ORIGINAL RESEARCH ARTICLE

A systematic review of clinical prediction rules for the diagnosis of chronic heart failure

Joe Gallagher[1,2]*, Darren McCormack[4], Shuaiwei Zhou[3], Fiona Ryan[3], Chris Watson[5], Kenneth McDonald[1] and Mark T. Ledwidge[1]

[1]*School of Medicine & Medical Sciences, University College Dublin, Belfield, Dublin 4, Ireland;* [2]*Irish College of General Practitioners, Lincoln Place, Dublin, Ireland;* [3]*Heartbeat Trust, Crofton Terrace, Dun Laoghaire, Co Dublin, Ireland;* [4]*gHealth Research Group, University College Dublin, Belfield, Dublin 4, Ireland;* [5]*Centre for Experimental Medicine, Queens University, Belfast, Ireland*

Abstract

Aims This study sought to review the literature for clinical prediction models for the diagnosis of patients with chronic heart failure in the community and to validate the models in a novel cohort of patients with a suspected diagnosis of chronic heart failure.

Methods and results MEDLINE and Embase were searched from 1946 to Q4 2017. Studies were eligible if they contained at least one multivariable model for the diagnosis of chronic heart failure applicable to the primary care setting. The CHARMS checklist was used to evaluate models. We also validated models, where possible, in a novel cohort of patients with a suspected diagnosis of heart failure referred to a rapid access diagnostic clinic.

In total, 5310 articles were identified with nine articles subsequently meeting the eligibility criteria. Three models had undergone internal validation, and four had undergone external validation. No clinical impact studies have been completed to date. Area under the curve (AUC) varied from 0.74 to 0.93 and from 0.60 to 0.65 in the novel cohort for clinical models alone with AUC up to 0.89 in combination with electrocardiogram and B-type natriuretic peptide (BNP). The AUC for BNP was 0.86 (95% confidence interval 83.3–88.6%).

Conclusions This review demonstrates that there are a number of clinical prediction rules relevant to the diagnosis of chronic heart failure in the literature. Clinical impact studies are required to compare the use of clinical prediction rules and biomarker strategies in this setting.

Introduction

The diagnosis of chronic heart failure poses major challenges. Studies have shown that clinical assessment alone is grossly insufficient in the diagnosis of heart failure with only half of patients diagnosed having echocardiography evidence of heart failure.[1] However, access to echocardiography is often limited.[2] In the early 2000s, only a third of those with a clinical diagnosis of heart failure had echocardiography or further referral,[3] and this has not changed significantly in more recent times.[4]

Clinical prediction rules are defined as 'a clinical tool that quantifies the individual contributions that various components of the history, physical examination and basic laboratory results make towards the diagnosis, prognosis, or likely response to treatment in an individual patient'.[5] Such rules have been proposed in a number of areas as a solution to help guide physicians as to which patients may need referral. An adapted clinical prediction rule has been integrated into the National Institute for Health and Care Excellence guidelines in the UK on the diagnosis of chronic heart failure.[6]

The aim of this study was to undertake a systematic review of the literature to identify current clinical prediction rules used for the diagnosis of chronic heart failure in the community and to undertake a validation of these rules,

where possible, in a novel cohort of patients referred to a rapid access heart failure diagnostic clinic by their general practitioner.

Methods

The systematic review was undertaken according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)[7] guidelines and the CHARMS checklist for systematic reviews of prediction models.[8]

Search strategy

The review question and design were framed using the CHARMS checklist for systematic reviews of prediction models[8] (*Appendix S1A*). A systematic search strategy was then constructed for use in MEDLINE and Embase (details on the search syntax can be found in *Appendix S1B*). In brief, a search string was developed to identify clinical prediction rules for the diagnosis of heart failure using terms including 'heart failure', 'cardiomyopathy' 'diagnosis', 'model', 'predict', 'decision', or 'score'. We searched for studies on the development, validation, and clinical impact of clinical prediction models for heart failure published between 1946 and Q4 2017. There was no language restriction.

Eligibility criteria

Studies were eligible if they reported multivariable models for the diagnosis of heart failure in patients in the community.

Inclusion criteria for the systematic review were

 (i) patient population: adult human patients,
 (ii) multivariable models relating to the diagnosis of heart failure,
 (iii) outcome measure: heart failure, and
 (iv) setting of care: hospital outpatient or primary care (studies of clinical prediction rules conducted in non-primary care settings were eligible for inclusion if they were relevant to primary care).

Exclusion criteria were

 (i) studies that developed clinical prediction rules for patients with acute decompensated heart failure only,
 (ii) models that were designed to identify patients with acute heart failure, and
 (iii) studies based on single predictors only (as they are prone to reporting overly optimistic findings owing to a number of methodological limitations).

