-positive cases (a specificity of 90%), of which 17 (59%) had at least one other echocardiographic abnormality. However, sensitivity dropped to 42% (n=43 patients with missed heart failure). Therefore, for subsequent analyses, the lower score of Oudejans' rule was used.

The rule of Oudejans yielded a significant NRI of 21% (95% CI 9% to 32%, p<0.001) in predicting heart failure compared with the ESC diagnostic algorithm (table 4). In Oudejans' rule, an NT-proBNP cut-off of 400 pg/mL is used in addition to clinical characteristics (see online supplementary appendix 1). However, the comparison between NT-proBNP as a stand-alone test (cut-off 400 pg/mL) and the full diagnostic rule of Oudejans regarding both diagnostic and prognostic accuracy yielded no significant differences (data not shown). Hence, the differences between the ESC algorithm and the diagnostic rule of Oudejans could be ascribed to differences in NT-proBNP cut-off values (125 vs 400 pg/mL).

The diagnostic rule of Kelder *et al*

By applying the higher score of Kelder's rule (summed score >54), 14 patients with heart failure were missed (sensitivity 81%). False positives were noted in 124 patients (67% of total referrals), corresponding to a specificity of 57%.

Using the lower score of Kelder's rule (summed score ≥ 24) led to zero false negatives (sensitivity 100%), but only 18 patients were not referred in that case (specificity 6%). Therefore, for subsequent analyses, the higher score of Kelder's rule was used.

The rule of Kelder also yielded a significant NRI of 12% (95% CI 0.7% to 22%, p=0.04) in predicting heart failure compared with the ESC diagnostic algorithm (table 4).

The diagnostic rule of Mant *et al*

Applying the diagnostic rule of Mant *et al* led to 15 patients with missed heart failure (sensitivity 80%) and 148 unnecessary referrals (71% of total referrals, specificity 59%). No significant NRI was noted compared with the ESC algorithm.

Five-year mortality predictions

During the 5-year follow-up period, 177 patients died (49%), including 78 (44%) who died of cardiovascular causes. There was no loss to follow-up for mortality. Of the patients with heart failure, 61% died (n=45). The

Table 3 Echocardiographic characteristics of patients according to the different decision rules and algorithms

	ESC guideline HF 2012 ^{4*}						Mant et al clinical decision rule ^{5§}		
	Non-referral n=106		Referral n=259		Kelder et al diagnostic rule ^{6†}		Oudejans et al diagnostic rule ^{7‡}		
	Non-referral n=18	Uncertain n=163	Referral n=184	Non-referral n=232	Uncertain n=73	Referral n=60	Non-referral n=158	Referral n=207	
Heart failure, n=74									
EF≤50, n (%)	1 (1.4)	14 (19)	0 (0)	1 (1.4)	14 (19)	4 (5.4)	4 (5.4)	7 (7.5)	2 (2.7)
Clinically relevant valvular disease,** n (%)	5 (6.8)	39 (53)	0 (0)	11 (15)	33 (45)	12 (16)	16 (22)	16 (22)	10 (14)
Severe diastolic dysfunction, n (%)	0 (0)	16 (22)	0 (0)	2 (2.7)	14 (19)	3 (4.1)	4 (5.4)	9 (12)	3 (4.1)
No heart failure, n=291									
No echocardiographic abnormalities, n=102	42 (41)	60 (59)	8 (2.7)	55 (19)	39 (13)	84 (29)	10 (3.4)	8 (2.7)	56 (19)
Other echocardiographic abnormalities, n=189	58 (31)	131 (69)	10 (5.3)	94 (50)	85 (45)	129 (44)	39 (13)	21 (7.2)	87 (30)
RWMA, n (%)	9 (4.7)	19 (10)	0 (0)	14 (7.4)	14 (7.4)	19 (10)	6 (3.2)	3 (1.6)	10 (5.3)
Left atrial enlargement,†† n (%)	30 (16)	76 (40)	3 (1.6)	50 (27)	53 (28)	65 (34)	25 (13)	16 (8.5)	46 (24)
Left ventricular hypertrophy,‡‡ n (%)	15 (7.9)	21 (11)	3 (1.6)	20 (11)	13 (6.9)	30 (16)	5 (2.6)	1 (0.5)	25 (13)
Estimated pulmonary artery pressure >30 mm Hg, n (%)	6 (4.6)	46 (35)	2 (1.1)	21 (11)	29 (15)	25 (19)	16 (12)	11 (8.5)	19 (15)
Moderate diastolic dysfunction, n (%)	21 (11)	46 (24)	6 (3.2)	32 (17)	29 (15)	51 (27)	11 (5.8)	5 (2.6)	35 (19)
									32 (17)

