Diagnosing Heart Failure in Primary Care

Johannes Christiaan Kelder

Diagnosing Heart Failure in Primary Care.
Thesis. Utrecht University, The Netherlands. With a summary in Dutch.

ISBN 978-94-6108-315-9
Author Johannes Christiaan Kelder
Cover Kees-Jan Kelder, Sandra Verheem, Sanne Kelder
Print Gildeprint Drukkerijen, Enschede, The Netherlands

Diagnosing Heart Failure in Primary Care

Diagnosticeren van hartfalen in de eerste lijn

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht

op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan,

ingevolge het besluit van het college voor promoties

in het openbaar te verdedigen op

dinsdag 26 juni 2012 des ochtends te 10.30 uur

door

Johannes Christiaan Kelder

geboren op 16 oktober 1958 te Bussum

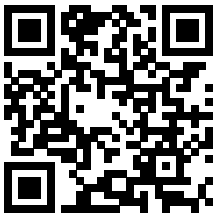
Promotoren: Prof. dr. A.W. Hoes

Prof. dr. D.E. Grobbee

Co-promotor: Dr. M.J.M. Cramer

Contents

General introduction



General introduction

Mrs. Robinson is a 76 years old widow and suffers from high blood pressure and diabetes, for which she is prescribed respectively chlortalidone and metformin besides a diet. She recently visited the office of her general practitioner (GP) because she isn't able to climb the stairs at her home or do her shopping anymore without losing her breath. She had just arrived as she enters the consulting room, still somewhat short of breath. As usual, she walks with a stick due to osteoarthritis of the left hip, for which she regularly uses a NSAID. In the following conversation ("history taking") it becomes clear that she is not in any pain (other than her osteoarthritis), there is no fever and there were no recent changes in medication. She lost some of her appetite, although gaining some weight, around 3 kilograms during the last few weeks. There was a period of coughing but this has almost resolved. Rhetorically she asks the GP whether her age is to blame for this discomfort.

As a diagnostician, her GP has already obtained -in part knowingly- much information to build on, but is this sufficient to make a diagnosis. Of note, the patient does not have a diagnosis, only the doctor may have.[1] Anyway, this initial diagnostic phase needs to end with a conclusion and a subsequent action, albeit that this conclusion could be to exclude any specific disease and the action could simply be reassurance of the patient or watchful waiting. The conclusion is hardly ever a diagnosis in the Aristotelean view, where the determination of the nature of a disease is picked from a pre-existing list of diseases. As such, diagnosis could be studied and modeled in abstraction from the rest of clinical reasoning. Rather, a diagnosis is made because of its role in informing etiology and prognosis, and the acts of altering prognosis: therapy and prevention.[2]

The beginning of the diagnostic phase is a patient suspected of the disease (which can be a disease, syndrome, as well as 'not otherwise specified'). The next part of the diagnostic phase is selecting tests (including additional items from history taking and physical examination) and interpreting the results of these tests conditional on the knowledge thus far. The diagnostic process involves sequential tests, which have a natural order. Starting with the information obtained from medical history taking (demographics, symptoms, past medical history, medication use), followed by physical examination. In the next step the diagnostician has to consider readily available laboratory test, such as blood sample tests , chest X-ray, ECG and pulmonary function tests. The end of the diagnostic phase is a decision based on the estimated probability of the disease, with one of three outcomes: sufficient certainty that the disease is present leading to targeted interventions (which can be no treatment), insufficient

certainty leading to the next diagnostic phase (often implying referral) or sufficient certainty the disease is absent.[3]

The disease of interest in this thesis is heart failure, which is defined as a syndrome in which patients have the following features[4]:

- Symptoms typical of heart failure (breathlessness at rest or on exercise, fatigue, tiredness, ankle swelling) and
- Signs typical of heart failure (tachycardia, tachypnoea, pulmonary rales, pleural effusion, raised jugular venous pressure, peripheral oedema, hepatomegaly) and
- Objective evidence of a structural or functional abnormality of the heart at rest (cardiomegaly, third heart sound, cardiac murmurs, abnormality on the echocardiogram, raised natriuretic peptide concentration).

Specifically the interest lies in the slow onset (or non-acute) manifestation of the syndrome as opposed to acute heart failure, where urgent therapeutic interventions are warranted. Typically, patients with slow onset heart failure present themselves to the GP first, thus the initial diagnostic assessment is made by the GP.

The aim of this thesis is to assess diagnostic strategies in patients suspected of heart failure in primary care by means of empirical diagnostic research. In the literature much attention has been paid to B-type Natriuretic Peptide (BNP, a blood sample measurement), as a diagnostic test for heart failure. It is nowadays an intricate part of the definition of heart failure (see above). In this thesis, BNP is used as a leverage to assess diagnostic strategies encompassing not only BNP but other diagnostic tests as well, with special emphasis on physical examination.

In Chapter 1 the focus is on design issues of clinically relevant diagnostic research. Four topics are discussed in more detail i) inclusion of the relevant patient population (domain); ii) the application of a multivariable approach; iii) the natural clinical hierarchy of diagnostic tests with emphasis on the added value of the new test and iv) does the study enable the prediction of probabilities of the disease? The diagnostic value of BNP in suspected heart failure is used as an example.

In Chapter 2 the example of the BNP test from chapter 1 is further explored in an individual patient data meta-analysis making use of three previously published diagnostic studies. The aim of this meta-analysis was to quantify the additional diagnostic value of BNP.

Chapter 3 introduces the design and presents the major findings of the Utrecht Heart Failure Organisation - Initial Assessment (UHFO-IA) study. All patients

presenting with symptoms and signs suggestive of heart failure (typically dyspnea, fatigue, signs of fluid retention) in primary care were eligible for inclusion in the study. Patients were referred to 1 of 8 rapid access outpatient clinics by primary care physicians based in the catchment area of the participating hospitals throughout the Netherlands. Seven hundred and twenty one patients were included and underwent a full diagnostic work-up, including chest X-ray, BNP, ECG, spirometry as well as echocardiography. In this large study different diagnostic strategies are explored, with the aim to derive a multivariable diagnostic model to help GPs to estimate an individual patient's probability of heart failure based on a limited number of diagnostic items.

In Chapter 4 we explore the question whether it is feasible in the initial diagnostic assessment to further specify the diagnosis of heart failure by phenotype, i.e. heart failure with reduced ejection fraction and heart failure with preserved ejection fraction.

The use of furosemide test treatment as a diagnostic test in suspected heart failure seems popular among Dutch GPs, but the diagnostic value of this test treatment has not been assessed. In Chapter 5, in a sub-study of the UHFO-IA, the diagnostic value of the use of a short course of administering furosemide in trying to establish a diagnosis of slow-onset heart failure is determined.

In another sub-study of the UHFO-IA, presented in Chapter 6, three different automated assays which measure BNP or NT-proBNP in a blood sample are compared and the implications for clinical practice of the differences are assessed.

In Chapter 7 the economic implications of several diagnostic strategies in suspected heart failure are explored. The full course of the patient is taken into account, starting from being suspected of the disease, being diagnosed timely or later, being treated and experiencing the vicious nature of heart failure until an often early death.

In the general discussion concluding remarks and advice for future diagnostic research are given, with special emphasis on the need for adequate sample size to identify the value of physical examination.

References

1. Miettinen OS. *Epidemiological Research: Terms and Concepts*. Dordrecht: Springer; 2011; Page 10.
2. Whitbeck C. What is diagnosis? Some critical reflections. Theoretical Medicine and Bioethics. 1981;2:319-329.
3. Grobbee DE, Hoes AW. *Clinical epidemiology: principles, methods, and applications for clinical research*. Boston: Jones & Bartlett Publishers; 2008; Page 58-102.
4. 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.

Chapter 1

Clinically relevant diagnostic research in primary care: the example of B-type natriuretic peptides in the detection of heart failure

based on:

Family Practice. 2009 Feb;26(1):69-74. Epub 2008 Dec 3.

Clinically relevant diagnostic research in primary care: the example of B-type natriuretic peptides in the detection of heart failure.

Kelder JC, Rutten FH, Hoes AW.

Abstract

With the emergence of novel diagnostic tests, e.g. point-of-care tests, clinically relevant empirical evidence is needed to assess whether such a test should be used in daily practice. With the example of the value of B-type natriuretic peptides (BNP) in the diagnostic assessment of suspected heart failure, we will discuss the major methodological issues crucial in diagnostic research; most notably the choice of the study population and the data analysis with a multivariable approach.

