

The Diagnostic Value of Physical Examination and Additional Testing in Primary Care Patients With Suspected Heart Failure

Johannes C. Kelder, MD; Maarten J. Cramer, MD, PhD; Jan van Wijngaarden, MD, PhD; Rob van Tooren, MD; Arend Mosterd, MD, PhD; Karel G.M. Moons, PhD; Jan W. Lammers, MD, PhD; Martin R. Cowie, MD; Diederick E. Grobbee, MD, PhD; Arno W. Hoes, MD, PhD

Background—Early diagnosis of nonacute heart failure is crucial because prompt initiation of evidence-based treatment can prevent or slow down further progression. To diagnose new-onset heart failure in primary care is challenging.

Methods and Results—This is a cross-sectional diagnostic accuracy study with external validation. Seven hundred twenty-one consecutive patients suspected of new-onset heart failure underwent standardized diagnostic work-up including chest x-ray, spirometry, ECG, N-terminal pro-B-type natriuretic peptide (NT-proBNP) measurement, and echocardiography in specially equipped outpatient diagnostic heart failure clinics. The presence of heart failure was determined by an outcome panel using the initial clinical data and 6-month follow-up data, blinded to biomarker data. Of the 721 patients, 207 (28.7%) had heart failure. The combination of 3 items from history (age, coronary artery disease, and loop diuretic use) plus 6 from physical examination (pulse rate and regularity, displaced apex beat, rales, heart murmur, and increased jugular vein pressure) showed independent diagnostic value (c-statistic 0.83). NT-proBNP was the most powerful supplementary diagnostic test, increasing the c-statistic to 0.86 and resulting in net reclassification improvement of 69% ($P < 0.0001$). A simplified diagnostic rule was applied to 2 external validation datasets, resulting in c-statistics of 0.95 and 0.88, confirming the results.

Conclusions—In this study, we estimated the quantitative diagnostic contribution of elements of the history and physical examination in the diagnosis of heart failure in primary care outpatients, which may help to improve clinical decision making. The largest additional quantitative diagnostic contribution to those elements was provided by measurement of NT-proBNP. For daily practice, a diagnostic rule was derived that may be useful to quantify the probability of heart failure in patients with new symptoms suggestive of heart failure. (*Circulation*. 2011;124:2865-2873.)

Key Words: diagnosis ■ heart failure ■ natriuretic peptides ■ primary care ■ tests

Early diagnosis of heart failure is crucial because prompt initiation of treatment can prevent or slow down further progression.^{1,2} However, diagnosing heart failure is a challenge, and, when based on presenting clinical features alone, there may remain considerable diagnostic uncertainty.³ Yet, the relative importance of diagnostic information obtained at the bedside and subsequent testing, as well as the optimal diagnostic strategy, is not well determined. This particularly holds for settings where more advanced diagnostic tests, notably echocardiography, are not readily available.

Clinical Perspective on p 2873

The diagnostic value of patient history and physical examination has been questioned,^{4,5} albeit studies are few and often

have methodological limitations. Importantly, some studies focus on particular diagnostic tests in isolation of other tests and tend to ignore that the diagnostic process always involves multiple sources of information, tests, although not all these tests are explicit (eg, age, comorbidity). Moreover, the selection of patients for diagnostic research can lead to biased estimates of the test's performance when, for example, patients with established heart failure are simply contrasted with subjects without this disease.^{6,7} Much attention has been paid to the neurohormone B-type natriuretic peptide (BNP) and the inactive N-terminal counterpart NT-proBNP as diagnostic markers for heart failure, as exemplified by the 2008 European Society of Cardiology (ESC) Heart Failure guideline.⁸ However, although

Continuing medical education (CME) credit is available for this article. Go to <http://cme.ahajournals.org> to take the quiz.

Received January 11, 2011; accepted October 20, 2011.

From the Julius Center for Health Sciences and Primary Care (J.C.K., A.M., K.G.M.M., D.G., A.W.H.), Department of Cardiology (M.J.C., A.M.), and Department of Pulmonary Diseases (J.W.L.), University Medical Center, Utrecht, The Netherlands; Department of Cardiology, St. Antonius Hospital, Nieuwegein/Utrecht, The Netherlands (J.C.K., R.v.T.); Department of Cardiology (J.W.), Deventer Hospital, Deventer, The Netherlands; Department of Cardiology (A.M.), Meander Medical Center, Amersfoort, The Netherlands; and National Heart and Lung Institute, Imperial College London (Royal Brompton Hospital), United Kingdom (M.R.C.).

Correspondence to J.C. Kelder, MD, Julius Center for Health Sciences and Primary Care, Room 6.101, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands, PO Box 85500, 3508 GA Utrecht, The Netherlands. E-mail hans.kelder@gmail.com

© 2011 American Heart Association, Inc.

Circulation is available at <http://circ.ahajournals.org>

DOI: 10.1161/CIRCULATIONAHA.111.019216

BNP indeed shows excellent diagnostic accuracy, it has mainly been studied without regard to other test results, including signs and symptoms.^{9,10} Consequently, the relevance of BNP measurements in patients with suspected heart failure with other clinical information already available is unknown.

We set out to determine the diagnostic value of history, physical examination, and subsequent additional testing including BNP measurements to efficiently and accurately establish a diagnosis of new-onset heart failure in the domain of outpatients presenting with nonacute symptoms.

Methods

Setting and Participants

All patients presenting with symptoms and signs suggestive of heart failure (typically dyspnea, fatigue, signs of fluid retention) were eligible for inclusion in the study. Patients were referred to 1 of the 8 rapid access outpatient clinics by primary care physicians based in the catchment area of 7 participating hospitals throughout The Netherlands. The rapid access outpatient clinics provide a "1-stop shop" diagnostic service to the patient and the referring physician and enabled standardized collection of diagnostic data for the present analyses.