*Referral, patients with non-acute onset suspected heart failure if NT-proBNP≥125 pg/ml.

†Non-referral, probability of heart failure <20%, <24 points; uncertain, probability of heart failure 20–70%, 24–54 points; referral, probability of heart failure >70%, >54 points.

‡Non-referral, geriatric patients with slow onset heart failure <16 points; uncertain, 16–32 points; referral >32 points.

§Referral, a history of myocardial infarction, basal crepitus, ankle oedema in a male OR female gender without ankle oedema if NT-proBNP >620 pg/mL, male gender without ankle oedema if NT-proBNP>390 pg/mL, female gender with ankle oedema if NT-proBNP>190 pg/mL.

**Any grade of mitral stenosis, moderate/severe mitral or aortic regurgitation or severe aortic stenosis.

††Left atrial volume index≥34 mL/m².‡‡Left ventricular mass index (LVMI)≥109 g/m² in women, LVMI≥132 g/m² in men.

EF, ejection fraction; ESC, European Society of Cardiology; HF, heart failure; RWMA, regional wall motion abnormality.



proportion of cardiovascular deaths within this group was 60% (n=27). HR for all-cause mortality of heart failure was 1.6 (95% CI 1.2 to 2.3, p=0.004), and for cardiovascular mortality 2.5 (95% CI 1.6 to 4.0, p<0.001). The algorithms and diagnostic rules successfully identified patients at risk for mortality (figure 1). HRs for all-cause mortality ranged from 1.9 (95% CI 1.4 to 2.5; Kelder) to 2.3 (95% CI 1.7 to 3.1; Oudejans), and from 1.7 (95% CI 1.0 to 2.6; Mant) to 2.9 (95% CI 1.9 to 4.5; Oudejans) for cardiovascular mortality.

The improvements in mortality risk classification of the diagnostic rules of Kelder and Mant were only minor compared with the ESC algorithm. The diagnostic rule of Oudejans exhibited a larger degree of

improvement, particularly with respect to cardiovascular mortality risk classification, although the said improvement was not statistically significant (NRI 10% (95% CI -3% to 22%, p=0.18; table 4).

DISCUSSION

The present study demonstrated that all diagnostic algorithms have their limitations in symptomatic patients aged 80 years and older. Therefore, choosing a diagnostic algorithm for non-acute HF in elderly patients represents a trade-off between sensitivity and specificity, mainly depending on differences between used cut-off values for natriuretic peptides. The ESC algorithm

Table 4 Reclassification table and NRI* for the identification of heart failure and the prediction of 5-year mortality