BNP have been studied extensively in the emergency care setting, and also several studies in the primary care are available. The usefulness of this test when applied in combination with other readily available tests is still not adequately addressed in the relevant patient domain, i.e. those who are clinically suspected of heart failure by their GP. Future diagnostic research in primary care should be targeted much more at answering the clinically relevant question 'Is it useful to add this (new) test to the other tests I usually perform, including history taking and physical examination, in patients I suspect of having a certain disease'.

Introduction

Diagnosing is a major and challenging task for general practitioners (GPs), and setting a diagnosis is the starting point for prognostication and adequate treatment. In daily practice, the diagnostic process starts with a patient presenting himself with certain complaints, and after history taking and physical examination, including a patient's medical history, the GP suspects a particular disorder or disease. This process is often implicit, as are the estimates of the probability of the suspected disease. The latter is partly due to the lack of evidence from adequate quantitative diagnostic research.[1] Estimated disease probabilities will guide the GP in choosing whether an additional test is needed, or whether the GP gained sufficient assurance to rule in or exclude the suspected disease. When subsequent tests provide information beyond already available test results (i.e. have added value), they 'update' the probability of the presence or absence of the disease. This process continues until the GP is confident enough to take clinical (treatment) decisions, for example because the GP's treatment threshold, say 75% or 90% probability, is reached.

With the example of B-type natriuretic peptide (BNP) in the diagnostic assessment of patients suspected of heart failure in primary care, we will discuss which evidence from diagnostic research should be available (and how such studies should be designed) before such a new diagnostic test can be qualified as an useful additional tool in every day primary care practice.

We will focus on the following design issues of such clinically relevant diagnostic research: (i) inclusion of the relevant patient population (domain); (ii) the application of a multivariable approach; (iii) the natural clinical hierarchy with emphasis on the *added* value of the new test and (iv) does the study enable the prediction of probabilities of the disease? (Boxes 1 and 2).

Box 1. An example

A company representative offers you some specimen of a new diagnostic test to detect heart failure in your practice; a bedside test to measure B-type natriuretic peptide (BNP). In short, BNP is a peptide that is mainly produced in the myocardium of the left ventricle in response to stretch, generally caused by increased volume and pressure overload. The latter is the mainstay of heart failure.[2] The representative shows you tables with very high values of sensitivity and specificity of both more than 90%! Just a few drops of capillary blood on a test strip put in the machine, and a few minutes later the test result is available. 'Something every general practitioner should have: You save the patient and yourself the inconvenience of referral for echocardiography.'

The next day, you suspect one of the patients presenting himself at your office of having heart failure: a man, 78 years of age, 30 pack years of cigarette smoking, who receives medication for diabetes and hypertension, and who had a myocardial infarction 10 years ago. He now complains of slowly progressive breathlessness while walking with his dog. Since three weeks he sleeps with an extra pillow. On physical examination, the blood pressure is 164/92 mmHg, the pulse frequency is regular with 92 beats/min and the patient has minor peripheral oedema at both legs. On palpation, the apex beat is broadened and sustained in left lateral position and with auscultation a holosystolic cardiac murmur at the apex, suggestive of mitral valve insufficiency, is audible. No abnormal pulmonary sounds are audible. Unable to rule in or exclude heart failure, you cannot decide about the initiation of treatment, and you want to perform a simple non-burdening additional investigation. You wonder whether the new BNP test offers help at this very moment. During coffee break, you do a quick internet search with Google and Pubmed. You find impressive numbers, 98% diagnostic accuracy, and you have more than 1000 hits with the search terms ‘diagnosis, BNP, heart failure’. Can you now conclude that this is a good test that perfectly fits as a first additional step in your diagnostic assessment of this type of patient?

Box 2. What is BNP?

Apart from its mechanical role, the heart also has an endocrine function. Myocytes produce natriuretic peptides, a family of vasoactive hormones, in response to myocyte stretch and increased wall tension, primarily caused by pressure and/or volume overload.[3] The atria mainly produce atrial natriuretic peptide, while BNP is produced mainly in the ventricles. The prehormone proBNP is cleaved upon release in the circulation in equal portions of the biological active hormone BNP and inactive split product amino-terminal proBNP (NT-proBNP). Because of different ways of clearance, the half-life of BNP is 20 minutes, and that of NT-proBNP is 120 minutes. The bio-active hormone BNP augments urinary volume and urinary sodium excretion, relaxes vascular smooth muscles and thus counterbalances the sympathetic nervous and renin-angiotensin system, trying to maintain circulatory homeostasis.[2]

Both B-type natriuretic peptides can be measured in plasma, serum, or full blood by fully automated and commercially available assays from different companies, including reliable point-of-care tests. Nowadays immunoassays are based on two monoclonal antibodies, and differences between the most often used assays of different companies are small and generally not of clinical relevance.[4] Blood concentrations of BNP and NT-proBNP are expressed in pg/mL or pmol/L. The conversion factor for BNP is 1 pg/mL = 0.289 pmol/L, and for NT-proBNP it is 1 pg/mL = 0.118 pmol/L. Nowadays, pg/mL is most often used. B-type natriuretic peptides can be used for diagnostic and prognostic purposes in different clinical settings, in particular in patients with (suspected) heart failure.[5-8] More recently it has also been suggested that B-type natriuretic peptides can be helpful in titrating drug treatment in heart failure patients.[9-11]

Research in the relevant clinical domain

First of all, the population under study should be the population where there is a diagnostic dilemma. There is a diagnostic problem neither in patients with overt (or established) disease nor in healthy people. Studies that 'screen' *all* patients in the community, or those in the community with certain risk factors for a disease also lack the 'diagnostic dilemma'. The aforementioned examples do not represent the domain of interest for clinically relevant diagnostic research, namely patients suspected of a certain disease by the doctor, based on symptoms and signs, in whom the doctor considers additional diagnostic testing. Thus, diagnostic accuracy data obtained in other populations, such as from the population at large of those with established disease, are not easily applicable to individual patients from the domain of interest; i.e. those who consult the practising clinician with symptoms suggestive of the disease of interest.

Consequently, diagnostic accuracy results with cut-off levels for BNP from population studies or studies in high risk patients for heart failure (without suspicion of having heart failure) not very helpful.[12,13]

In addition, results from diagnostic research from secondary care cannot be applied uncritically to the primary care setting, even when performed in patients in whom a diagnostic problem exists. On the one hand this is caused by higher prevalence rates ((that is, higher pretest likelihood or prior probability) of the relevant diseases in suspected patients in secondary care compared to those in primary care, due to the selection process of referral. As a result, positive predictive values of tests in secondary care are higher, and negative predictive tests lower than in primary care. Moreover, more advanced disease stages with higher levels of diagnostic markers will be presented in secondary care compared to primary care. This will in general provide higher sensitivity and lower specificity in secondary care compared to primary care, because sensitivity and specificity are not 'fixed' test characteristics, and vary according to differences in severity of disease.[14] In laboratory tests with continuous values such as with BNP, this will lead to (much) higher values. When BNP is digitomized, the ‘optimal’ (highest sum of sensitivity and specificity) cutpoint will be much higher in the secondary care setting compared to primary care (see also Box 3 about cutpoints of BNP).

In the example of heart failure, the population under study should be patients from primary care suspected of heart failure because of typical symptoms and signs such as (slowly progressive) shortness of breath, fatigue, and/or peripheral oedema. In those patients a diagnostic challenge exists and additional tests such as BNP and electrocardiography can be considered.

Because heart failure is a chronic progressive disease and effective morbidity and mortality reducing treatment is available, the GP has the difficult task to detect early stages of the disease, at a point of time when symptoms and signs are less overt and specific and plasma levels of BNP or the N-terminal fraction of BNP (NT-proBNP) often may not differ much from normal values;[4] a situation very different from secondary care, where (NT-pro)BNP is typically measured in acute dyspnoeic patients visiting the emergency department, with typical signs and symptoms and often very high values of (NT-pro)BNP in those with heart failure.[15,16]

Thus, the usefulness of a diagnostic test such as (NT-pro)BNP in primary care can only be assessed when the diagnostic research is performed in close adherence to daily primary care practice. For this, consecutive patients suspected of heart failure by the GP should be investigated with as few as possible exclusion criteria to ensure wide applicability of the findings.