We used broad inclusion criteria to acknowledge variation in the same-day diagnostic tests that are available across countries, with the understanding that not all countries will have access to the same services and that the role of primary care clinicians also differs internationally. Given the scope of the systematic review, we assessed all diagnostic prediction models retrieved by our search for their applicability to primary care. In particular, it must have been possible for diagnostic predictors or variables to be assessed in the community or at a general practitioner's surgery. Variables such as symptoms and signs, history taking, electrocardiogram (ECG), and blood tests are easily obtained, whereas items such as echocardiography or other advanced imaging are not.

Critical appraisal

We critically appraised studies using the CHARMS checklist (*Appendix S1D–F*).

Two reviewers (J. G. and D. M.) screened all titles and abstracts and made decisions regarding potential eligibility after full-text review. In case of doubt, a third reviewer was consulted.

The reference lists of eligible studies as well as reviews relating to this subject were hand searched for identification of additional relevant publications. We also requested recognized experts to suggest other articles or sources.

For each included study, the following information was extracted on the basis of the CHARMS checklist by J. G.: source of data, participant characteristics, outcomes to be predicted, candidate predictors, sample size, handling of missing data, model development, model performance (discrimination and calibration), and model evaluation (internal or external validation) and results.

The methodological quality of each article was assessed using criteria based on a previous systematic review of clinical prediction rules in primary care based on modified McGinn criteria[5] in which the criterion concerning '100% follow-up' was changed to 'adequate follow-up' and was defined as ≥80% follow-up of study participants.[5,9] There were eight criteria assessing internal and external validity for derivation studies, and for validation studies, there were five criteria (*Appendix S1C*).

External validation

Rules were validated in a dataset from the new heart failure diagnostic clinic in the St. Vincent's Healthcare Group, Dublin. Only those rules where the variables in the rule were available in the validation dataset were included. General practitioners referred suspected new-onset cases to the rapid access clinic, which is part of the heart failure unit. Patients

Table 1 Summary of clinical prediction rule

	Fahey et al.[10]	Rutten et al.[11]	Yamamoto et al.[12]	Nielsen et al.[13]	Oudejan et al.[15]	Kelder et al.[14]	Roalfe et al.[16]	Boonman-de Winter et al.[17]	Van Riet et al.[18]
Year	2007	2005	2000	2000	2011	2011	2012	2015	2016
Stage of development	Derivation and external validation	Derivation and internal validation	Derivation	Derivation	Derivation and internal validation	Derivation and external validation	Derivation and external validation	Derivation and external validation	Validate, update, and extend a rule for patients with COPD
Clinical setting	26 general practices	51 GP	Single echocardiography clinic	Three GPs	Geriatric outpatient clinic	HF outpatient clinic	HF outpatient clinic	GP	30 general practices
Design	Prospective	Cross-sectional	Prospective	Prospective	Prospective	Prospective	Prospective	Prospective	Prospective
Outcome	HF-REF	Heart failure	HF-REF	HF-REF	Heart failure	Heart failure	Heart failure	Heart failure	Heart failure
Patient age	10% age > 65 years	73 years	65 years	71	82 years	70.7 years	71.5 years	71.6 years	74.1 years
Male	40%	55.10%	55%	55%	62%	35.40%	41%	53%	45.5%
Derivation cohort	458	405	466	126	206	721	306	581	
Validation cohort	536	—	—	—	—	407	936		585
Model discrimination	0.88	0.76	0.83 for HF-REF, 0.78 for ALVSD. 0.90 for EF < 35%	—	0.90 (0.86–0.94)		0.88 in derivation, 0.84–0.93 in validation	0.82 clinical, 0.86 with clinical, ECG, and NT-proBNP, 0.74 in external validation cohort	0.84 (range 0.80–0.85) of original rule 0.88 (0.85–0.90) of updated rule
Calibration	HL = 234.6	—	Not assessed	—	HL and calibration plots	—	Calibration plots	HL > 0.3 and calibration plots	HL 0.02 poor calibration in original rule HL 0.56 good calibration in updated rule
Comment		COPD patients only			Geriatric OPD referrals			Diabetes patients only	

ALVSD, asymptomatic left ventricular systolic dysfunction; COPD, chronic obstructive pulmonary disorder; GP, general practitioner; HL, Hosmer-Lemeshow test; OPD, outpatient department.

whose presentation indicated an acute onset were advised to go to the emergency department and were not included in this study. On attending the clinic, all patients underwent a systematic assessment including medical history and physical examination by a cardiology registrar and heart failure nurse specialist. Routine phlebotomy (haematology, renal profile, and natriuretic peptide) was also performed. Doppler echocardiography was performed in those with a B-type natriuretic peptide (BNP) > 100 pg/mL; in those with a BNP between 50 and 100 pg/mL already treated with diuretic and/or with a body mass index (BMI) > 30; and in those with a BNP < 50 pg/mL with further cardiovascular reason for echocardiographic assessment. The patient was then reviewed by a consultant cardiologist, and a decision was made to confirm or rule out the diagnosis of heart failure or, in equivocal situations, to send the patient for additional tests. Patients requiring additional tests were scheduled for review to confirm or rule out the diagnosis.