	All	Reclassified up	Reclassified down	NRI (95% CI; p Value)
<i>Kelder diagnostic rule⁷†</i>				
All-cause mortality				
Dead, n (%)	177	10 (5.6)	44 (25)	-0.19 (-0.27 to -0.11; <0.001)
Alive, n (%)	188	8 (3.7)	49 (26)	-0.22 (-0.29 to -0.14; <0.001)
Total, n (%)	365	18 (4.9)	93 (25)	0.026 (-0.075 to 0.13; 0.65)
Cardiovascular mortality				
Dead, n (%)	78	2 (2.6)	20 (26)	-0.23 (-0.34 to -0.13; <0.001)
Alive, n (%)	287	16 (5.6)	73 (25)	-0.20 (-0.31 to -0.19; <0.001)
Total, n (%)	365	18 (4.9)	93 (25)	-0.03 (-0.15 to 0.09; 0.64)
Heart failure				
Present, n (%)	80	2 (2.7)	10 (14)	-0.11 (-0.21 to -0.02; 0.001)
Absent, n (%)	285	16 (5.5)	83 (29)	-0.23 (-0.29 to -0.17; <0.001)
Total, n (%)	365	18 (4.9)	93 (25)	0.12 (0.007 to 0.22; 0.04)
<i>Oudejans diagnostic rule^{8‡}</i>				
All-cause mortality				
Dead, n (%)	177	2 (1.1)	57 (32)	-0.31 (-0.38 to -0.24; <0.001)
Alive, n (%)	188	1 (0.5)	72 (38)	-0.38 (-0.45 to -0.31; <0.001)
Total, n (%)	365	3 (0.8)	129 (35)	0.07 (-0.03 to 0.17; 0.28)
Cardiovascular mortality				
Dead, n (%)	78	1 (1.3)	22 (28)	-0.27 (-0.38 to -0.16; <0.001)
Alive, n (%)	287	2 (0.7)	107 (37)	-0.37 (-0.42 to -0.31; <0.001)
Total, n (%)	365	3 (0.8)	129 (35)	0.10 (-0.03 to 0.22; 0.18)
Heart failure				
Present, n (%)	74	1 (1.3)	14 (19)	-0.18 (-0.28 to -0.08; <0.001)
Absent, n (%)	291	2 (0.7)	115 (40)	-0.39 (-0.45 to -0.33; <0.001)
Total, n (%)	365	3 (0.8)	129 (35)	0.21 (0.09 to 0.32; <0.001)
<i>Diagnostic rule of Mant et al⁶</i>				
All-cause mortality				
Dead, n (%)	177	12 (6.8)	32 (18)	-0.11 (-0.18 to -0.04; 0.002)
Alive, n (%)	188	19 (10)	51 (27)	-0.17 (-0.25 to -0.09; <0.001)
Total, n (%)	365	31 (8)	83 (23)	0.06 (-0.05 to 0.16; 0.32)
Cardiovascular mortality				
Dead, n (%)	78	1 (1.3)	15 (19)	-0.18 (-0.28 to -0.09; <0.001)
Alive, n (%)	287	30 (10)	68 (24)	-0.13 (-0.19 to -0.07; <0.001)
Total, n (%)	365	31 (8)	83 (23)	-0.05 (-0.17 to 0.06; 0.52)
Heart failure				
Present, n (%)	74	2 (2.7)	11 (15)	-0.12 (-0.22 to -0.03; 0.005)
Absent, n (%)	291	29 (10)	72 (25)	-0.15 (-0.21 to -0.09; <0.001)
Total, n (%)	365	31 (8)	83 (23)	0.026 (-0.09 to 0.13; 0.66)

*Comparison between the rules of Kelder, Oudejans and Mant with the ESC 2012 diagnostic algorithm.

†Dichotomous variable with a cut-off >54.

‡Dichotomous variable with a cut-off >16.

ESC, European Society of Cardiology; NRI, net reclassification improvement.