Box 3. Cutpoints of B-type natriuretic peptides

Blood levels of BNP and NT-proBNP are continuous. To facilitate the use of (NT-pro)BNP, single cutpoints have been proposed. The disadvantage, however, is the dramatic reduction of available information, since levels in daily practice range from nearly zero to ten thousands of pg/ml. A partial solution to this problem is to use two cutpoints, one exclusionary (rule out) and one confirmatory (rule in) threshold. Thus, more patients can be correctly classified as no heart failure on the one hand and (probable) heart failure on the other. In between the cut-points, however, a grey zone is then created in which (combinations of) other diseases and phenomena such as renal disease, pulmonary embolism, pulmonary artery hypertension, cardiac ischaemia, and advanced age (>75 years) have to be considered as causes of increased blood (NT-pro)BNP values.[17]

Many cutpoints for (NT-pro)BNP have been published, mainly depending on the patient domain. Studies from secondary care, assessing patients with acute dyspnoea published 'ruling out' cutpoints of NT-proBNP of 300 pg/ml (or 400 pg/ml) and BNP of 100 pg/ml.[15,16,20] Studies from primary care assessing patients with slowly progressive dyspnoea and suspected of slow onset new heart failure published (much) lower values, with 'optimal single' cutpoints of about 125 pg/ml for NT-proBNP and 35–77 pg/ml for BNP, ruling out heart failure >95%.[15;16;18;19] For the 'ruling in' cutpoint 1800 pg/ml (or 2000 pg/ml) for NT-proBNP and 35-77 pg/ml for BNP has been suggested for the secondary care population[15,20], while such a cutpoint is lacking for primary care patients. The biological basis for different cutpoints between secondary and primary care is disease severity, with high intracardiac pressures and volume overload and thus high plasma levels of BNP and NT-proBNP in acute dyspnoeic patients (suspected of acute new onset heart failure) assessed in the secondary care setting versus low intracardiac pressures and volumes in patients with chronic or slowly progressive dyspnoea patients from primary care and suspected of slow onset new heart failure.

Multivariable approach, natural clinical hierarchy, and the prediction of absolute probabilities

In daily practice, a diagnosis is never based on a single diagnostic test. Rather, multiple tests are applied, including age, gender, history taking, physical examination and most often simple laboratory tests. As a consequence, also diagnostic research should take multiple tests into account. This *multivariable* diagnostic process has a natural hierarchy. First, age, sex and earlier diagnoses (such as myocardial infarction) will be considered, since these parameters are known at the time of or even before the presentation of the complaints. Subsequently history taking and physical examination will be performed, followed by easily available additional test (e.g. ECG, lung function tests or some laboratory tests). Finally, more complicated, expensive and patient-burdening tests may be considered (e.g. chest X-ray, echocardiography) that often require patient referral. The data analysis in diagnostic research should therefore incorporate multivariable techniques in which multiple diagnostic tests in their proper sequence (natural clinical hierarchy) can be assessed.[21] Thus, before a new diagnostic test can be applied in every day practice, its value in *improving* the prediction of the *absolute* probability of the suspected disease in a clinically relevant manner, *beyond* what can be derived from readily available information obtained from the clinical assessment (notably signs and symptoms). In other words, a new diagnostic test should prove to have *added* value.

Apart from showing that a (new) test independently contributes to the detection or exclusion of a disease, the extent of this contribution can and should be quantified. This can be done by using multivariable techniques (usually multiple logistic regression analysis) with the disease as the dependent variable and the diagnostic tests (including signs and symptoms and the new test(s)) as independent variables. Once the optimal combination of tests is established, each test gets a 'weight' according to its contribution to detect or exclude a disease, and thus a diagnostic score can be calculated in which an individual score translates to an absolute probability of the disease (ranging from 0% to 100%). In a *diagnostic or clinical prediction rule*, diagnostic score thresholds can be related to subsequent patient management, e.g. ruling out of the disease/assuring the patient in case of low scores and considering the disease to be present/initiating targeted drug therapy in case of high scores. Applying such a (simplified) rule in daily practice may assist the physician in his decision making.[22] Although prediction rules are important tools, the physician's own judgement (based on clinical assessment of the patient and the doctor's clinical experience) should not be discarded and could even overrule the clinical prediction rule.

Examples of widely used decision rules are Well's rule to diagnose or exclude deep venous thrombosis, a similar, adapted rule for the primary care setting by Oudega et al. and the Ottawa Ankle rule.[23-25]

In the example of heart failure, the practising GP should not be interested to know how good a single BNP test diagnoses or rules out a disease, (nor in this test's accuracy characteristics such as sensitivity, specificity, positive and negative predictive value), but needs to know whether BNP has *added* value beyond the clinical assessment in patients he suspects of having heart failure, and whether electrocardiography and chest X-ray are viable alternatives. The additional tests either decrease or increase the estimated probability of heart failure in case these additional tests are negative or positive respectively. When such probabilities then reach a threshold below which the physician is confident that the disease can be ruled out or beyond which she considers the disease present (and initiates therapy) the clinical relevance of the additional test is evident.[26] In addition, she may wish to have an easy to use diagnostic rule to calculate the absolute risk of heart failure in order to estimate the level of certainty, end the diagnostic process, and enter the prognostic and therapeutic realm.

The 'evidence': diagnostic studies with BNP to detect heart failure in the primary care setting

We did not apply a standard systematic review methodology, but used available information of three recently published systematic reviews by Doust et al.,[27] Hill et al.[28] and Davenport et al.[6] of the diagnostic accuracy of B-type natriuretic peptides (BNP or NT-proBNP) in detecting heart failure or reduced left ventricular ejection fraction (LVEF). We selected studies in the relevant domain, that is, patients suspected of heart failure in the primary care setting (leaving out studies that screened high risk or subpopulations or studies performed in the hospital setting). Furthermore, we selected studies in which the outcome was heart failure (with or without reduced LVEF), that is, signs and/or symptoms of heart failure in combination with objective evidence of (systolic and/or diastolic) ventricular dysfunction, preferably using a panel of clinicians to assess the final diagnosis.[9,19,21] Thus, studies with as outcome heart failure with reduced LVEF (and thus missing about half of all patients with heart failure, namely those with a preserved LVEF) were excluded. Only four studies met the two selection criteria.[18,29,30,31] Table 1 presents relevant characteristics of the four studies. The study patients were predominantly older than 70 years of age, as is typical for patients suspected of heart failure in primary care.[32] The pretest likelihood of the study patients ranged from 23% to 34%, representing the prior probability for a patient the GP suspects of having heart failure.

None of the studies performed a complete multivariable analysis, providing data of the added value of either (NT-pro)BNP on top of signs and symptoms. As a consequence, none of the studies provided absolute probabilities of heart failure or a diagnostic prediction rule applicable in the primary care setting.

Conclusions

Diagnostic studies in primary care should provide answers to diagnostic dilemmas of practising GPs. For this, they should be performed in patients suspected of a certain disease by their GP usually as a result of prior history, signs and symptoms. A multivariable data approach should be undertaken, quantifying the *added* value of a test on top of readily available information such as demographic data and the clinical assessment (i.e. taking te natural hierarchy of the diagnostic process into account). Although this seems reasonable and straightforward, with the example of heart failure BNP (or NT-proBNP) in the primary care setting, we showed that studies executed in the aforementioned preferable way are lacking in literature.

BNP has been studied extensively since its first description in 1988 in many different patient populations, including primary care patients. However, the majority of studies was 'test research', providing accuracy data. (NT-pro)BNP showed to have high negative predictive values as a single test, both in the primary and secondary care setting. Importantly, however, at completely different 'exclusionary' cutpoints. As such, the test may be useful as a ruling out test, however, at different cutpoints in primary care compared to secondary care.

Importantly, a straightforward answer to the question how should GPs use this test in combination with his clinical assessment and other readily available tests such as electrocardiography and chest X-ray, is still lacking.