Patients with acute heart failure were excluded because they require immediate therapeutic intervention. Patients with known, established heart failure were also excluded.¹¹

Diagnostic Procedures

The standardized diagnostic work-up included: (1) patient history (including age, gender, dyspnea on exertion, orthopnea, fatigue, edema, medical history, medication, and smoking); (2) physical examination (including palpation of the apex beat, auscultation of the heart, pulmonary and abdominal examination, edema, and estimation of the jugular venous pressure, qualitatively as distension of the external jugular vein); (3) 12-lead ECG; (4) chest x-ray (cardiothoracic ratio and pulmonary edema as reported by the local specialist); (5) spirometry (forced expiration volume in 1 second, peak flow, and vital capacity); and (6) standard laboratory assessments (including hemoglobin, electrolytes, C-reactive protein, kidney and liver function parameters). In addition, plasma NT-proBNP was determined using validated commercially available immunoassay kits (Roche Diagnostics GmbH, Mannheim, Germany) and assessed at a central laboratory where plasma samples were stored at -80°C . Finally, echocardiography (M-mode, 2D, and Doppler flow) were performed locally in accordance with the American Society of Echocardiography guidelines.¹² Left ventricular ejection fraction was assessed semiquantitatively. Diastolic function was categorized as normal, impaired relaxation, or restrictive filling by a combination of left ventricular wall thickness, transmitral and pulmonary vein flow patterns, and left atrial volume, according to international guidelines,¹³ with the addition of mitral inflow pattern during Valsalva maneuver or tissue Doppler-derived E' when available, as well as assessing noncardiac explanation of the signs and symptoms in the individual patient, notably pulmonary dysfunction. All images were stored digitally for offline assessments.

Follow-Up

Data on the clinical course during the subsequent 6 months, including the response to therapy, were reported by the referring physician. This follow-up information was collected to provide definitive information on the presence or absence of heart failure for those cases with uncertainty at the time of the initial assessment and was used by the outcome panel during the consensus meetings (see below).

Diagnostic Outcome (Reference Standard)

The primary outcome of the study was the presence of heart failure at the time of initial presentation. Because a uniform reference test is

lacking for heart failure, we chose a consensus diagnosis as formal reference standard, in analogy with earlier studies and as recommended by recent diagnostic research guidelines.^{7,14,15} An outcome panel judged all the available diagnostic and 6-month follow-up information from each patient to determine the final diagnosis. The result of the NT-proBNP was not available during the consensus meetings, however, because a specific aim of our study was to quantify the added diagnostic value of this test and inclusion of the NT-proBNP result would lead to incorporation bias and overestimate its accuracy.¹⁶ Similarly, we did not assess the added diagnostic value of echocardiography, as it played a crucial role in setting the final diagnosis and has very limited availability in primary care, and because ECG, chest x-ray, and BNP are typically used before considering echocardiography.^{17,18}

The outcome panel comprised 1 of 4 cardiologists, 1 of 4 pulmonologists, and an outpatient heart failure clinic physician (J.C.K.). The latter member was present at all panel meetings, with a panel always consisting of 1 cardiologist and 1 pulmonologist. The approach to arrive at a final diagnosis was by discussion and then consensus. The panel assessed each patient for the presence of heart failure following the criteria and approach outlined by the ESC.¹¹ For each patient, confirmation of cardiac dysfunction, as well as signs and symptoms suggestive of heart failure, was assessed by the panel. In the event of doubt, the response to treatment during the 6-month follow-up period information was used, as outlined by the ESC guideline. If the panel decided heart failure was present, the potential cause and the presence of systolic and/or diastolic ventricular dysfunction were determined.

Statistical Analysis

The analysis was prespecified as a hierarchical multivariable logistic regression model. The aim was to estimate the contribution of sequential diagnostic information from history, physical examination, and additional testing and to summarize this in a diagnostic algorithm that would allow for optimal estimation of the absolute probability of heart failure being present using a minimal number of patient-burdening or expensive diagnostic tests. We first quantified the univariable association of all variables with the primary outcome. Subsequently, multivariable analysis was done, following the order in which tests are applied in practice and without preselection of variables based on univariable analysis, as this can lead to unstable models.¹⁹ The first model included all variables available from history taking and physical examination. Noncontributing tests were manually (1 by 1) excluded using the likelihood ratio test²⁰ and resulted in model 1. Second-stage models were conditional on model 1 to compute the added value of each of the supplementary tests, ie, ECG, chest x-ray, spirometry, plasma parameters (other than NT-proBNP), and NT-proBNP, resulting in models 2a to 2e. The final stage for our analysis was the assessment of the additional value of a second supplementary test conditional on model 2, resulting in models 3a to 3d. Tests were considered to provide added diagnostic value if the likelihood ratio test probability value comparing models was <0.05 .

The discriminative ability of each model was quantified by the c-statistic and the category-free net reclassification improvement,²¹ which measures reclassification in the right direction, for cases upward and for noncases downward the probability of heart failure scale. The calibration was tested with the Hosmer-Lemeshow statistic.²²

To adjust the coefficients and c-statistic for overoptimistic performance, a shrinkage factor was obtained from a bootstrap analysis (1 000 samples).²³ Finally, a diagnostic rule was derived from the shrunken, rounded, multivariable coefficients to estimate the probability of heart failure presence, ranging from 0% to 100%. Score thresholds for ruling in and ruling out heart failure were introduced based on clinically acceptable probabilities of false-positive (20% and 30%) and false-negative (10% and 20%) diagnoses. The average proportion of missing values was 4%, and we imputed missing values rather than performing a complete case analysis.²⁴ The imputation method used a multiple imputation technique as imple-