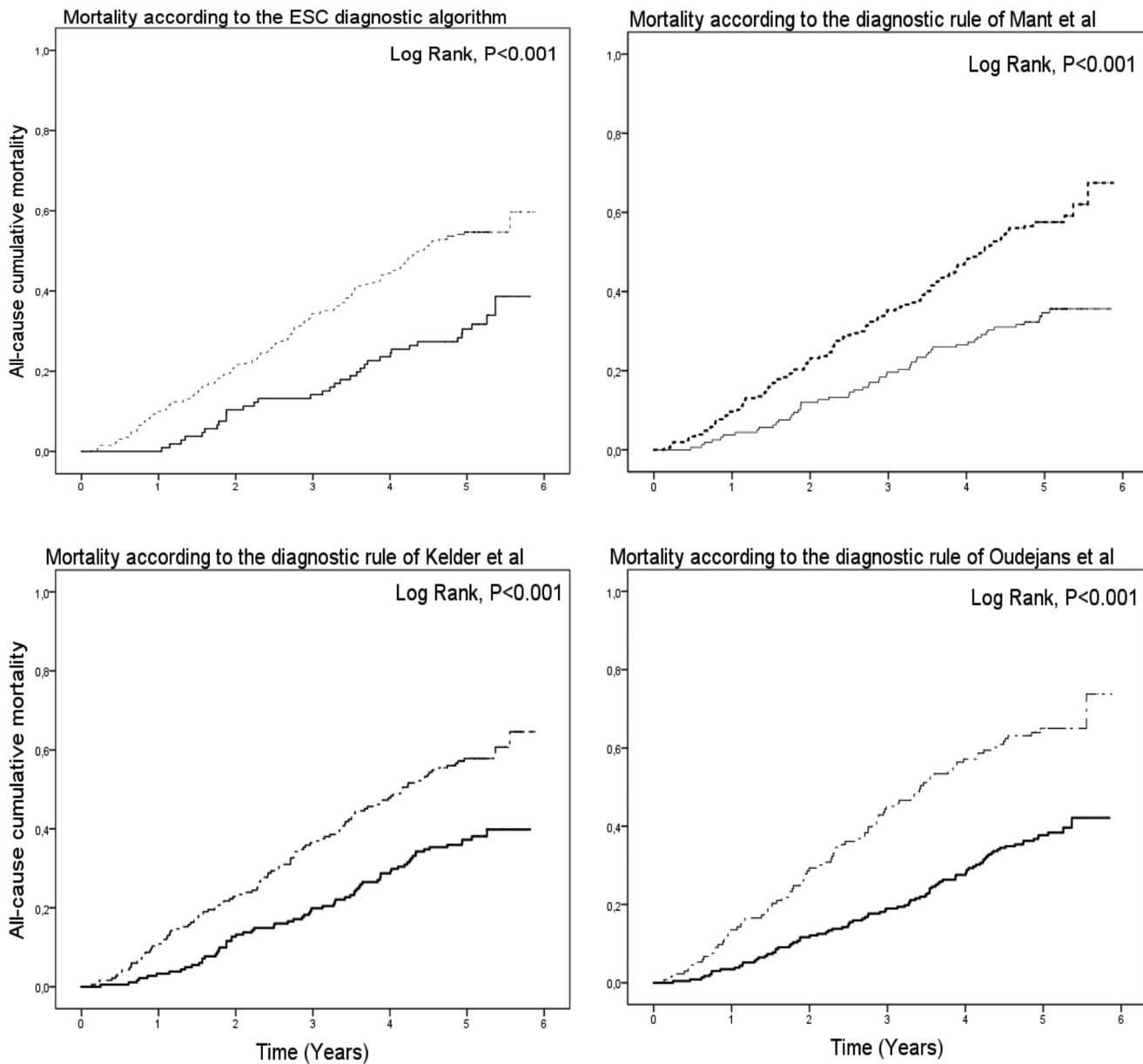


Figure 1 All-cause mortality risk stratification based on the diagnostic algorithms and rules. ESC, European Society of Cardiology.

exhibited the highest sensitivity. However, implementing this algorithm in this age group also resulted in high referral rates. In contrast, the diagnostic rule of Oudejans exhibited the highest accuracy with reasonable referral rates (36%, n=133). The said rule was a superior predictor of heart failure compared with the ESC algorithm because of a strong gain in specificity, however, in spite of a decrease in sensitivity. Furthermore, all algorithms were able to identify patients at high risk for 5-year mortality.