Heart failure was diagnosed on the basis of typical symptoms, with signs of fluid overload and/or an elevated BNP (>100 or >50 pg/mL with new diuretic use and/or elevated BMI), with Doppler echocardiographic evidence of systolic and/or diastolic dysfunction of the left ventricle or significant valvular disease.

Data analysis

It was possible to validate four rules in our novel cohort on the basis of the variables available in the cohort and those in the rules selected. For the four rules to be validated in our novel cohort, we retrospectively calculated the individual score or threshold of each included patient on the basis of models' parameters. The area under the curve of the receiver operating characteristics curve, with 95% confidence intervals (CIs) were used to compare the overall discriminative ability of rules of the models. We also calculated the common diagnostic accuracy measures for the rules of the models, i.e. sensitivity, specificity, positive predictive values, and negative predictive values as well as their corresponding 95% CIs.

All calculations including statistical analysis were carried out using R language version 3.2.3.

Figure 1 PRISMA flow diagram.
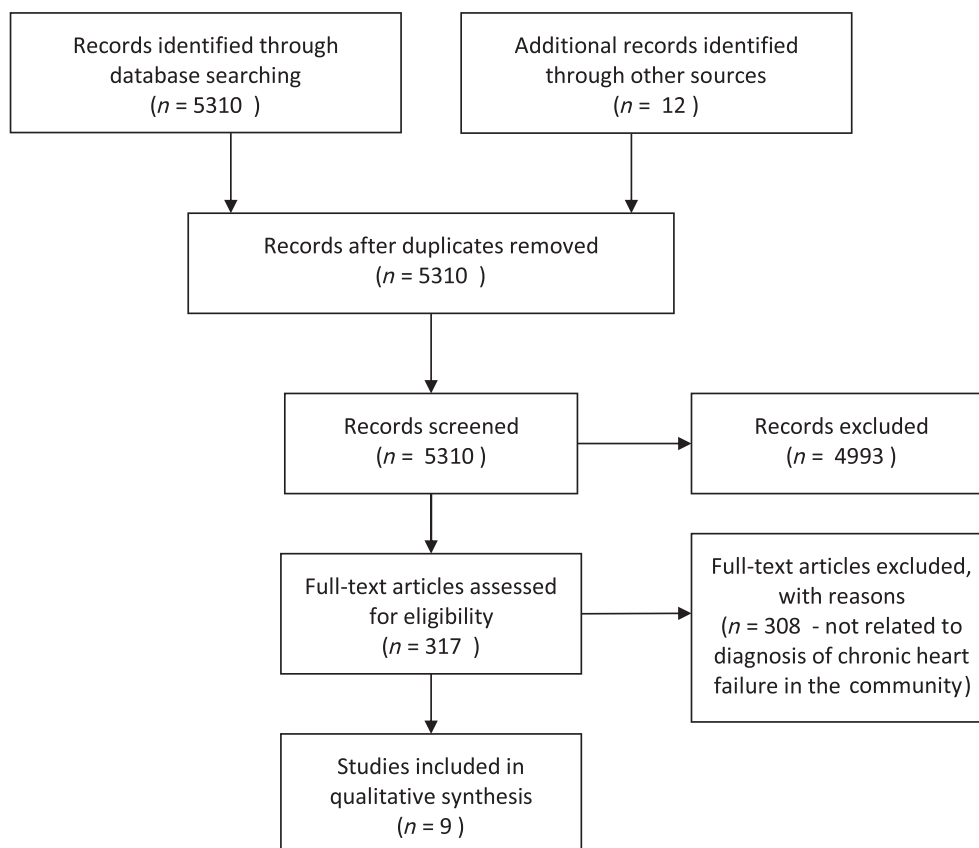
Results

A review of 5310 abstracts led to the identification of eight main models for the diagnosis of heart failure in the community and one updated model (*Table 1*).[10–18] The search results are outlined in the PRISMA flow diagram (*Figure 1*). One modification of a rule was identified.[18] One trial protocol for a clinical impact study of a clinical prediction rule was found.[19]

Clinical settings for model development

Candidate variables varied widely from five to 36. Final predictors in the model varied from five to 11. Four rules used symptoms, eight rules used physical examination findings, five rules used ECG findings, and five rules used biomarker data (all relating to natriuretic peptides). The year of publication of the rules ranged from 2000 to 2016 and became more common in later years.