Table 1. Clinical Characteristics of the Study Population

	All n=721	Heart Failure		P Value
		Yes n=207 (28.7%)	No n=514	
Female gender	466 (64.6%)	125 (60.4%)	341 (66.3%)	0.14
Age (years)	70.7±11.8	75.5±9.7	68.8±12.1	<0.01
Medical history				
MI, CABG, or PCI	48 (6.7%)	26 (12.6%)	22 (4.3%)	<0.01
Hypertension	373 (51.7%)	115 (55.6%)	258 (50.2%)	0.22
Diabetes	110 (15.3%)	42 (20.3%)	68 (13.2%)	0.02
COPD	188 (26.1%)	52 (25.1%)	136 (26.5%)	0.78
History taking				
Dyspnea at <1 flight of stairs	438 (60.8%)	150 (72.5%)	288 (56.0%)	<0.01
Orthopnea	158 (21.9%)	61 (29.5%)	97 (18.9%)	<0.01
Paroxysmal nocturnal dyspnea	138 (19.1%)	55 (26.6%)	83 (16.2%)	<0.01
Nocturia (more than once)	280 (38.8%)	94 (45.4%)	186 (36.2%)	0.02
Smoker-never	275 (38.1%)	74 (35.8%)	201 (39.1%)	0.45
Medication use				
Loop diuretic	233 (32.3%)	114 (55.1%)	119 (23.2%)	<0.01
ACE-I or angiotensin II receptor blocker	165 (22.9%)	76 (36.7%)	89 (17.3%)	<0.01
Digoxin	40 (5.6%)	23 (11.1%)	17 (3.3%)	<0.01
NSAID	49 (6.8%)	17 (8.2%)	32 (6.2%)	0.33
Physical examination				
BMI (kg/m ²)	29.4±6.0	29.1±6.1	29.5±5.9	0.48
Blood pressure systolic (mm Hg)	156.0±26.7	153.8±30.0	157.0±25.3	0.14
Blood pressure diastolic (mm Hg)	87.0±13.2	86.5±15.0	87.1±12.4	0.60
Pulse rate (bpm)	77.2±14.8	82.5±16.8	75.1±13.3	<0.01
Wheezing or rhonchi	58 (8.0%)	15 (7.3%)	43 (8.4%)	0.76
Rales basal or more	99 (13.7%)	54 (26.1%)	45 (8.8%)	<0.01
Irregularly irregular pulse	72 (10.0%)	51 (24.6%)	21 (4.1%)	<0.01
Displaced apex beat	70 (9.7%)	51 (24.6%)	19 (2.6%)	<0.01
Third heart sound	10 (1.4%)	9 (4.4%)	1 (0.2%)	<0.01
Heart murmur suggesting mitral regurgitation	77 (10.7%)	46 (22.2%)	31 (6.0%)	<0.01
Elevated jugular venous pressure	56 (7.8%)	37 (17.9%)	19 (3.7%)	<0.01
Ankle swelling (bilateral)	197 (27.3%)	84 (40.6%)	113 (22.0%)	<0.01
Laboratory values				
Hemoglobin (mmol/L)	8.6±0.9	8.5±1.0	8.6±0.8	0.64
CRP >6 mg/(vs ≤6 mg/L)	333 (49.3%)	109 (56.2%)	224 (46.6%)	0.03
eGFR MDRD (mL/min/1.73 m ²)	65±16	62±17	67±16	<0.01
ALT >2 times upper limit of normal	26 (3.7%)	8 (3.9%)	18 (3.6%)	0.83
GGT >2 times upper limit of normal	38 (6.0%)	25 (13.2%)	13 (3.0%)	<0.01
Chest x-ray				
Cor-thorax ratio				
≤0.50	426 (66.8%)	74 (17.4%)	}	P for
0.5–0.55	104 (16.3%)	33 (18.3%)	}	trend
>0.60	108 (16.9%)	73 (67.6%)	}	<0.01
Pleural fluid right, left, or both	40 (5.6%)	28 (13.5%)	12 (2.3%)	<0.01
Pulmonary vascular redistribution	70 (9.7%)	52 (25.1%)	18 (3.5%)	<0.01
Kerley B lines	26 (3.6%)	19 (9.2%)	7 (1.4%)	<0.01

(Continued)

Table 1. Continued

		Heart Failure		
	All n=721	Yes n=207 (28.7%)	No n=514	P Value
ECG				
Rhythm				
Sinus rhythm	619 (87.9%)	141 (22.8%)	}	Exact
Atrial fibrillation	58 (8.2%)	43 (74.1%)	}	P
Other	27 (3.8%)	15 (55.6%)	}	<0.01
LBBB complete	29 (4.1%)	20 (10.0%)	9 (1.8%)	<0.01
LVH	79 (11.3%)	52 (26.4%)	27 (5.4%)	<0.01
Q waves inferior	39 (5.6%)	22 (11.1%)	17 (3.4%)	<0.01
Q waves anterior	27 (3.8%)	20 (10.1%)	7 (1.4%)	<0.01
Normal ECG	185 (26.5%)	15 (7.7%)	170 (33.9%)	<0.01
Spirometry				
Vital capacity (% predicted)	97.4±20.3	89.3±21.9	100.5±18.7	<0.01
FEV1 (% predicted)	88.4±24.7	80.9±25.5	91.3±23.7	<0.01
FEV1 (% of VC)	71.2±14.8	69.8±14.2	71.8±15.0	0.11
BNP				
Log NT-proBNP (log pg/ml)	3.40±1.70	4.85±1.86	2.84±1.28	<0.01
Echocardiogram				
LV systolic function “eyeball”				
Normal	532 (75.4%)	72 (34.8%)	460 (89.5%)	
Mild dysfunction	90 (12.7%)	50 (24.2%)	40 (7.8%)	
Moderate dysfunction	49 (6.9%)	46 (22.2%)	3 (0.6%)	
Severe dysfunction	35 (5.0%)	35 (16.9%)	0 (0%)	
LV diastolic function				
Normal	432 (59.9%)	56 (27.1%)	376 (73.2%)	
Impaired relaxation	161 (22.3%)	76 (36.7%)	85 (16.5%)	
Restrictive pattern	21 (2.9%)	20 (9.7%)	1 (0%)	
Missing data	107 (14.8%)	55 (26.6%)	52 (10.1%)	
LVH				
No	490 (68.0%)	121 (58.5%)	369 (71.8%)	
Mild	144 (20.0%)	49 (23.7%)	95 (18.5%)	
Moderate or severe	50 (6.9%)	26 (12.6%)	24 (4.7%)	
Missing data	37 (5.1%)	11 (5.3%)	26 (5.1%)	

Values are presented as n (%) or mean±SD.