The prevalence of heart failure in this population-based cohort study of symptomatic elderly individuals was 20%, as could be expected in this age group.²⁴ The 2008 ESC algorithm with an NT-proBNP cut-off point of 400 pg/mL, the clinical decision rule of Mant and the

Dutch guideline (either NT-proBNP \geq 125 pg/mL or a positive ECG) were validated by Oudejans *et al*¹² in a cohort of geriatric outpatients (patients referred to a geriatrician on suspicion of HF). The prevalence of heart failure was 45% in this cohort, resulting in higher negative predictive values compared with this study (Mant *et al*: 95% vs 91%), as well as higher positive predictive values (Mant *et al*: 62% vs 29%) and higher referral rates (Mant *et al*: 70% vs 57%). They concluded that the 2008 ESC algorithm performed the best.¹² However, the Newcastle 85+ study demonstrated, in line with our results, that both proposed NT-proBNP cut-off points (125 and 400 pg/mL) had their limitations in an elderly population. Their used reference standard was LV dysfunction.¹³ However, both studies agreed that algorithms

using natriuretic peptide levels alone performed better than the algorithms incorporating patients' clinical characteristics.^{12 13}

For clinicians, the decision regarding the referral of patients suspected of having a diagnosis of heart failure will always be a trade-off between the potential benefits afforded by additional investigations and the burdens the said investigations impose on affected patients.¹² In elderly patients, this choice is more difficult, as the benefits may be smaller,^{5 25 26} and the burdens larger. Therefore, large numbers of false positives should be avoided, a scenario made possible with the use of the diagnostic rule of Oudejans. On the other hand, effective treatments exist for heart failure characterised by either a reduced ejection fraction or valvular heart disease,^{5 25} highlighting the importance of reducing false-negative cases, an area in which the ESC algorithm excels. However, no differences in the identification of patients at high risk for 5-year mortality were observed among the different rules and algorithms. Another difficulty with an elderly population is the high prevalence of echocardiographic abnormalities overall. Only 25–35% of the false-positive cases did not exhibit any echocardiographic abnormalities. Additional research is warranted regarding the clinical significance of these cardiac phenotypes among elderly patients.

NT-proBNP levels played a crucial role in all decision rules and algorithms. Oudejans selected 400 pg/mL as a cut-off point, whereas the ESC algorithm selected 125 pg/mL. Since the comparison between NT-proBNP as a stand-alone test (cut-off 400 pg/mL) and the full diagnostic rule of Oudejans regarding both diagnostic and prognostic accuracy yielded no significant differences, the distinction between the ESC algorithm and the diagnostic rule of Oudejans could be ascribed to differences in NT-proBNP cut-off values. The diagnostic value of natriuretic peptides in heart failure has been proved, even in the elderly.^{14 27} However, controversy remains about optimal NT-proBNP cut-off values among elderly patients.^{9 10 13 28–30} Previous studies have advocated the use of age-dependent values.^{10 29 30} However, these values were highly dependent on settings, populations and reference standards, and warrant further investigation. Although the benefits of extensive diagnostic rules with clinical variables were not proven, it is important to note that natriuretic peptide levels must never be seen as a stand-alone test. A thorough clinical assessment must always take precedence.

This was the first comprehensive study to simulate and compare the ability of the diagnostic rules of Oudejans and Kelder to detect heart failure among symptomatic elderly patients. Additionally, this was the first study to describe and compare the abilities of different algorithms and clinical decision rules in identifying patients at high risk for mortality. Moreover, community studies involving elderly patients are scarce, which made our data particularly valuable. However, a few limitations of this study should be noted. First, heart failure was

chosen as the reference standard and defined according to the ESC definition of heart failure. No difference was made between heart failure with reduced ejection fraction and heart failure with preserved ejection fraction. Although these entities have different therapeutic implications, both are important to detect. Debate is possible about which echocardiographic abnormalities should be regarded as objectified cardiac dysfunction. Therefore, a nuanced presentation of all echocardiographic abnormalities present in the group of patients with and without heart failure was given in table 3. Second, all patients with heart failure symptoms were included, not just those suspected of heart failure by their GPs. This approach was chosen since suspicion of heart failure by GPs is subjective and patients could be misclassified based on suspicion. Third, the clinical assessment was performed by each patient's personal physician, which may have increased the heterogeneity of our data. However, all participating GPs were trained, and the assessments of their patients were standardised to limit the risk of heterogeneity as much as possible.¹⁵