TABLE 1 Diagnostic studies with BNP in patients suspected of heart failure from primary care with as outcome heart failure (with or without reduced LVEF)

Study	Number of participants, mean age	Prevalence of heart failure	Data analysis
Cowie et al.[18]	106, age range 24-87 years[a]	27%	Univariably, partly multivariable; BNP, N-ANP, ANP and chest radiograph
Nielsen et al.[30]	345, median age 65 years (range 18-89)	23%	Univariably
Wright et al.[29]	305, mean age 72 years (range 40-95)	25%	Randomized trial in which addition of BNP to signs and symptoms was assessed; multivariably on post-assessment variables
Zaphiriou et al.[31]	305, mean age 72 years (range 40-95)	34%	Univariably, however, the additional value of electrocardiogram on top of BNP was considered

References

1. Richardson WS. Five Uneasy Pieces About Pre-test Probability. Journal of General Internal Medicine 2002; 17(11):891-892.
2. Levin ER, Gardner DG, Samson WK. Natriuretic peptides. N Engl J Med 1998; 339(5):321-328.
3. Weber M, Hamm C. Role of B-type natriuretic peptide (BNP) and NT-proBNP in clinical routine. Heart 2006; 92(6):843-849.
4. Rutten FH, Cramer MJ, Zuithoff NP, Lammers JW, Verweij W, Grobbee DE et al. Comparison of B-type natriuretic peptide assays for identifying heart failure in stable elderly patients with a clinical diagnosis of chronic obstructive pulmonary disease. Eur J Heart Fail 2007; 9(6-7):651-659.
5. Battaglia M, Pewsner D, Juni P, Egger M, Bucher HC, Bachmann LM. Accuracy of B-Type Natriuretic Peptide Tests to Exclude Congestive Heart Failure: Systematic Review of Test Accuracy Studies. Archives of Internal Medicine 2006; 166(10):1073-1080.
6. Davenport C, Cheng EYL, Kwok YTT, Lai AHO, Wakabayashi T, Hyde C et al. Assessing the diagnostic test accuracy of natriuretic peptides and ECG in the diagnosis of left ventricular systolic dysfunction: a systematic review and meta-analysis. British Journal of General Practice 2006; 56:48-56.
7. Mair FS, Crowley TS, Bundred PE. Prevalence, aetiology and management of heart failure in general practice. Br J Gen Pract 1996; 46(403):77-79.
8. Struthers A, Lang C. The potential to improve primary prevention in the future by using BNP/N-BNP as an indicator of silent 'pancardiac' target organ damage: BNP/N-BNP could become for the heart what microalbuminuria is for the kidney. Eur Heart J 2007; 28(14):1678-1682.
9. Swedberg K, Cleland J, Dargie H, Drexler H, Follath F, Komajda M et al. 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(11):1115-1140.
10. Troughton RW, Frampton CM, Yandle TG, Espiner EA, Nicholls MG, Richards AM. Treatment of heart failure guided by plasma aminoterminal brain natriuretic peptide (N-BNP) concentrations. Lancet 2000; 355(9210):1126-1130.
11. Jourdain P, Jondeau G, Funck F, Gueffet P, Le Helloco A, Donal E et al. Plasma brain natriuretic peptide-guided therapy to improve outcome in heart failure: the STARS-BNP Multicenter Study. J Am Coll Cardiol 2007; 49(16):1733-1739.

12. Hobbs FD, Davis RC, Roalfe AK, Hare R, Davies MK, Kenkre JE. Reliability of N-terminal pro-brain natriuretic peptide assay in diagnosis of heart failure: cohort study in representative and high risk community populations. BMJ 2002; 324(7352):1498.
13. McDonagh TA, Robb SD, Murdoch DR, Morton JJ, Ford I, Morrison CE et al. Biochemical detection of left-ventricular systolic dysfunction. Lancet 1998; 351(9095):9-13.
14. Moons KG, van Es GA, Deckers JW, Habbema JD, Grobbee DE. Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: a clinical example. Epidemiology 1997; 8(1):12-17.
15. Januzzi JL, van Kimmenade R, Lainchbury J, Bayes-Genis A, Ordonez-Llanos J, Santalo-Bel M et al. NT-proBNP testing for diagnosis and short-term prognosis in acute destabilized heart failure: an international pooled analysis of 1256 patients: The International Collaborative of NT-proBNP Study. Eur Heart J 2006; 27(3):330-337.
16. Maisel AS, Krishnaswamy P, Nowak RM, McCord J, Hollander JE, Duc P et al. Rapid Measurement of B-Type Natriuretic Peptide in the Emergency Diagnosis of Heart Failure. N Engl J Med 2002; 347(3):161-167.
17. de Lemos JA, McGuire DK, Drazner MH. B-type natriuretic peptide in cardiovascular disease. Lancet 2003; 362(9380):316-322.
18. Cowie MR, Struthers AD, Wood DA, Coats AJS, Thompson SG, Poole-Wilson PA et al. Value of natriuretic peptides in assessment of patients with possible new heart failure in primary care. Lancet 1997; 350(9088):1349-1353.
19. Rutten FH, Moons KGM, Cramer MJ, Grobbee DE, Zuithoff NPA, Lammers JW et al. Recognising heart failure in elderly patients with stable chronic obstructive pulmonary disease in primary care: cross sectional diagnostic study. BMJ 2005; 331(7529):1379.
20. 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.
21. Moons KG, Grobbee DE. Diagnostic studies as multivariable, prediction research. J Epidemiol Community Health 2002; 56(5):337-338.
22. Moons KG, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. Clin Chem 2004; 50(3):473-476.

23. Wells PS, Anderson DR, Bormanis J, Guy F, Mitchell M, Gray L et al. Value of assessment of pretest probability of deep-vein thrombosis in clinical management. Lancet 1997; 350(9094):1795-1798.
24. Oudega R, Moons KG, Hoes AW. Ruling out deep venous thrombosis in primary care. A simple diagnostic algorithm including D-dimer testing. Thromb Haemost 2005; 94(1):200-205.
25. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. Ann Emerg Med 1992; 21(4):384-390.
26. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med 1980; 302(20):1109-1117.
27. 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(18):1978-1984.
28. Hill SA, Balion CM, Santaguida P, McQueen MJ, Ismaila AS, Reichert SM et al. Evidence for the use of B-type natriuretic peptides for screening asymptomatic populations and for diagnosis in primary care. Clinical Biochemistry 2008; 41(4-5):240-249.
29. Wright SP, Doughty RN, Pearl A, Gamble GD, Whalley GA, Walsh HJ et al. Plasma amino-terminal pro-brain natriuretic peptide and accuracy of heart-failure diagnosis in primary care: a randomized, controlled trial. JACC 2003; 42(10):1793-1800.
30. Nielsen LS, Svanegaard J, Anders Klitgaard N, Egeblad H. N-terminal pro-brain natriuretic peptide for discriminating between cardiac and non-cardiac dyspnoea. Eur J Heart Fail 2004; 6(1):63-70.
31. Zaphiriou A, Robb S, Murray-Thomas T, Mendez G, Fox K, McDonagh T et al. 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(4):537-541.
32. 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(3):337-344.

Chapter 2

Quantifying the added value of BNP in suspected heart failure in general practice: an individual patient data meta-analysis

based on:

Heart. 2011 Jun;97(12):959-63. Epub 2011 Apr 8.

Quantifying the added value of BNP in suspected heart failure in general practice: an individual patient data meta-analysis.

Kelder JC, Cowie MR, McDonagh TA, Hardman SM, Grobbee DE, Cost B, Hoes AW.

Abstract

Objectives

Diagnosing early stages of heart failure with mild symptoms is difficult. B-type Natriuretic Peptide (BNP) has promising biochemical test characteristics, but its diagnostic yield on top of readily available diagnostic knowledge has not been sufficiently quantified in early stages of heart failure.

Design

Individual patient data meta analysis followed by external validation. The additional diagnostic yield of BNP above standard clinical information was compared to ECG and chest X-ray.

Patients and Methods

Derivation was performed on two existing datasets from Hillingdon (n=127) and Rotterdam (n=149) while the UK Natriuretic Peptide Study (n=306) served as validation dataset. Included were patients with suspected heart failure referred to a rapid-access diagnostic outpatient clinic. Case definition was according to the ESC guideline. Logistic regression was used to assess discrimination (with the c-statistic) and calibration.

Results

Of the 276 patients in the derivation set 30.8% had heart failure. The clinical model (encompassing age, gender, known coronary artery disease, diabetes, orthopnea, elevated jugular venous pressure, crackles, pitting edema and S3 gallop) had a c statistic of 0.79. Adding respectively chest X-ray, ECG or BNP to the clinical model increased the c statistic to 0.84, 0.85 and 0.92. Neither ECG nor chest X ray added significantly to the 'clinical plus BNP' model. All models had adequate calibration.