ACE-I indicates angiotensin converting enzyme inhibitor; ALT, alanine amino transferase; BMI, body mass index; BNP, B-type natriuretic peptide; bpm, beats per minute; CABG, coronary artery bypass grafting; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein; ECG, electrocardiogram; eGFR MDRD, estimated glomerular filtration rate according to the Modification of Diet in Renal Disease Study Group equation; FEV1, forced expiration volume in 1 second; GGT, gamma-glutamyltransferase; LBBB, left bundle-branch block; LV, left ventricle; LVH, left ventricular hypertrophy; MI, myocardial infarction; NSAID, nonsteroidal anti-inflammatory drug; and PCI, percutaneous coronary intervention.

mented in SAS (version 9.1.3 [SAS Corp, Cary, NC] and R version 2.11 [http://www.r-project.org]).

External Validation

The final diagnostic rule was externally validated in the datasets from the UKNP²⁵ and Hillingdon²⁶ studies from the United Kingdom. The probability of heart failure was calculated for each patient in the respective datasets applying the diagnostic rule we derived and per dataset the c-statistic was calculated. External validation of the rule was hampered by the fact that not all variables were readily available in the 2 datasets. We only

included patients with nonmissing data on the NT-proBNP measurement. For the Hillingdon study validation, we used the subset of patients not assessed in the emergency room. The displaced apex beat was derived from the chest x-ray and defined as a cardiothoracic ratio being >60. The BNP was converted to NT-proBNP by means of the formula given by Alibay et al: $_{10}\log(\text{NT-proBNP}) = 1.1 * _{10}\log(\text{BNP}) + 0.57$; units: pg/mL.²⁷ The UKNP study lacked the "heart murmur suggestive of mitral regurgitation" variable, for which we used the echocardiographic mitral valve regurgitation to impute, with a penalty of scoring only 5 points instead of 10.

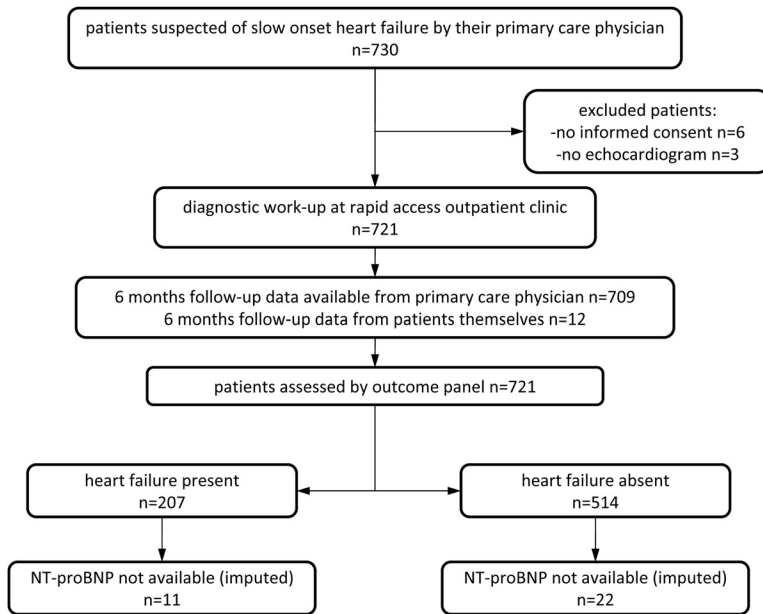


Figure 1. Flow chart.

Results

The characteristics of the 721 patients enrolled are shown in Table 1 and Figure 1; 6 patients refused to participate (Figure 2). Patients were on average 70.7 years of age (64.6% female). Of the patients, 51.7% had hypertension, 26.1% had chronic obstructive pulmonary disease, and 15.3% had diabetes mellitus. All patients had 1 or more complaints compatible with heart failure, notably dyspnea on exertion, ankle swelling, fatigue, or orthopnea. Of the patients, 7.8% had elevated jugular venous pressure, 9.7% a displaced apex beat, and 27.3% bilateral ankle swelling. In 10 patients, a third heart sound was heard.

A final diagnosis of heart failure was made in 207 patients (prevalence 28.7%; 95% CI 25.4–32.2), of whom 72 (34.8%) had a normal systolic function. Follow-up information on the clinical course from the general practitioner at 6 months was unavailable for 12 patients; these patients were contacted individually. Table 1 also shows the univariable associations of all potential diagnostic predictors.

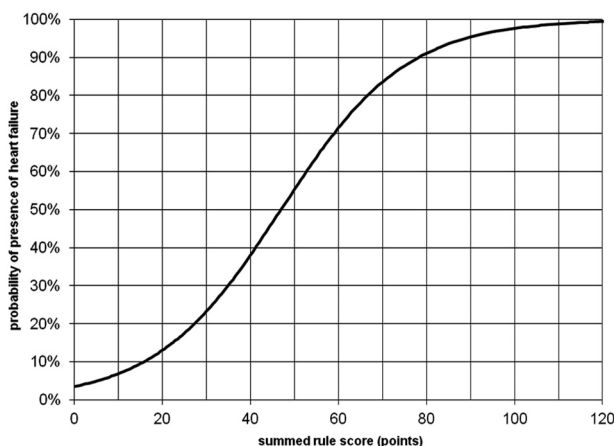


Figure 2. Relationship of summed diagnostic rule score with probability of presence of heart failure.