Table 2 Candidate and final predictors

	Fahey et al.	Rutten et al.	Yamamoto et al.	Nielsen et al.	Kelder et al.	Roalfe et al.	Boonman-de Winter et al.	Oudejan et al.	Van Riet et al.
Age		✓			✓✓	✓	✓	✓✓	✓✓
Gender	✓✓				✓	✓✓	✓	✓✓	
MI, CABG, or PCI	✓✓		✓✓	✓	✓✓	✓✓	✓		
Dyspnoea, orthopnoea, and PND	✓✓		✓✓		✓	✓		✓✓	
Smoking status		✓			✓	✓	✓		✓
IHD		✓✓					✓		✓✓
Medication	✓	✓			✓✓	✓			✓
Abnormal ECG	✓✓	✓✓	✓✓	✓✓	✓	✓	✓✓	✓	✓✓
BNP/NT-proBNP level		✓✓		✓	✓✓	✓✓	✓✓	✓✓	✓✓
Angina	✓					✓	✓	✓	
CCF			✓✓	✓					
BMI/obesity	✓	✓✓			✓	✓		✓✓	✓✓
Elevated JVP	✓✓	✓			✓✓		✓	✓	✓
Displaced apex beat	✓	✓✓			✓✓				✓✓
Abnormal chest X-ray			✓✓		✓			✓	
Peripheral oedema		✓			✓	✓✓	✓		✓✓
Asthma or COPD					✓		✓		
NT-ANP				✓✓					
Heart rate		✓✓		✓	✓✓				✓✓
Blood pressure	✓				✓	✓	✓		
Wheeze	✓				✓			✓✓	
heart murmur	✓				✓✓				✓✓
Hyperlipidaemia	✓					✓			
Crepitations						✓✓	✓		
TIA/stroke						✓	✓		
Basal rales					✓✓			✓	
Irregularly irregular pulse					✓✓			✓	
Diabetes					✓	✓			
Coronary revascularization	✓								
Previous echocardiogram	✓								
Cardiovascular co-morbidity		✓							✓
Nocturia		✓			✓			✓	✓
C-reactive protein		✓			✓				✓
Haemoglobin					✓				
eGFR					✓				
ALT > 2 ULN					✓				
GGT > 2 ULN					✓				
Abnormal spirometry					✓				
Alcohol consumption						✓			
Fatigue						✓	✓		
PTCA						✓	✓		
Peripheral vascular disease						✓			
Atrial fibrillation							✓	✓	
Loss of appetite								✓✓	
Digitalis								✓	
Oral anticoagulants								✓	

✓, candidate variables; ✓✓, final variables; ALT, alanine transaminase; ANP, atrial natriuretic peptide; BNP, B-type naturitic peptide; CABG, coronary artery bypass graft; CCF, Chronic cardiac failure; COPD, chronic obstructive pulmonary disorder; ECG, electrocardiogram; eGFR, estimated glomerular filtration rate; GGT, gamma-glutamyl transpeptidase; IHD, ischaemic heart disease; JVP, jugular venous pressure; MI, myocardial infarction; PCI, percutaneous coronary intervention; PND, paroxmal nocturnal dyspnoea; PTCA, percutaneous transluminal coronary angioplasty; TIA, transient ischaemic attack; ULN, upper limit of normal.

Heart failure subtypes

Six rules addressed heart failure in general [heart failure with reduced ejection fraction (HF-REF) and heart failure with preserved ejection fraction (HF-PEF)], and three rules were for the prediction of HF-REF only. No rule predicted HF-PEF only. Three rules were developed for specific patient populations (diabetes, geriatric, and chronic obstructive pulmonary disorder).

Populations

Mean age was reported in eight rules; it varied from 65 to 82 years. One rule reported that only 10% of the patients were >65 years. The proportion of male patients ranged from 35.4% to 55.1%. Four rules were derived from the general practice setting, two from referrals to a heart failure outpatient clinic, and one from an echocardiography unit. Six rules were derived from prospectively collected data.

Year of publication

All the studies included in our literature review were published between 2000 and 2016.

Stage of development

All clinical prediction articles involved derivation; two rules were solely derived, two had derivation and internal validation, and four had derivation and external validation. There were no clinical impact studies retrieved. However, one clinical prediction rule had a protocol for a clinical impact study published.[19]

Clinical setting

The most common setting was general practice (five studies), followed by heart failure outpatient clinics (two studies), an echocardiography clinic (one study), and a geriatric outpatient clinic (one study).

Variables/predictors

There was a wide range of both candidate and final variables used in the studies, which are represented in *Table 2*.

Bias

Factors affecting bias are outlined in *Appendix S1C* and in the CHARMS data table (*Appendix S1D–F*). A number of studies involved an unfavourable number of events per candidate variable, which can lead to overestimation of effect. In a number of studies, it was also not possible to blind assessors to the outcome, as this involved a consensus panel to reach the diagnosis of heart failure. The use of a consensus panel to establish the diagnosis of heart failure may have resulted in incorporation bias, by knowing the results of the diagnostic