The 'clinical plus BNP' diagnostic model performed well in an independent cohort with comparable inclusion criteria (c-statistic=0.91 and adequate calibration). Using separate cut off values for 'ruling in' (typically implying referral for echocardiography) and for 'ruling out' heart failure creating a grey zone resulted in insufficient proportions of patients with a correct diagnosis.

Conclusion

BNP has considerable diagnostic value in addition to signs and symptoms in patients suspected of heart failure in primary care. However, using BNP alone with the currently recommended cut off levels is not sufficient to make a reliable diagnosis of heart failure.

Introduction

Heart failure is a clinical diagnosis with significant consequences for survival and quality of life. Early diagnosis is important in order to improve prognosis by initiating targeted therapy, notably ACE-inhibitors, beta-blockers and aldosterone antagonists as exemplified in various national and international clinical guidelines.[1, 2] Heart failure is, however, difficult to diagnose, especially in primary care where the diagnosis is usually based on signs and symptoms alone. Recognizing heart failure is often hindered by a lack of specificity of signs and symptoms, particularly in the early stages in a typically elderly population with comorbid conditions.[3] In general practice, "the goal is not to distinguish patients with heart failure from young healthy people but to distinguish patients with heart failure from elderly patients who have multiple cardiovascular risk factors and conditions but who do not have heart failure".[4]

B-type natriuretic peptide (BNP)

BNP and the inactive N-terminal fragment of the precursor (NT-proBNP) have emerged as likely candidates for the 'ultimate' biochemical test for heart failure. This is partly attributable to the straightforward physiologic foundation of the marker, i.e. stretch is the stimulus for BNP production in the ventricular myocyte[5] reflecting the filling pressure.[6] Nevertheless there are caveats in translating available research on the diagnostic value of BNP into daily practice.

Doust et al.[7], Hill et al.[8] and Mant et al.[9] recently performed systematic reviews of the published research regarding the accuracy of BNP as a diagnostic test for heart failure. They concluded that BNP was an accurate biomarker for heart failure. Most of the individual studies, however, have serious limitations. First, as an overall measure of diagnostic accuracy the diagnostic odds ratio, sensitivity and specificity were chosen. These parameters quantify diagnostic characteristics of the *single* BNP test as if the process of diagnosis is a univariable, screening-like activity, and prior clinical knowledge from history taking and physical examination does not play a role.[10] In daily practice, however, a diagnosis is based on multiple diagnostic tests that are performed in a natural hierarchical order. Demographic characteristics (notably age, sex), prior medical history and findings from history taking can be considered first line diagnostic 'tests', followed by physical examination. Consequently, the value of BNP or other additional tests should be expressed in terms of the *added* diagnostic yield, on top of the diagnostic knowledge, usually from history taking and physical examination, that is available to the clinician as part of a routine history and examination.

Second, the setting of studies differed considerably. Patients who present themselves at the emergency department with acute dyspnea are not comparable

to patients presenting to the general practitioner (GP) with gradually increasing complaints (slow onset), which has important implications for the diagnostic value of the BNP test.[11] Accordingly the ESC guideline for heart failure 2008 explicitly makes the distinction between new-onset heart failure (acute or slow), transient heart failure, and chronic heart failure.

Third, many diagnostic studies lack power to perform a multivariable analysis. To achieve sufficient power to detect a meaningful added value in a multivariable analysis the generally adopted heuristic states that for each variable considered one needs at least 10 patients with the disease. For example when the prevalence of disease is 20% one needs 250 patients to properly consider 5 diagnostic variables.

Fourth, external validation is an essential step before implementation, since diagnostic models will generally perform better in the database from which they were derived than in new patient populations.[12]

In an attempt to follow this line of investigation we performed an individual patient data (IPD) meta-analysis[9, 13], with data from two separate studies, in order to quantify the added diagnostic value of BNP for the diagnosis of heart failure in a population relevant to GPs and validated our findings in an independent primary care patient population.

Methods

Selection of studies

For the derivation dataset individual patient data were combined form two existing datasets, from Hillingdon, West London (n=127)[14] and Rotterdam (n=149)[15]. These datasets were chosen because of the comparable base population, health system and study design. In brief: GPs were invited to refer all suspected cases of new heart failure to a rapid-access outpatient clinic. The standardized diagnostic work-up included history taking and physical examination, electrocardiography (ECG), chest radiography, and transthoracic echocardiography. In addition, blood samples were taken.

For the external validation we used the UK Natriuretic Peptide (UKNP) study, representing 306 patients referred by their GP to a rapid access heart failure clinic.[16]

Patients

We defined the clinical domain as patients suspected of heart failure by the GP. Patients referred to the emergency department are in need of urgent hospital

based care and were therefore excluded; furthermore patients with established heart failure were excluded.

Case definition

In all three studies the diagnosis of heart failure was based on the recommendations from the European Society of Cardiology heart failure guidelines. According to these guidelines, two criteria have to be fulfilled (1) symptoms of heart failure (at rest or during exercise) and (2) objective evidence of cardiac dysfunction (at rest); while in cases where the diagnosis remains in doubt the response to treatment directed towards heart failure can be taken into account.[17] Both derivation studies used a panel of experts with access to all diagnostic data (with the exception of the BNP data) to assess the final diagnosis of heart failure[18]. An underlying abnormality of cardiac structure or function was necessary to confirm a case as heart failure, but echocardiographic abnormalities alone were not sufficient to diagnose heart failure; patients had to satisfy the full case definition.

BNP

The Hillingdon and Rotterdam studies both used the same method for BNP measurement: blood was drawn into a tube containing edetic acid and aprotinin. Plasma was stored at -80°C. Plasma concentration was measured by means of a commercially available radioimmunoassay kit (Peninsula Laboratories, Belmont, CA, USA). In the UKNP study BNP was assayed using a point-of-care fluorescence immunoassay (Biosite Diagnostics, Velizy, France). BNP values were log transformed because of skewness in the distribution, effectively changing the scale from interval to ratio.

Data analysis

Standard univariable analyses were performed. For the multivariable regression analyses we used logistic regression analyses with heart failure as the dependent variable and signs, symptoms and BNP levels (and other potential additional diagnostic tests, such as an electrocardiogram) as independent variables. In the case of missing values a 'complete cases only' procedure to treat missing values may result in biased estimates[19], so we used the multiple imputation method for reconstructing the incomplete data in order to obtain multiple complete data sets to produce more accurate estimates.

In the analyses we followed the natural hierarchy of a diagnostic work-up. First, a model was constructed from all readily available variables -that is gender, age, medical history and physical examination- the clinical model. The criterion for variables to be included in this clinical model was a p-value <0.10 in the univariable analyses.[20]

The second step was used to estimate the added value of BNP conditional on the clinical model; the same procedure was followed for the comparators chest X-ray and ECG. ECG was condensed into one binary item: any ECG abnormality (yes/no); chest X-ray into cardio-thoracic ratio >0.55 (yes/no). Subsequently the best combination of second step tests was sought.

The statistical significance of the added value of the respective variables in the second step ware computed by means of a likelihood ratio test.[21]

The third step was to estimate the accuracy of a model by its calibration (ie, agreement between observed and predicted probabilities of heart failure and formally tested with the le Cessie-van Houwelingen-Copas[22] (CHC) method to evaluate the overall goodness of fit) and discrimination (tested using the c-statistic from the logistic regression).

The fourth step was external validation. We tested the diagnostic model in an independent population, the UKNP study, and computed calibration and discrimination measures.

BNP was analyzed as a ratio-scale variable in the above analyses. We also took the approach of analyzing the BNP as a single test for all patients suspected of heart failure, thereby *not* taking the added value into account on top of weighing the readily available items. In that analysis, we used the cut-offs advised by the ESC: i.e. to rule out when the BNP level is <100 pg/ml and to 'rule in' (typically implying referral for additional investigations, notably echocardiography) when BNP is >400 pg/ml.[1] For comparison, we used the lower thresholds of <35 pg/ml to rule out heart failure and levels >100 pg/ml to rule in (implying referral for echocardiography) heart failure, proposed by the recent Dutch multidisciplinary guideline for heart failure 2010.[23]

All analyses were performed in R (version 2.10; The R Foundation for Statistical Computing, Vienna, Austria).