Age, history of coronary artery disease, use of a loop diuretic, and 6 items from the physical examination (displaced apex beat, irregular pulse, rales, pulse rate, heart murmur suggestive of mitral regurgitation, elevated jugular venous pressure) were independently associated with the presence of heart failure. Without any further testing, this yielded a c-statistic of 0.83 (model 1, Table 2).

The only standard laboratory test (ie, excluding NT-proBNP) with sufficient added diagnostic value was “GGT >2 times upper limit of normal” (model 2a). The addition of chest x-ray (cardiothoracic ratio and pulmonary vascular redistribution, model 2b) to model 1 also contributed to the diagnosis, as did the single ECG item “normal” ECG (model 2c) and spirometry (model 2d). The largest added value was provided by the log-transformed NT-proBNP (model 2e), reaching a c-statistic of 0.86, which translates to a net reclassification improvement of 69%, ie, the sum of 21% (net) of patients with heart failure having a higher probability of heart failure plus 48% (net) of patients without heart failure having a lower probability of heart failure. When applying a 20% threshold below which heart failure is considered excluded (for the time being) and 70% above which heart failure is considered present, then the net reclassification improvement for these three categories is 17%, ie, 13% (net) of patients with heart failure move up at least one category, whereas 4% (net) of patients without heart failure move down at least one category. In addition to the model including signs and symptoms and NT-proBNP (model 2e), the chest x-ray conferred the largest diagnostic improvement (model 3b). Of note, the addition of ECG to model 3b did not increase the diagnostic accuracy (data not shown).

All multivariable statistical models were found to have adequate calibration, with probability values from the Hosmer-Lemeshow test all >0.3 and chi-squares all <9.

We recalibrated model 2e with the computed shrinkage factor of 0.92 to account for overfitting and finally rounded the coefficients to formulate the diagnostic rule (Table 3, c-statistic 0.85). The summed score can be transformed into

Table 2. Multivariable Models Predicting the Presence of Heart Failure

	OR	(95% CI)	c-Statistic	NRI	P Value Likelihood Ratio Test
Model 1			0.83	Reference	Reference
Age (years)	x 1.52	(1.25–1.86)	x=per 10 y		
MI, CABG, or PCI	4.48	(2.25–8.92)			
Loop diuretic	2.16	(1.44–3.26)			
Displaced apex beat	5.37	(2.75–10.5)			
Irregularly irregular pulse	4.08	(2.19–7.62)			
Rales basal or more	2.12	(1.31–3.75)			
Pulse rate (bpm)	x 1.27	(1.11–1.45)	x=per 10 bpm		
Heart murmur suggestive of mitral regurgitation	2.45	(1.34–4.49)			
Elevated jugular venous pressure	2.78	(1.35–5.70)			
Model 2a (Model 1+standard laboratory values)			0.83	0.36	0.0019
GGT >2 times upper limit of normal	3.75	(1.60–8.80)			
Model 2b (Model 1+chest x-ray)			0.85	0.63	<0.0001
Cor-thorax ratio					
≤0.50	1.00				
0.50–0.55	1.45	(0.82–2.57)			
0.55–0.60	2.13	(1.07–4.26)			
>0.60	9.38	(2.85–30.9)			
Pulmonary vascular redistribution	4.31	(1.96–9.47)			
Model 2c (Model 1+ECG)			0.84	0.28	<0.0001
“Normal” ECG	0.32	(0.18–0.60)			
Model 2d (Model 1+spirometry)			0.83	−0.01	0.0133
Vital capacity (% predicted)	x 0.88	(0.79–0.97)	x=per 10%		
Model 2e (Model 1+NT-proBNP)			0.86	0.69	<0.0001
Log NT-proBNP (log pg/mL)	x 1.71	(1.47–1.98)	x=per factor 2		
Model 3a (Model 2e+lab values)			0.87	0.83	0.0009
GGT >2 times upper limit of normal	3.75	(1.60–8.80)			
Model 3b (Model 2e+chest x-ray)			0.88	0.78	<0.0001
Cor-thorax ratio					
≤0.50	1.00				
0.50–0.55	0.97	(0.51–1.87)			
0.55–0.60	1.20	(0.53–2.72)			
>0.60	4.59	(1.27–16.5)			
Pulmonary vascular redistribution	5.97	(2.36–15.1)			
Model 3c (Model 2e+ECG)			0.86	0.73	0.0360
“Normal” ECG	0.52	(0.27–0.98)			
Model 3d (Model 2e+spirometry)			0.86	0.64	0.0776
Vital capacity (% predicted)	x 0.90	(0.80–1.01)	x=per 10%		

NRI indicates net reclassification improvement; BNP, B-type natriuretic peptide; bpm, beats per minute; CABG, coronary artery bypass grafting; CI, confidence interval; ECG, electrocardiogram; GGT, gamma-glutamyltransferase; MI, myocardial infarction; OR, odds ratio; and PCI, percutaneous coronary intervention.

the absolute probability of the presence of heart failure by reading from Figure 2.

The use of the diagnostic rule in 105 patients, of whom 29 had heart failure in the Hillingdon study, resulted in a c-statistic of 0.950. The use of the diagnostic rule in 302 patients from the UKNP study, by means of external validation, resulted in a c-statistic of 0.884. Even when all these patients were set at “no heart murmur suggestive of mitral regurgitation,” the c-statistic was 0.882.

Discussion

The findings among a large unselected group of patients suspected of having new-onset nonacute heart failure in primary care shows that a history taking and physical examination will markedly reduce diagnostic uncertainty. Of the potential additional tests, NT-proBNP yielded the highest additional diagnostic value. The validation datasets confirmed these findings (Figure 3).