Table 3 Demographics

	Not heart failure	Heart failure		
		All	Reduced EF	Preserved EF
n	448	285	116	169
Age[a]	75.5 [67.9:81.5]	79.4 [73.0:84.3]	77.7 [70.4:84.2]	80.1 [73.5:84.3]
Male[a,b]	41.1%	50.5%	61.2%	43.2%
Atrial fibrillation[a,b]	11.4%	53.7%	44.8%	59.8%
Cancer	7.8%	12.3%	14.7%	10.7%
Cardiovascular disease[a,b]	16.7%	34%	47.4%	24.9%
Chronic renal failure[a]	1.6%	5.6%	8.6%	3.6%
COPD	14.1%	16.1%	19.8%	13.6%
Diabetes	11.6%	16.5%	12.9%	18.9%
Hypertension[b]	58.5%	61.8%	49.1%	70.4%
Stroke/TIA	7.6%	8.1%	7.8%	8.3%
Valvular disease	1.8%	3.5%	4.3%	3%
Baseline BMI (kg/m^2)	29.6 [26.6:34.0]	29.3 [25.5:34.0]	28.1 [25.2:31.8]	30.0 [26.1:34.6]
Baseline SBP[b] (mmHg)	141 [129:158]	137 [125:157]	130 [120:151]	143 [129:160]
Baseline DBP (mmHg)	79 [71:87]	78 [68:89]	77 [68:89]	78 [69:90]
Baseline creatinine[a] (mg/dL)	1.10 [0.92:1.32]	1.25 [1.05:1.60]	1.30 [1.09:1.73]	1.19 [1.04:1.47]
Primary aetiology				
Hypertensive		43.5%	17.2%	61.5%
Ischaemic		35.8%	59.5%	19.5%
Valvular		9.5%	6.0%	11.8%
Alcohol/drugs		1.8%	4.3%	0%
Idiopathic		4.9%	10.3%	1.2%
Unknown		4.6%	2.6%	5.9%

Bonferroni correction with *n* = 2 and *α* = 0.05. Cardiovascular disease includes ischaemic heart disease, peripheral arterial disease and abdominal aortic aneurysm.
COPD, chronic obstructive pulmonary disorder; DBP, diastolic blood pressure; SBP, systolic blood pressure; TIA, transient ischaemic attack.
[a]Statistically significant for not heart failure vs. heart failure.
[b]Statistically significant for reduced EF vs. preserved EF.

tests under study. However, 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.

External validation

It was possible to validate four of the rules in our cohort. The demographics of the validation cohort are presented in *Table 3*.

Results of external validation show that clinical prediction rules perform similarly in the novel cohort as in their original cohorts (*Table 4*). However, their performance is similar to that of BNP alone in this cohort (*Figure 2*).

Discussion

This study sought to review the literature for clinical prediction rules for the diagnosis of patients with chronic heart failure in the community and validate them, where possible, in a novel cohort.

Eight previously developed rules for the diagnosis of chronic heart failure in the community and one adaptation of a rule were identified. Four were externally validated, and one is undergoing a clinical impact study. In those studies that estimated the effect of natriuretic peptide on the performance of the clinical prediction rule, and in our own validation study, this biomarker performed as well as, and indeed better than in some cases, the proposed clinical prediction rules. It may be that clinical prediction rules based on symptoms and signs may not add to the diagnostic process and a natriuretic peptide-based approach in those patients in whom the general practitioner considers heart failure as a diagnosis may be as useful. Although there are concerns about overestimation of single-marker models, a biomarker-based approach should be investigated in comparison with the use of clinical prediction rules based on symptoms and clinical signs only for the diagnosis of heart failure in the community in a clinical impact study. Also the use of multimarker panels should be researched further. A paper from the REFER study supports this as natriuretic peptide testing alone performed as well as the validated clinical prediction rule in determining which patients presenting with symptoms went on to have a diagnosis of heart failure. It was also noted that the current N-terminal proBNP (NT-proBNP) cut-off level of 400 pg/mL used in the UK is too high and means that one in five patients with heart failure may not be appropriately referred for further investigation and diagnosis.[20]

This systematic review of clinical prediction rules in the diagnosis of chronic heart failure in the community used an extensive literature search and structured critical appraisal of the models using the guidance of the CHARMS checklist of the Cochrane collaboration. The validation cohort in this study is large compared with that used for the derivation of these