Results

Results from all measurements across the three contributing studies are summarized in table 1. The derivation dataset comprised of 276 patients suspected of new heart failure. The population under study was elderly with a mean over 70 years of age, female majority of 60% and more than half had a history of hypertension. All patients had one or more signs or symptoms compatible with heart failure, but only 85 (30.8%) patients were ultimately diagnosed with heart failure in the derivation datasets and 104 (34.0%) in the validation dataset. Table 2 shows the crude association of patient characteristics and test results with the presence of heart failure for the derivation datasets. Of note, the rate of missing values for all variables was <5%.

The multivariable clinical model resulting from all readily available clinical items contained age, gender, history of coronary artery disease (myocardial infarction, percutaneous coronary intervention or coronary artery bypass graft), diabetes, orthopnea, elevated jugular venous pressure, pulmonary crackles, peripheral pitting edema and S3 gallop rhythm. The clinical model, combining these 9 items, had a c-statistic of 0.79.

In the next hierarchic level we evaluated additional diagnostic tests available to the GP –namely ECG, chest X-ray and BNP. The univariable analyses (ie, these tests were considered as solitary tests) revealed all three tests to be statistically significant predictors of the presence of heart failure (table 2). Adding the cardiac thoracic ratio >0.55 to the clinical model increased the c-statistic to 0.84 (LR test $p<0.0001$). Addition of the ECG (normal versus any abnormality) and the log-transformed BNP to the clinical model the c-statistic increased to 0.85 (LR test $p<0.0001$) and 0.92 (LR test $p<0.0001$), respectively (table 3).

The next hierarchical level is the independent added value of adding diagnostic tests to the clinical model plus BNP. Further addition of cardio-thoracic ratio or ECG did not contribute significantly (LR test $p>0.9$ for both additions).

All models were found to have the a CHC p-value >0.5, indicating calibration was satisfactory.

Applying the clinical model plus BNP in the validation dataset showed that the model performed well in another comparable primary care population suspected of heart failure (ie, c-statistic = 0.91 and CHC p-value = 0.65).

Table 1. Patient characteristics and test results

	Derivation		Validation
	Hillingdon[14]	Rotterdam[15]	UKNP[16]
Characteristics			
N	127 (46%)	149 (54%)	306
Heart failure present (%)	32.3	29.5	34.0
Heart failure present (N)	41	44	104
Female (%)	52.8	60.4	57.5
Age (mean, years)	66.8	76.6	71.7
Medical history (%)			
diabetes	4.8	11.4	19.0
MI, PCI or CABG	7.1	14.1	16.7
COPD	10.2	32.9	19.0
hypertension	46.8	51.7	54.9
Symptoms (%)			
orthopnea	22.0	20.8	na
exertional dyspnea	89.0	75.2	94.8
dyspnea at rest	10.2	24.2	na
swollen ankles	51.2	51.0	64.4
coughing	17.3	28.9	na
Physical examination (%)			
BMI(kg/m^2, mean)	27.8	27.7	29.4
BP systolic (mmHg, mean)	153	161	145
BP diastolic (mmHg, mean)	90	90	79
wheezing	13.4	39.6	6.5
JVP elevated	17.1	12.8	17.0
pulmonary crackles	72.4	40.3	30.0
rhonchi	44.4	8.7	na
peripheral pitting edema	45.7	41.6	46.4
S3 gallop rhythm	4.7	2.7	12.4
systolic murmur	38.6	30.9	na
ECG (%)			
any ECG abnormality	45.7	36.2	53.3
Chest X–ray (%)			
cardio–thoracic ratio >0.55	50.8	20.1	44.2
log BNP (mean, log$_e$(pg/ml))	4.4	4.1	3.9

MI, PCI or CABG = myocardial infarction, percutaneous coronary intervention or coronary artery bypass grafting; COPD = Chronic Obstructive Pulmonary Disease; BMI = body mass index; BP = blood pressure; JVP = jugular venous pressure; BNP = B-type natriuretic peptide.

Table 2. Univariable predictors of heart failure in the derivation dataset

	heart failure present N (%)				OR	(95% CI)
	No		Yes			
	N=191 (69.2)		N=85 (30.8)			
Female	122	63.9	35	41.2	0.40	(0.23 - 0.67)
Age (year)*	71.1	(10.8)	74.2	(10.0)	^1.35	(1.03 - 1.76)
Patient history						
Hypertension	94	49.2	42	49.4	1.00	(0.60 - 1.67)
Diabetes	12	6.3	11	12.9	2.16	(0.91 - 5.12)
MI	12	6.3	13	15.3	2.69	(1.17 - 6.18)
PCI	3	1.6	1	1.2	0.75	(0.08 - 7.28)
CABG	3	1.6	5	5.9	3.92	(0.91 - 16.8)
MI, PCI or CABG	15	7.9	15	17.6	2.51	(1.17 - 5.40)
COPD	43	22.5	19	22.4	0.99	(0.54 - 1.83)
Symptoms						
Dyspnea on exertion	152	79.6	73	85.9	1.56	(0.77 - 3.16)
Orthopnea	30	15.7	29	34.1	2.78	(1.53 - 5.03)
Coughing or wheezing	73	38.2	28	32.9	0.79	(0.46 - 1.36)
Swollen ankles	92	48.2	49	57.6	1.46	(0.87 - 2.45)
Physical exam						
BMI (kg/m^2)*	27.8	(4.7)	27.7	(4.7)	1.00	(0.94 - 1.05)
BP systolic (mmHg)*	158	(23)	155	(27)	^0.95	(0.85 - 1.06)
BP diastolic (mmHg)*	90	(11)	90	(13)	^1.02	(0.81 - 1.27)
JVP elevated (yes/no)	19	9.9	22	25.9	3.88	(1.77 - 8.49)
Pulmonary crackles	73	38.2	58	68.2	3.27	(1.73 - 6.16)
Peripheral pitting edema	73	38.2	47	55.3	2.00	(1.19 - 3.36)
S3 gallop rhythm	2	1.0	8	9.4	9.82	(2.04 - 47.3)
Systolic murmur	63	33.0	32	37.6	1.23	(0.72 - 2.09)
Cardiothorax ratio[#]>0.55	41	21.5	49	57.6	5.19	(2.97 - 9.08)
Any ECG abnormality	48	25.1	64	75.3	9.08	(5.03 - 16.4)
log BNP (log_e(pg/ml))*	3.8	(0.5)	5.1	(0.8)	11.1	(5.73 - 21.4)

[#] on chest X–ray; * mean (S.D.); OR denotes Odds Ratio; ^ denotes OR per 10 units

Finally the results of applying the 'rule in' and 'rule out' cut-off levels from the ESC guideline and the lower alternative are presented in table 4. The ESC suggested decision cut points resulted in 46% of patients with a BNP level <100 pg/ml of whom 11% were false negative, 31% were in the intermediate 'grey' zone and 24% had a BNP level >400 pg/ml of whom 36% did not have heart failure . The alternative rule resulted in 15% of patients with a BNP level <35 pg/ml of whom 3.5% were false negative, also 31% in the 'grey' zone and 54% in the >100 pg/ml zone (requiring referral for echocardiography) of whom 50% did not have heart failure.

Table 3. Discriminative capacity of different models predicting the presence of heart failure

	c–statistic
Clinical model, combining 9 items readily available in the GP office	0.79
Clinical model plus chest X–ray cardio thoracic ratio >0.55	0.84
Clinical model plus ECG normal versus ECG abnormal	0.85
Clinical model plus BNP	0.92
Clinical model plus BNP in the validation dataset	0.91

Table 4. Presence of heart failure according to different cut-off levels for BNP to 'rule in' (usually implying referral for echocardiography) and 'rule out' heart failure

	heart failure present		
	yes (%)	no (%)	all (%)
All	32	68	
ESC guideline cut-off levels			
BNP < 100	5	41	46
BNP 100-400	12	18	31
BNP > 400	15	8	24
Alternative cut-off levels			
BNP < 35	1	14	15
BNP 35-100	5	26	31
BNP > 100	27	27	54

levels in pg/ml

Discussion

In this individual patient data meta-analysis BNP had the highest diagnostic value in addition to signs and symptoms: the knowledge of BNP significantly increased the c-statistic from 0.79 to 0.92. Validation in a third cohort confirmed the added value of BNP.