To appreciate these results, some aspects of the study need to be addressed. Previous reports on the diagnostic power of

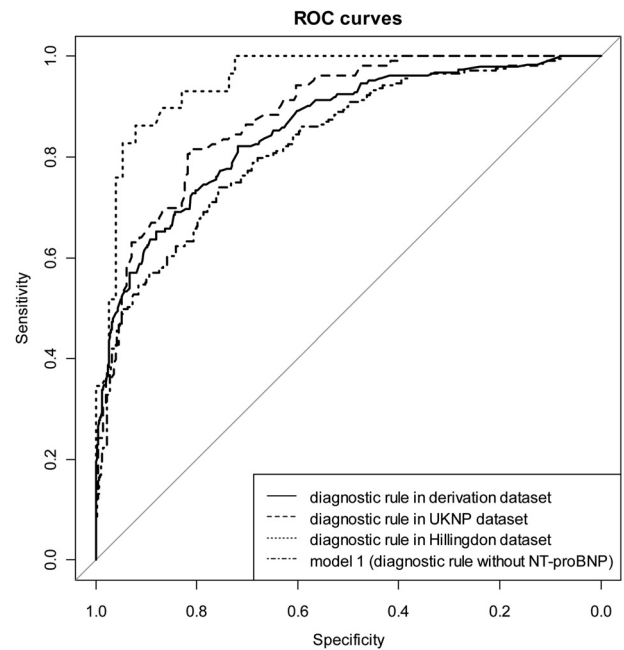
Table 3. Diagnostic Rule

Rule Score: Summation of Points	Points
Age (years)	
<60	0
60–70	4
70–80	7
>80	10
MI, CABG or PCI	
Present	15
Loop diuretic	
Present	10
Displaced apex beat	
Present	20
Rales basal or more	
Present	14
Irregularly irregular pulse	
Present	11
Heart murmur suggestive of mitral regurgitation	
Present	10
Pulse rate (bpm)	(bpm >60)/3
Elevated jugular venous pressure	
Present	12
NT-proBNP (pg/ml)	
<100	0
100–200	8
200–400	16
400–800	24
800–1600	32
1600–3200	40
>3200	48

c-statistic=0.85.

BNP indicates B-type natriuretic peptide; bpm, beats per minute; CABG, coronary artery bypass grafting; MI, myocardial infarction; and PCI, percutaneous coronary intervention.

the physical examination of suspected heart failure are scarce. Warnings have even been given for the serious limitations of physical examination,³ and very low univariable positive predictive values of solitary physical examination items have been published in relatively small studies.²⁸ In the present study, however, pulse rate, pulse regularity, pulmonary rales, displaced apex beat, heart murmur suggestive of mitral regurgitation, and distended jugular veins were all univariably associated with the presence of heart failure and retained that association in the multivariable analyses. We decided to exclude the third heart sound from our multivariable models mainly due to its low prevalence (10 patients). By itself, however, it holds a positive predictive value of 0.90, making it a specific marker of heart failure. When using only the history taking and physical examination findings in the validation datasets, the c-statistic reached 0.87 in both the Hillingdon and UKNP study (data not shown), albeit incorporation bias cannot be ruled out in all 3 studies. One could reasonably argue that it is important for diagnosticians to maintain their physical examination skills.

**Figure 3.** Receiver operating characteristic curves.

Owing to the large study size, we were able to assess the value of a range of tests potentially associated with the presence of heart failure in multivariable analyses, thereby following the natural hierarchy of daily clinical practice. For example, the chest x-ray contributed to the diagnostic accuracy when added to the clinical assessment, also in case NT-proBNP is known (model 3b), which concurs with a study by Cowie et al²⁶ that reported a (statistically nonsignificant) multivariable odds ratio of 2.65 for the cardiothoracic ratio. Badgett et al reviewed 29 studies and concluded that chest x-ray alone cannot adequately exclude or confirm left ventricular dysfunction in clinical settings.²⁹ According to several guidelines, a normal ECG rules out heart failure, based on the high negative predictive value (univariable) for echocardiographic systolic dysfunction of 98%, studied in a population with 18% systolic dysfunction and 51% normal ECG.³⁰ Others have also shown that a normal ECG may be useful to exclude heart failure.^{5,31} In the present study, the negative predictive value was 94% in a population with 24% systolic dysfunction and only 27% normal ECGs. In our study, an ECG added significant diagnostic accuracy to readily available diagnostic tests (model 2c) even when NT-proBNP was included in the model (3c). In contrast, one of the few diagnostic studies in this field that performed a multivariable analysis concluded that the ECG did not contribute to a diagnostic assessment including NT-proBNP.²⁵

Knowledge of the plasma level of NT-proBNP resulted in the largest gain in diagnostic accuracy. Three other studies of similar patients comparing BNP with other readily available diagnostic tests concluded that BNP is the most powerful diagnostic test.^{25,26,32} The vast majority of other studies assessing the diagnostic value of NT-proBNP were, however, performed in hospital emergency rooms and not in an outpatient or primary care setting. It should be emphasized that univariable diagnostic characteristics, such as sensitivity and specificity, cannot be routinely generalized from 1 setting to another.³³

Table 4a. Application of the Diagnostic Rule Resulting in Absence or Presence of Heart Failure

	Summed Score From Diagnostic Rule	Probability of Heart Failure Estimated by Rule	Total		Prevalence		False		
			n	%	n	%	n	%	95% CI
Assume heart failure absent	<13 points	<10%	233	32.3	11	4.7	11	4.7	2.4–8.3
	<24 points	<20%	403	55.9	38	9.4	38	9.4	6.8–12.3
Assume heart failure present	>54 points	>70%	96	13.3	84	87.5	12	12.5	7.2–22.4
	>63 points	>80%	69	9.6	63	91.3	6	8.7	3.3–18.0

4b. Application of the Diagnostic Rule Resulting in Diagnostic Uncertainty

Diagnostic uncertainty zone	13–54 points	10%–70%	392	54.4	112	28.6			
	24–54 points	20%–70%	222	30.8	85	38.3			
	13–63 points	10%–80%	419	58.1	133	31.7			
	24–63 points	20%–80%	249	34.5	106	42.6			

CI indicates confidence interval.