In conclusion, in a cohort of symptomatic patients aged 80 years and older, all diagnostic algorithms had their limitations. Therefore, choosing a diagnostic algorithm for non-acute HF in elderly patients represents a trade-off between sensitivity (2012 ESC algorithm) and specificity (diagnostic rule of Oudejans *et al*), depending on differences between used cut-off values for natriuretic peptides. The rules and algorithms were all able to detect patients at high risk for mortality without significant differences between the rules. Further research is warranted to identify strategies that optimise both sensitivity and specificity in a heterogeneous population of elderly patients.

Author affiliations

¹Department of Public Health and Primary Care, KU Leuven (KUL), Leuven, Belgium

²Institute of Health and Society, Université Catholique de Louvain (UCL), Brussels, Belgium

³Department of Cardiovascular Diseases, Universitair Ziekenhuis Gasthuisberg, KU Leuven (KUL), Leuven, Belgium

⁴Laboratory of Analytical Biochemistry, Cliniques Universitaires St Luc, Université Catholique de Louvain (UCL), Brussels, Belgium

⁵Department of Cardiology, Cliniques Universitaires Saint-Luc, Université Catholique de Louvain (UCL), Brussels, Belgium

Acknowledgements This study was made possible by the participating GPs.

Contributors All named authors have made substantial contributions to the conception and design of the BELFRAIL study and have seen and approved the final version of the manuscript. MS analysed and interpreted the data and drafted the work. BV and JD acquired and interpreted the data and critically revised the work. PW and J-LV acquired the data and critically revised the work. SJ, CM and BA critically revised the work.

Funding The BELFRAIL study (B40320084685) was supported by an unconditional grant from the Foundation Louvain. The Foundation Louvain is the support unit of the Université Catholique de Louvain and is charged with developing the educational and research projects of the university by collecting gifts from corporations, foundations and alumni.

Competing interests None declared.



Patient consent Obtained.

Ethics approval All participants provided informed consent, and the Biomedical Ethics Committee of the Medical School of the Université Catholique de Louvain (UCL) of Brussels approved the study. All clinical and laboratory tests were performed in accordance with the Declaration of Helsinki.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All authors had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