BNP can be considered a robust additional marker which is illustrated by the fact that its diagnostic power was retained, in spite of the varying case-definitions in the individual studies. This is also exemplified by the fact that although in the validation study another method to measure plasma BNP was used than in the derivation studies, validation confirmed the diagnostic value of BNP.

Certain limitations of our study should be discussed. Heart failure has no clearcut objective diagnostic reference standard, hence our studies all used an expert panel to arrive at the final diagnosis. The Hillingdon[14] study used a panel of three cardiologists; the Rotterdam[15] study a panel consisting of a general practitioner, cardiologist and internist; the UKNP[16] validation study used the judgment of the treating cardiologist. All final diagnoses were made without the knowledge of the BNP results, precluding incorporation bias. The influence of different panels on the final diagnosis is unknown, but the reproducibility was checked in a sample of a similar study and found to be excellent (Cohen's kappa = 1.00).[24]

The generalisability of our findings should also be discussed. The clinical domain of the study, patients suspected of heart failure by the GP, seems adequately captured as illustrated by patient characteristics (e.g. mean age of 74 years) and the prior probability of heart failure of 32.5%. One of the strengths of our study is that we only included patients suspected of heart failure, in contrast to several earlier studies that (also) included patients with established heart failure, e.g.[25] or non-diseased persons, e.g.[26], in whom no diagnostic dilemma exists. The latter leads to an overestimation of the diagnostic accuracy of BNP, while our findings are applicable to daily clinical practice where BNP is typically determined in patients suspected of heart failure.

The use of the c-statistic has been questioned lately.[27] The interpretation of the c-statistic is the probability of correctly identifying one pair of patients, one with the disease, the other without. Thus the range of the c-statistic is between 0.5 and 1.[28] It cannot easily be translated in daily practice. The c-statistic summarizes diagnostic discrimination across the whole range of disease probabilities, whereas certain tests may be of greatest value in certain probability areas, for example around the threshold where the disease is considered absent and treatment will be withheld. Such thresholds, however, are unidentified for heart failure. For patients suspected of heart failure the ESC guidelines for the diagnosis and

treatment of acute and chronic heart failure 2008 propose a threshold of <100 pg/ml for unlikely and >400 pg/ml for likely chronic heart failure.[1] One can argue that in general practice more emphasis should be given to a high negative predicted value and thus other thresholds should apply. For that purpose we also used BNP levels below 35 pg/ml to rule out heart failure and levels exceeding 100 pg/ml to 'rule in' heart failure as recommended in the recently published Dutch guideline.[29] Both scenarios led to 31% of patients in the grey zone and consequently 69% of patients with heart failure ruled in or out. In the ESC scenario 5% of all patients will have heart failure but will not be treated and 8% of all patients 'ruled in' (most of whom will be referred for echocardiography) will not have heart failure. In the 'low threshold' scenario 27% of 'ruled in' patients will not have heart failure and 1% of all patients with will not receive heart failure treatment although heart failure is present. In general practice the use of two thresholds as a single test for 'ruling in' or 'ruling out' heart failure will lead to better diagnosis, but still more than 40% of patients will require referral for additional investigations (or when not referred will receive unnecessary heart failure therapy) if the test is used in isolation of clinical findings. These findings again illustrate that BNP combined with clinical assessment can be very useful to rule out heart failure and aid the physician's diagnostic management.[30]

From our meta-analysis it can be concluded that the addition of BNP results in a significant improvement in the discrimination between those with and without heart failure and yields more adequate probability estimations of the presence of heart failure, as illustrated by the excellent calibration. The aim of our study was to assess the value of BNP in addition to signs and symptoms in suspected heart failure and not to provide the practicing clinician with a diagnostic rule. Some argue that in daily practice, BNP cut-off values should be adjusted according to age, gender or renal function.[31] Others advice using multivariable diagnostic algorithms for daily practice, using BNP levels (without cut-off points), and including age, gender or renal function to take into account the influence of these items on BNP levels.[32] This calls for the next step: to develop a diagnostic rule, obtained from the logistic regression model, offering the clinician an estimation of the probability of the presence of heart failure. This is likely to be more informative to the clinician and the patient than a BNP value above or below an age, gender and renal function adjusted cut-off level. Unfortunately the contributing studies were not designed to build such a diagnostic model; this would require an even larger study with standardized information on the whole range of potentially useful diagnostic items.

Based on the UKNP dataset (n=306), Mant et al. recently published and validated a decision tree to guide physicians in their referral for echocardiography[9], which was incorporated into the 2010 update of the National Institute for Clinical Excellence (NICE) clinical guideline "Chronic heart failure: the management of adults with chronic heart failure in primary and secondary care"[33]. Although the Mant et al. study was also limited in that standardized assessment in large numbers of suspected patients of clinical signs and symptoms was lacking, the study confirmed the potential value of (NT-pro)BNP. For primary care the NICE recommendations on diagnosing heart failure are in essence rules of referral for specialist assessment and Doppler echocardiography, taking into account diagnostic, prognostic and organizational information, such as the very limited availability of echocardiography for GPs in the UK. Based on that guideline the GP can rule out heart failure in patients without a previous myocardial infarction and a BNP level below 100 pg/ml, while all patients with either a previous myocardial infarction or BNP above 100 pg/ml should be referred.[33] In primary care, these patients, especially the very old with multiple co-morbidity, will not always be referred. Obviously, echocardiography performed in all patients with BNP levels above the cut-off point will reduce the number of false-positive diagnoses, but the applicability of such a strategy critically depends on the availability of echocardiography services for GPs, related costs and patient burden.

The *additional* diagnostic value of BNP on top of readily available was addressed in a few studies in the primary care setting. Besides the three studies used in this meta analysis, a study by Landray et al. refers to this issue.[34] Notwithstanding the limited power of the Landray study, BNP provided statistically significant additional diagnostic value. Of note, the 'disease' being identified in this study was echocardiographic left ventricular systolic dysfunction as opposed to the clinical syndrome of heart failure, the object of the present study.

We conclude that BNP has considerable diagnostic value *in addition* to signs and symptoms in patients suspected of heart failure in primary care. Nonetheless the simple method of applying one BNP test with one threshold for ruling in and one threshold for ruling out chronic heart failure for each patient suspected of heart failure will lead to insufficient proportions of patients correctly identified. Inclusion of available diagnostic information from history taking and physical examination together with the BNP level in one multivariable clinical prediction rule will lead to improved diagnostic accuracy.

References

1. 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.
2. Hunt SA, Abraham WTM, Chin MHM, Feldman AMM, Francis GS, Ganiats TGM, Jessup M, Konstam MAM, Mancini DM, Michl K, Oates JAM, Rahko PS, Silver MA, Stevenson LW, Yancy CW, Task Force MEMB, Antman EM, Smith SCJ, Adams CDM, Anderson JLM, Faxon DPM, Fuster V, Halperin JL, Hiratzka LFM, Jacobs AKM, Nishimura RA, Ornato JPM, Page RLM, Riegel B. ACC/AHA 2005 Guideline Update for the Diagnosis and Management of Chronic Heart Failure in the Adult-Summary Article: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Update the 2001 Guidelines for the Evaluation and Management of Heart Failure). Circulation. 2005;112:1825-1852.
3. Fonseca C, Morais H, Mota T, Matias F, Costa C, Gouveia-Oliveira A, Ceia F. The diagnosis of heart failure in primary care: value of symptoms and signs. Eur J Heart Fail. 2004;6:795-800.
4. Packer M. Should B-Type Natriuretic Peptide Be Measured Routinely to Guide the Diagnosis and Management of Chronic Heart Failure? Circulation. 2003;108:2950-2953.
5. Wiese S, Breyer T, Dragu A, Wakili R, Burkard T, Schmidt-Schweda S, Fuchtbauer EM, Dohrmann U, Beyersdorf F, Radicke D, Holubarsch CJF. Gene Expression of Brain Natriuretic Peptide in Isolated Atrial and Ventricular Human Myocardium: Influence of Angiotensin II and Diastolic Fiber Length. Circulation. 2000;102:3074-3079.
6. de Lemos JA, McGuire DK, Drazner MH. B-type natriuretic peptide in cardiovascular disease. Lancet. 2003;362:316-322.
7. 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.
8. Hill SA, Balion CM, Santaguida P, McQueen MJ, Ismaila 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. Clinical Biochemistry. 2008;41:240-249.