We selected the diagnostic rule including NT-proBNP as an additional test; it adds considerable diagnostic power, blood can be drawn at the patient's home, and the measurement can be performed in most laboratories. In addition, point-of-care tests for NT-proBNP are available, further facilitating the application of the diagnostic rule. Because there were only gradual differences with other diagnostic models (see Table 2, for example, one including ECG or chest x-ray instead of NT-proBNP), alternative strategies exist should NT-proBNP measurement not be available. We believe our study population is representative of, and our study results generalizable to, the large population of patients with suspected heart failure, ie, a population characterized by a mean age of approximately 70 years, predominantly white and female, often with considerable comorbidity and receiving polypharmaceutical therapy with a mixture of signs and symptoms compatible with heart failure. The prevalence of heart failure was 28.7%, which is similar to comparable previous studies.^{25,26,31,32} The use of a loop diuretic was a predictor of the presence of heart failure in our study, which seems somewhat surprising as its prescription is often related to heart failure, but is understandable from a clinical point of view. Patients presenting with symptoms of volume overload will often be treated with loop diuretics to alleviate symptoms, usually before tests to confirm heart failure are performed.

In the present study, we did not assess observer variability within a panel regarding the classification of heart failure as present or absent. In another study from our group, a comparable method was used, and a sample of 41 cases was reassessed resulting in disagreement in only 1 diagnosis.¹⁴ The outcome panel used all available information excluding the NT-proBNP values. Incorporation bias cannot be excluded for other variables.

Finally, we analyzed a substantial number of diagnostic tests, and although we adjusted for overfitting and optimism as much as possible, the accuracy of our rule should be further assessed by means of external validation studies. The external validation in 2 independent external datasets demonstrates the robustness of the diagnostic rule.

For daily clinical practice, the estimation of the probability of the presence of heart failure using the point score estimated in this analysis may help to make decision making more explicit. Consider thresholds whereby a disease probability of

<10% (13 points) is considered to rule out heart failure and a probability of >70% (54 points) as ruling in heart failure (Table 4). Application would have resulted in 96 (13.3%) patients assumed to have heart failure and 233 (32.3%) patients assumed not to have heart failure, leaving 392 (54.4%) patients in the intermediate no-decision group. The false-positive and false-negative rates are 12.5% (12 of 96) and 4.7% (11 of 233), respectively.

In conclusion, the findings obtained in this study conducted in a large unselected group of patients suspected of having new-onset heart failure in primary care support the view that history taking and, in particular, physical examination are major sources of diagnostic information. The estimation of their diagnostic contribution will help to improve clinical decision making. The largest additional diagnostic contribution is provided by measurement of NT-proBNP. For daily practice, these findings as well as the diagnostic rule derived from them, may be useful to quantify the probability of heart failure in patients with new symptoms suggestive of heart failure.

Sources of Funding

This study was funded by the Dutch Ministry of Health, ZON-MW grant 945-02-014. Roche Diagnostics supplied the kits for the NT-proBNP assessments. The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Disclosures

None.

References

1. Dargie HJ. Effect of carvedilol on outcome after myocardial infarction in patients with left-ventricular dysfunction: the CAPRICORN randomised trial. *Lancet*. 2001;357:1385–1390.
2. Jong P, Yusuf S, Rousseau MF, Ahn SA, Bangdiwala SI. Effect of enalapril on 12-year survival and life expectancy in patients with left ventricular systolic dysfunction: a follow-up study. *Lancet*. 2003;361:1843–1848.
3. Watson RD, Gibbs CR, Lip GYH. ABC of heart failure: clinical features and complications. *BMJ*. 2000;320:236–239.
4. Khunti K, Baker R, Grimshaw G. Diagnosis of patients with chronic heart failure in primary care: usefulness of history, examination, and investigations. *Br J Gen Pract*. 2000;50:50–54.
5. Nielsen OW, Hansen JF, Hilden J, Larsen CT, Svanegaard J. Risk assessment of left ventricular systolic dysfunction in primary care: cross sectional study evaluating a range of diagnostic tests. *BMJ*. 2000;320:220–224.