Table 4 Performance of clinical prediction rules in the novel cohort

Author	Subsample/threshold	AUC	Sensitivity	Specificity	NPV	PPV
BNP alone	BNP	85.9% [83.3–88.6%]	80.1% [75.6–84.7%]	80.2% [76.5–84.0%]	85.4% [82.0–88.8%]	73.7% [68.9–78.5%]
Roalfe[16]	MICE	64.9% [61.0–68.8%]	61.8% [56.3–67.2%]	59.7% [55.2–64.2%]	69.8% [65.3–74.4%]	50.8% [45.7–55.9%]
Roalfe	MICE and ECG	70.4% [66.5–74.3%]	64.3% [58.570.1%]	63.8% [59.1–68.5%]	73.4% [68.8–78.0%]	53.5% [48.059.0%]
Roalfe	MICE and BNP	86.6% [84.1–89.2%]	79.8% [75.2–84.4%]	79.8% [76.0–83.6%]	85.1% [81.6–88.6%]	73.1% [68.3–78.0%]
Roalfe	MICE and ECG and BNP	88.6% [86.1–91.1%]	81.5% [76.7–86.2%]	81.3% [77.4–85.2%]	86.7% [83.2–90.2%]	74.6% [69.5–79.6%]
Fahey[10]	MIJO	65.1% [61.2–68.9%]	66.0% [60.7–71.3%]	54.8% [50.3–59.4%]	70.5% [65.8–75.3%]	49.6% [44.8–54.5%]
Fahey	MIJO and ECG	69.4% [65.5–73.4%]	61.2% [55.3–67.1%]	65.3% [60.6–69.9%]	72.2% [67.6–76.8%]	53.3% [47.7–58.9%]
Fahey	MIJO and BNP	86.9% [84.3–89.4%]	79.5% [74.9–84.1%]	79.5% [75.7–83.3%]	84.9% [81.4–88.4%]	72.8% [68.0–77.7%]
Fahey	MIJO and ECG and BNP	88.7% [86.2–91.2%]	81.5% [76.7–86.2%]	81.6% [77.7–85.4%]	86.7% [83.2–90.2%]	74.8% [69.8–79.9%]
Boonman-de Winter[17]	Clinical score	64.1% [60.8–67.4%]	71.6% [67.7–75.6%]	49.8% [45.7–53.9%]	66.7% [62.2–71.2%]	55.6% [51.8–59.6%]
Boonman-de Winter	Clinical score and BNP					
Yamamoto[12]	Clinical score	60.7% [56.5–64.9%]	70.8% [65.5–76.1%]	45.7% [40.9–50.5%]	69.4% [63.9–74.9%]	47.4% [42.6–52.2%]
Yamamoto	Any 1–4 and BNP	65.0% [62.3–67.7%]	58.2% [54.2–62.2%]	71.8% [68.0–75.5%]	62.3% [58.6–66.1%]	68.1% [64.0–72.2%]

AUC, area under the curve; MICE, male, infarction, crepitations, oedema; NPV, negative predictive value; PPV, positive predictive value.

Figure 2 Calibration plots for BNP alone (A), MICE[16] rule with BNP (B), and MIJO[10] rule with BNP (C). BNP, B-type natriuretic peptide.
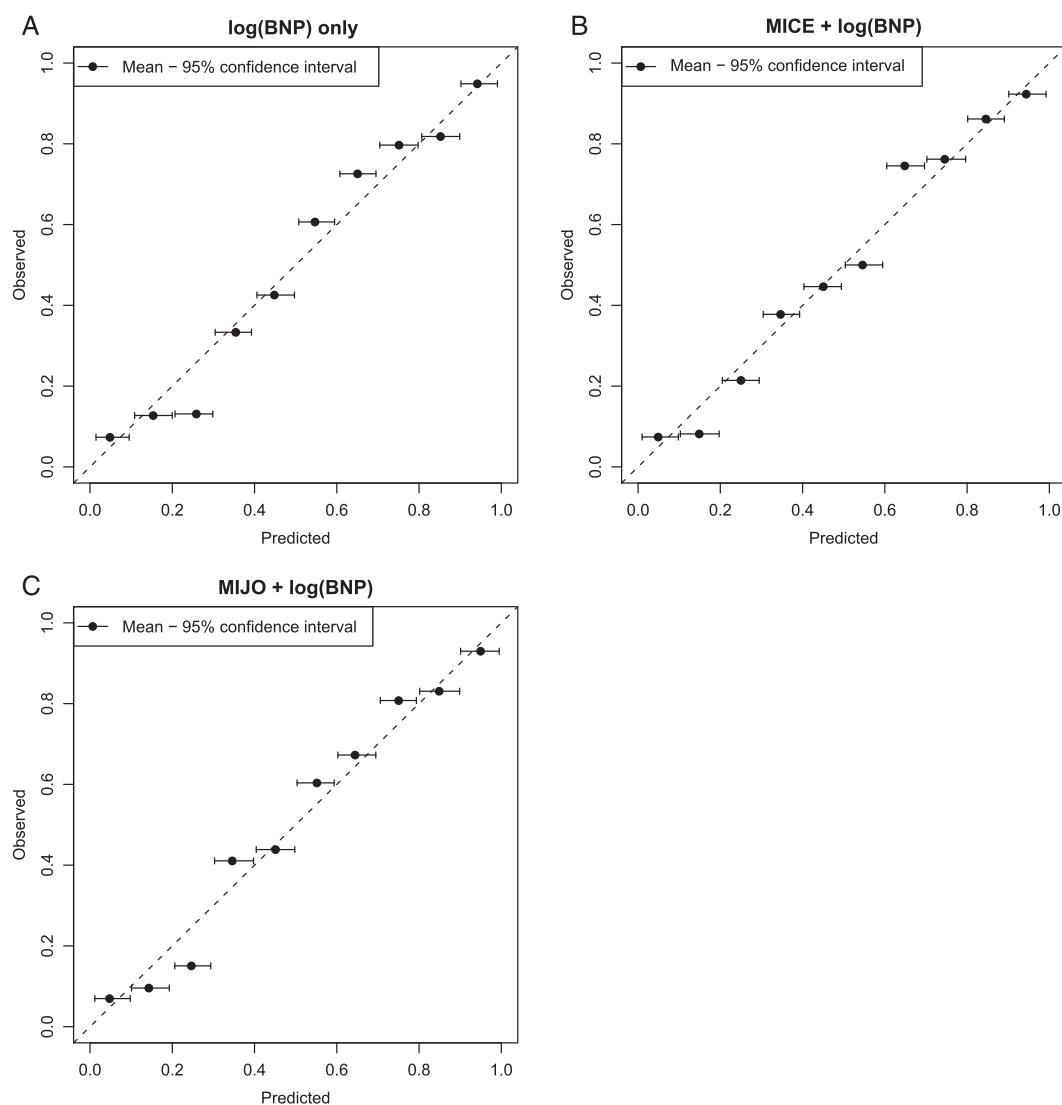
rules. However, the validation study was conducted in Ireland in which general practitioners are health care gatekeepers. Therefore, the results may be less generalizable to health care settings where primary care does not play such a role. We were unable to validate four of the rules as some variables were not available. However, the missing variables such as cardiothoracic ratio and biomarkers such as adrenomedullin or atrial natriuretic peptide are not part of current guidelines and not widely used on a clinical basis. Also, following guidelines on validation of clinical prediction models, we wanted to refrain from simplification of these models by excluding those predictors that were not available and then re-fitting new model on our data, as this would lead to the development of even more new models. Simply omitting these missing predictors from the full models without further re-fitting would lead to structural underestimation of the risk and poor clinical