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

- Fonseca C. Diagnosis of heart failure in primary care. *Heart Fail Rev* 2006;11:95–107.
- Smeets M, Degryse J, Aertgeerts B, et al. General practitioners' judgement of chronic heart failure in the oldest old: insights from the BELFRAIL study. *Int J Cardiol* 2015;191:120–7.
- Devroey D, Van Casteren V. The incidence and first-year mortality of heart failure in Belgium: a 2-year nationwide prospective registration. *Int J Clin Pract* 2010;64:330–5.
- Hancock HC, Close H, Mason JM, et al. High prevalence of undetected heart failure in long-term care residents: findings from the Heart Failure in Care Homes (HFinCH) study. *Eur J Heart Fail* 2013;15:158–65.
- 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.
- Tait L, Roalfe AK, Mant J, et al. 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. Rationale and design. *BMC Cardiovasc Disord* 2012;12:97.
- 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.
- 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.
- Vaes B, Gruson D, Van Pottelbergh G, et al. The impact of confounders on the test performance of natriuretic peptides for cardiac dysfunction in subjects aged 80 and older. *Peptides* 2012;38:118–26.
- Hildebrandt P, Collinson PO, Doughty RN, et al. Age-dependent values of N-terminal pro-B-type natriuretic peptide are superior to a single cut-point for ruling out suspected systolic dysfunction in primary care. *Eur Heart J* 2010;31:1881–9.
- 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;13:1–207, iii.
- Oudejans I, Mosterd A, Zuijthoff NP, et al. Applicability of current diagnostic algorithms in geriatric patients suspected of new, slow onset heart failure. *Age Ageing* 2012;41:309–16.
- Collerton J, Kingston A, Yousaf F, et al. Utility of NT-proBNP as a rule-out test for left ventricular dysfunction in very old people with limiting dyspnoea: the Newcastle 85+ Study. *BMC Cardiovasc Disord* 2014;14:128.
- Vaes B, Boland B, Scavee C, et al. Value of adding natriuretic peptides and electrocardiographic findings to assess the presence of cardiac dysfunction in patients >/=80 years of age. *Am J Cardiol* 2013;111:1198–208.
- Vaes B, Pasquet A, Wallemacq P, et al. The BELFRAIL (BFC80+) study: a population-based prospective cohort study of the very elderly in Belgium. *BMC Geriatr* 2010;10:39.
- Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527.
- Fletcher CM. The clinical diagnosis of pulmonary emphysema; an experimental study. *Proc R Soc Med* 1952;45:577–84.
- Pahor M, Chrischilles EA, Guralnik JM, et al. Drug data coding and analysis in epidemiologic studies. *Eur J Epidemiol* 1994;10:405–11.
- Lang RM, Bierig M, Devereux RB, et al. Recommendations for chamber quantification. *Eur J Echocardiogr* 2006;7:79–108.
- Vaes B, Rezzouq N, Pasquet A, et al. The prevalence of cardiac dysfunction and the correlation with poor functioning among the very elderly. *Int J Cardiol* 2012;155:134–43.
- Quinones MA, Otto CM, Stoddard M, et al. Recommendations for quantification of Doppler echocardiography: a report from the Doppler Quantification Task Force of the Nomenclature and Standards Committee of the American Society of Echocardiography. *J Am Soc Echocardiogr* 2002;15:167–84.
- Zoghbi WA, Enriquez-Sarano M, Foster E, et al. Recommendations for evaluation of the severity of native valvular regurgitation with two-dimensional and Doppler echocardiography. *J Am Soc Echocardiogr* 2003;16:777–802.
- Nagueh SF, Appleton CP, Gillette TC, et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography. *Eur J Echocardiogr* 2009;10:165–93.
- Bleumink GS, Knetsch AM, Sturkenboom MC, et al. Quantifying the heart failure epidemic: prevalence, incidence rate, lifetime risk and prognosis of heart failure the Rotterdam Study. *Eur Heart J* 2004;25:1614–19.
- Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. *Circulation* 2013;128:e240–327.
- Owan TE, Hodge DO, Herges RM, et al. Trends in prevalence and outcome of heart failure with preserved ejection fraction. *N Engl J Med* 2006;355:251–9.
- Vaes B, Delgado V, Bax J, et al. Diagnostic accuracy of plasma NT-proBNP levels for excluding cardiac abnormalities in the very elderly. *BMC Geriatr* 2010;10:85.
- Gustafsson F, Steensgaard-Hansen F, Badskjaer J, et al. Diagnostic and prognostic performance of N-terminal ProBNP in primary care patients with suspected heart failure. *J Card Fail* 2005;11(5 Suppl): S15–20.
- Olofsson M, Boman K. Usefulness of natriuretic peptides in primary health care: an exploratory study in elderly patients. *Scand J Prim Health Care* 2010;28:29–35.
- Fazal IA, Bhagra SK, Bailey KM, et al. Impact of using different guideline recommended serum natriuretic peptide thresholds on the diagnosis and referral rates of a diagnostic heart failure clinic. *Int J Clin Pract* 2015;69:1349–56.