6. Moons KGM, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. *Clin Chem*. 2004;50:473–476.
7. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Ann Intern Med*. 2003;138:40–44.
8. Dickstein K, Cohen-Solal A, Filippatos G, McMurray JJV, Ponikowski P, Poole-Wilson PA, Stromberg A, van Veldhuisen DJ, Atar D, Hoes AW, Keren A, Mebazaa A, Nieminen M, Priori SG, Swedberg K. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2008: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2008 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association of the ESC (HFA) and endorsed by the European Society of Intensive Care Medicine (ESICM). *Eur Heart J*. 2008;29:2388–2442.
9. Doust JA, Glasziou PP, Pietrzak E, Dobson AJ. A systematic review of the diagnostic accuracy of natriuretic peptides for heart failure. *Arch Intern Med*. 2004;164:1978–1984.
10. Hill SA, Balion CM, Santaguida P, McQueen MJ, Ismail AS, Reichert SM, McKelvie R, Worster A, Raina PS. Evidence for the use of B-type natriuretic peptides for screening asymptomatic populations and for diagnosis in primary care. *Clin Biochem*. 2008;41:240–249.
11. Swedberg K, Cleland J, Dargie H, Drexler H, Follath F, Komajda M, Tavazzi L, Smiseth OA, Gavazzi A, Haverich A, Hoes A, Jaarsma T, Korewicki J, Levy S, Linde C, Lopez-Sendon JL, Nieminen MS, Pierard L, Remme WJ. Guidelines for the diagnosis and treatment of chronic heart failure: executive summary (update 2005): The Task Force for the Diagnosis and Treatment of Chronic Heart Failure of the European Society of Cardiology. *Eur Heart J*. 2005;26:1115–1140.
12. Schiller NB, Shah PM, Crawford M, DeMaria A, Devereux R, Feigenbaum H, Gutgesell H, Reichek N, Sahn D, Schnittger I. Recommendations for quantitation of the left ventricle by two-dimensional echocardiography. American Society of Echocardiography Committee on Standards, Subcommittee on Quantitation of Two-Dimensional Echocardiograms. *J Am Soc Echocardiogr*. 1989;2:358–367.
13. Oh JK, Appleton CP, Hatle LK, Nishimura RA, Seward JB, Tajik AJ. The noninvasive assessment of left ventricular diastolic function with two-dimensional and Doppler echocardiography. *J Am Soc Echocardiogr*. 1997;10:246–270.
14. Rutten FH, Moons KGM, Cramer MJ, Grobbee DE, Zuithoff NPA, Lammers JW, Hoes AW. Recognising heart failure in elderly patients with stable chronic obstructive pulmonary disease in primary care: cross sectional diagnostic study. *BMJ*. 2005;331:1379.
15. Cowie MR, Wood DA, Coats AJ, Thompson SG, Poole-Wilson PA, Suresh V, Sutton GC. Incidence and aetiology of heart failure; a population-based study. *Eur Heart J*. 1999;20:421–428.
16. Moons KGM, Grobbee DE. When should we remain blind and when should our eyes remain open in diagnostic studies? *J Clin Epidemiol*. 2002;55:633–636.
17. Rutten FH, Grobbee DE, Hoes AW. Differences between general practitioners and cardiologists in diagnosis and management of heart failure: a survey in every-day practice. *Eur J Heart Fail*. 2003;5:337–344.
18. Hobbs FDR, Jones MI, Allan TF, Wilson S, Tobias R. European survey of primary care physician perceptions on heart failure diagnosis and management (Euro-HF). *Eur Heart J*. 2000;21:1877–1887.
19. Moons KGM, Donders AR, Steyerberg EW, Harrell FE. Penalized maximum likelihood estimation to directly adjust diagnostic and prognostic prediction models for overoptimism: a clinical example. *J Clin Epidemiol*. 2004;57:1262–1270.
20. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007;115:928–935.
21. Pencina MJ, D'Agostino RB, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011;30:11–21.
22. 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;15:361–387.
23. Steyerberg EW. *Clinical Prediction Models*. New York: Springer, 2009.
24. Donders AR, van der Heijden GJ, Stijnen T, Moons KGM. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*. 2006;59:1087–1091.
25. Zaphiriou A, Robb S, Murray-Thomas T, Mendez G, Fox K, McDonagh T, Hardman SMC, Dargie HJ, Cowie MR. The diagnostic accuracy of plasma BNP and NT-proBNP in patients referred from primary care with suspected heart failure: results of the UK natriuretic peptide study. *Eur J Heart Fail*. 2005;7:537–541.
26. Cowie MR, Struthers AD, Wood DA, Coats AJ, Thompson SG, Poole-Wilson PA, Sutton GC. Value of natriuretic peptides in assessment of patients with possible new heart failure in primary care. *Lancet*. 1997;350:1349–1353.
27. Alibay Y, Beauchet A, El Mahmoud R, Schmitt C, Brun-Ney D, Benoit M-O, Dubourg O, Boileau C, Jondeau G, Puy H. Plasma N-terminal pro-brain natriuretic peptide and brain natriuretic peptide in assessment of acute dyspnea. *Biomed Pharmacother*. 2005;59:20–24.
28. Harlan WR, Oberman A, Grimm R, Rosati RA. Chronic congestive heart failure in coronary artery disease: clinical criteria. *Ann Intern Med*. 1977;86:133–138.
29. Badgett RG, Mulrow CD, Otto PM, Ramirez G. How well can the chest radiograph diagnose left ventricular dysfunction? *J Gen Intern Med*. 1996;11:625–634.
30. Davie AP, Francis CM, Love MP, Caruana L, Starkey IR, Shaw TRD, Sutherland GR, McMurray JJV. Value of the electrocardiogram in identifying heart failure due to left ventricular systolic dysfunction. *BMJ*. 1996;312:222.
31. Fox KF, Cowie MR, Wood DA, Coats AJ, Poole-Wilson PA, Sutton GC. A rapid access heart failure clinic provides a prompt diagnosis and appropriate management of new heart failure presenting in the community. *Eur J Heart Fail*. 2000;2:423–429.
32. Landray MJ, Lehman R, Arnold I. Measuring brain natriuretic peptide in suspected left ventricular systolic dysfunction in general practice: cross-sectional study. *BMJ*. 2000;320:985–986.
33. Knottnerus JA, Leffers P. The influence of referral patterns on the characteristics of diagnostic tests. *J Clin Epidemiol*. 1992;45:1143–1154.

CLINICAL PERSPECTIVE

The initial diagnosis of new nonacute heart failure is as difficult as it is important, especially in settings where more advanced diagnostic tests, notably echocardiography, are not readily available. This diagnostic study demonstrates the importance of history taking and, in particular, physical examination in a large cohort of 721 patients suspected of having nonacute heart failure, 207 (28.7%) of whom were diagnosed with heart failure. When a primary care physician is confronted with a patient suspected of having nonacute heart failure, a diagnostic rule including age, objective evidence of prior coronary artery disease, use of a loop diuretic, displaced apex beat, basal or more pulmonary rales, irregularly irregular pulse, heart murmur suggestive of mitral regurgitation, pulse rate, and elevated jugular venous pressure, plus a single additional test, the NT-proBNP plasma level, can accurately predict the presence or absence of heart failure numerically. The rule had comparable performance in 2 validation datasets. Chest x-ray or ECG can be used in addition to or instead of the NT-proBNP measurement depending on local availability and other patient characteristics. It remains important for diagnosticians to maintain basic physical examination skill.