performance. Therefore, validation of these diagnostic prediction models was not undertaken. Because BNP was used in some cases to identify those who may not require echocardiography, this may have led to an overestimation of its predictive ability and highlights also the difficulty with the use of single markers in diagnostic prediction. It highlights another feature of clinical prediction rules in that often features associated with the diagnosis are part of the clinical prediction rule, and therefore, caution should be exercised when using features of a clinical prediction rule on their own rather than as part of the rule. However, the results here are similar to those of a recently published paper highlighting that NT-proBNP testing alone performed as well as the MICE (male, infarction, crepitations, oedema) rule in determining which patients presenting with possible heart failure symptoms went on to have a diagnosis of heart failure.[20]

Other systematic reviews on clinical prediction rules in heart failure have focused on risk prediction, rehospitalization, and diagnosis of acute heart failure.[21,22] These studies have also shown a lack of impact analysis, and this has been noted in other areas of cardiovascular disease also.[23] This lack of clinical application of these rules makes it difficult to determine the impact that they have on patient care, physician behaviour, and health care costs. This review clearly indicates that many of the studies are similar in design as corresponding models often include similar predictors and include similar patient populations. Therefore, researchers are often repeating the same process and mostly introduce implicit knowledge when developing a prediction model from scratch. Over the past few decades, statistical methods for building prediction models using established knowledge have substantially improved, and these can be achieved by refining, updating, extending, and even combining the most promising existing models for diagnosis of heart failure in the general population, as was undertaken in the most recent model.[18]

Ideally, systematic reviews also guide evidence informed health decision making, in this case leading to recommendations on which models to advocate or even use in guidelines for the diagnosis of heart failure. Given the lack of clinical impact studies, we believe it is impossible to recommend which specific model or models should be used in which setting or location. Similarly, the adaptation of models for use in guidelines[6] without validation may lead to diagnostic inaccuracy and should be discouraged.

The validity, and thus potential impact, of models for the diagnosis of heart failure could substantially be improved by making better use of existing evidence, rather than starting from scratch to develop yet another model and undertaking clinical impact studies to determine the utility of these models. The most suitable and promising models for a particular targeted population should be identified and subsequently be validated (and if necessary tailored to the situation at hand), allowing for head-to-head comparisons such as previously performed for prediction models for the development of type 2 diabetes.[24]

Conclusions

The use of clinical prediction rules in chronic heart failure may help improve access to diagnostics for those most likely to have significant findings. However, the rules identified during this review require further validation and clinical impact studies before integration into guidelines. In particular, the approach of using a clinical prediction rule compared with natriuretic peptide testing or use of a multimarker panel as a rule-out test in people presenting with symptoms of chronic heart failure should be evaluated.

Conflict of interest

None declared.

Funding

Supporting information

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

Appendix S1. (A) CHARMS key items. (B) Search strategy. (C) Assessment of bias in retrieved clinical prediction. (D-F) CHARMS checklist.

References

1. Remes J, Miettinen H, Reunanen A, Pyorala K. Validity of clinical diagnosis of heart failure in primary health care. *Eur Heart J* 1991; **12**: 315–321.
2. Cowie MR. The heart failure epidemic: a UK perspective. *Echo Res Pract* 2017; **4**: R15–R20.
3. Hobbs FD, 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.
4. Dahlstrom U, Hakansson J, Swedberg K, Waldenstrom A. Adequacy of diagnosis and treatment of chronic heart failure in primary health care in Sweden. *Eur J Heart Fail* 2009; **11**: 92–98.
5. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA* 2000; **284**: 79–84.
6. Excellence NIfC. Chronic heart failure national clinical guideline for diagnosis and management in primary and secondary care NICE Clinical Guideline No 108 [Available from: https://www.nice.org.uk/guidance/cg108/evidence/full-guideline-136060525.
7. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; **6**: e1000097.

8. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, Reitsma JB, Collins GS. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 2014; **11**: e1001744.

9. Keogh C, Wallace E, O'Brien KK, Galvin R, Smith SM, Lewis C, Cummins A, Cousins G, Dimitrov BD, Fahey T. Developing an international register of clinical prediction rules for use in primary care: a descriptive analysis. *Ann Fam Med* 2014; **12**: 359–366.

10. Fahey T, Jeyaseelan S, McCowan C, Carr E, Goudie BM, Pringle SD, Donnan P, Sullivan F, Struthers A. Diagnosis of left ventricular systolic dysfunction (LVSD): development and validation of a clinical prediction rule in primary care. *Fam Pract* 2007; **24**: 628–635.

11. Rutten FH, Moons KG, Cramer MJ, Grobbee DE, Zuithoff NP, 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.

12. Yamamoto K, Burnett JC Jr, Bermudez EA, Jougasaki M, Bailey KR, Redfield MM. Clinical criteria and biochemical markers for the detection of systolic dysfunction. *J Card Fail* 2000; **6**: 194–200.

13. 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.

14. Kelder JC, Cramer MJ, van Wijngaarden J, van Tooren R, Mosterd A, Moons KG, Lammers JW, Cowie MR, Grobbee DE, Hoes AW. The diagnostic value of physical examination and additional testing in primary care patients with suspected heart failure. *Circulation* 2011; **124**: 2865–2873.

15. Oudejans I, Mosterd A, Bloemen JA, Valk MJ, van Velzen E, Wielders JP, Zuithoff NP, Rutten FH, Hoes AW. Clinical evaluation of geriatric outpatients with suspected heart failure: value of symptoms, signs, and additional tests. *Eur J Heart Fail* 2011; **13**: 518–527.

16. Roalfe AK, Mant J, Doust JA, Barton P, Cowie MR, Glasziou P, Mant D, McManus RJ, Holder R, Deeks JJ, Doughty RN, Hoes AW, Fletcher K, Hobbs FDR. Development and initial validation of a simple clinical decision tool to predict the presence of heart failure in primary care: the MICE (male, infarction, crepitations, edema) rule. *Eur J Heart Fail* 2012; **14**: 1000–1008.

17. Boonman-de Winter LJ, Rutten FH, Cramer MJ, Landman MJ, Zuithoff NP, Liem AH, Hoes AW. Efficiently screening heart failure in patients with type 2 diabetes. *Eur J Heart Fail* 2015; **17**: 187–195.

18. van Riet EE, Hoes AW, Limburg A, Landman MA, Kemperman H, Rutten FH. 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.

19. Tait L, Roalfe AK, Mant J, Cowie MR, Deeks JJ, Iles R, Barton PM, Taylor CJ, Derit M, Hobbs FDR. The REFER (REFer for EchocaRdiogram) protocol: a prospective validation of a clinical decision rule, NT-proBNP, or their combination, in the diagnosis of heart failure in primary care. Rationale and design. *BMC Cardiovasc Disord* 2012; **12**: 97.

20. Taylor CJ, Roalfe AK, Iles R, Hobbs FR, investigators R, Barton P, Deeks J, McCahon D, Cowie MR, Sutton G. Primary care REFerral for EchocaRdiogram (REFER) in heart failure: a diagnostic accuracy study. *Br J Gen Pract* 2017; **67**: e94–e102.

21. Rahimi K, Bennett D, Conrad N, Williams TM, Basu J, Dwight J, Woodward M, Patel A, McMurray J, MacMahon S. Risk prediction in patients with heart failure: a systematic review and analysis. *JACC Heart failure* 2014; **2**: 440–446.

22. Martindale JL, Wakai A, Collins SP, Levy PD, Diercks D, Hiestand BC, Fermann GJ, deSouza I, Sinert R. Diagnosing acute heart failure in the emergency department: a systematic review and meta-analysis. *Acad Emerg Med* 2016; **23**: 223–242.

23. Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I Lassale CM, Siontis GC, Chiocchia V, Roberts C, Schlüssel MM, Gerry S, Heus P, van der Schouw YT, Peelen LM, Moons KG. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 2016; **353**: i2416.

24. Abbasi A, Peelen LM, Corpeleijn E, van der Schouw YT, Stolk RP, Spijkerman AM, van der A DL, Moons KG, Navis G, Bakker SJL, Beulens JWJ. Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. *BMJ* 2012; **345**: e5900.