9. Mant J, Doust J, Roalfe A, Barton P, Cowie MR, Glasziou P, Mant D, McManus RJ, Holder R, Deeks J, Fletcher K, Qume M, Sohanpal S, Sanders S, Hobbs FD. 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.
10. Moons KGM, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. Clin Chem. 2004;50:473-476.
11. Knottnerus JA, Leffers P. The influence of referral patterns on the characteristics of diagnostic tests. Journal of Clinical Epidemiology. 1992;45:1143-1154.
12. Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derksen-Lubsen G, Grobbee DE, Moons KGM. External validation is necessary in prediction research: A clinical example. Journal of Clinical Epidemiology. 2003;56:826-832.
13. Lyman G, Kuderer N. The strengths and limitations of meta-analyses based on aggregate data. BMC Medical Research Methodology. 2005;5:14.
14. Cowie MR, Struthers AD, Wood DA, Coats AJS, 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.
15. Cost B. *Heart failure in the elderly. Thesis*. Rotterdam, Erasmus University; 2000.
16. 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.
17. The Task Force on Heart Failure of the European Society of C. Guidelines for the diagnosis of heart failure. European Heart Journal. 1995;16:741-751.
18. Weller SC, Mann NC. Assessing Rater Performance without a "Gold Standard" Using Consensus Theory. Medical Decision Making. 1997;17:71-79.
19. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. American Journal of Epidemiology. 1995;142:1255-1264.
20. 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. Journal of Clinical Epidemiology. 2004;57:1262-1270.
21. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine. 1996;15:361-387.

22. Cessie Sl, Houwelingen JCv. A Goodness-of-Fit Test for Binary Regression Models, Based on Smoothing Methods. Biometrics. 1991;47:1267-1282.
23. Hoes AW, Voors AA, Rutten FH, Van Lieshout J, Janssen PGH, Walma EP. Multidisciplinaire richtlijn hartfalen. Huisarts en Wetenschap. 2010;53:368-389 (in Dutch).
24. Fox KF, Cowie MR, Wood DA, Coats AJ, Gibbs JS, Underwood SR, Turner RM, Poole-Wilson PA, Davies SW, Sutton GC. Coronary artery disease as the cause of incident heart failure in the population. European Heart Journal. 2001;22:228-236.
25. Lainchbury JG, Campbell E, Frampton CM, Yandle TG, Nicholls MG, Richards AM. Brain natriuretic peptide and n-terminal brain natriuretic peptide in the diagnosis of heart failure in patients with acute shortness of breath. Journal of the American College of Cardiology. 2003;42:728-735.
26. McDonagh TA, Robb SD, Murdoch DR, Morton JJ, Ford I, Morrison CE, Tunstall-Pedoe H, McMurray JJ, Dargie HJ. Biochemical detection of left-ventricular systolic dysfunction. Lancet. 1998;351:9-13.
27. Cook NR. Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction. Circulation. 2007;115:928-935.
28. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29-36.
29. Hoes A, Walma E, Rutten F, Twickler T, Rohling R, Jansen R, Jaarsma T, Feenstra T, Bouvy M, Buskens E, De Graaff M, De Kok I, Dijkgraaf R, Duin M, Fischer E, Flikweert S, Hammelburg R, Honig A, Hulzebos E, In den Bosch H, Janssen-Boyne J, Koers H, Kortrijk M, Ninaber P, Poot E, Post P, Rosenbrand C, Schiffer A, Van Dijk J, Van Dijk P, Van Erp J, Van Erven L, Van Leen M, Van Lieshout J, Van Veldhuisen D, Weerts M, Voors A. Multidisciplinaire richtlijn Hartfalen 2010. (in Dutch). Available at: http://www.cbo.nl/Downloads/1081/rl_hartfalen_2010.pdf. Accessed 16 Nov 2010.
30. Hobbs FDR, Doust J, Mant J, Cowie MR. Heart failure: Diagnosis of heart failure in primary care. Heart. 2010;96:1773-1777.
31. Raymond I, Groenning BA, Hildebrandt PR, Nilsson JC, Baumann M, Trawinski J, Pedersen F. The influence of age, sex and other variables on the plasma level of N-terminal pro brain natriuretic peptide in a large sample of the general population. Heart. 2003;89:745-751.
32. Schou M, Alehagen U, Goetze JP, Gustafsson F, Dahlstrom U. Effect of estimated glomerular filtration rate on plasma concentrations of B-type natriuretic peptides measured with multiple immunoassays in elderly individuals. Heart. 2009;95:1514-1519.

33. National Institute for Clinical Excellence. Chronic heart failure: the management of adults with chronic heart failure in primary and secondary care (partial update). (Clinical guideline 108) 2010. *http://www.nice.org.uk/CG108*. Available at: http://www.nice.org.uk/CG108. Accessed 2/2/2011.
34. 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.

Chapter 3

The diagnostic value of physical examination and additional testing in primary care patients with suspected heart failure

based on:

Circulation. 2011 Dec 20;124(25):2865-73. Epub 2011 Nov 21.

The diagnostic value of physical examination and additional testing in primary care patients with suspected heart failure.

Kelder JC, Cramer MJ, van Wijngaarden J, van Tooren R, Mosterd A, Moons KG, Lammers JW, Cowie MR, Grobbee DE, Hoes AW.

Abstract

Background

Early diagnosis of non-acute heart failure is crucial since 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. 721 consecutive patients suspected of new onset heart failure underwent standardized diagnostic work-up including chest X-ray, spirometry, ECG, 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. 207 (28.7%) patients 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

This study estimated the quantitative diagnostic contribution of elements of the history and physical examination in the diagnosis of heart failure in primary care outpatients, and 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.

Introduction

Early diagnosis of heart failure is crucial since 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.

The diagnostic value of patient history and physical examination has been questioned,[4, 5] albeit that 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 non-diseased subjects.[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 BNP indeed shows excellent diagnostic accuracy this has mainly been studied without regard to other test results, including signs and symptoms.[9, 10] Consequently, the relevance of BNP measurements in patients suspected of 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 non-acute 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 one-

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, since these patients require immediate therapeutic intervention. Also, patients with known, established heart failure were 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 (cardiac-thoracic ratio and pulmonary edema as reported by the local specialist); 5) spirometry (forced expiration volume in 1 second (FEV1), peak flow and vital capacity); 6) standard laboratory assessments (including hemoglobin, electrolytes, C-reactive protein (CRP), 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° Celsius. 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 semi-quantitatively. 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 non-cardiac explanation of the signs and symptoms in the individual patient, notably pulmonary dysfunction. All images were stored digitally for off-line assessments.

Follow-up

Data on the clinical course during the subsequent 6 months were reported by the referring physician including the response to therapy. 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, since 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, 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 of 1 of 4 cardiologists, 1 of 4 pulmonologists and an outpatient heart failure clinic physician (JCK). 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 European Society of Cardiology (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 pre-specified 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 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. Non-contributing tests were manually (1 by 1) excluded using the likelihood ratio test[20] and resulted in model 1. Second stage models were conditional on model 1 to compute the added value of each of the supplementary

tests, i.e. 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 (NRI)[21], which measures reclassification in the right direction, for cases upwards and for non-cases downwards the probability of heart failure scale. The calibration was tested with the Hosmer-Lemeshow statistic.[22]

To adjust the coefficients and c-statistic for over-optimistic performance a shrinkage factor was obtained from a bootstrap analysis (1000 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 implemented in SAS®. We used 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 UKNP[25] and Hillingdon[26] 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 non-missing data on the (NT-pro)BNP measurement. For the Hillingdon study validation we used the subset of patients not assessed in the ER. The displaced apex beat was derived from the chest X-ray and defined as a cardio-thoracic ratio being larger than 0.60. The BNP was converted to NT-proBNP by means of the formula given by Alibay et al.: $_{10}$log(NT-proBNP)=1.1*$_{10}$log(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.

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), 51.7% had hypertension, 26.1% COPD and 15.3% diabetes. All patients had one or more complaints compatible with heart failure, notably dyspnea on exertion, ankle swelling, fatigue or orthopnea. 7.8% of patients 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.

Age, history of coronary artery disease, use of a loop diuretic and six items from 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 (i.e. excluding NT-proBNP) with sufficient added diagnostic value was “GGT >2 times upper limit of normal” (model 2a). Addition of chest X-ray (CT-ratio and pulmonary vascular redistribution, model 2b) to model 1 also contributed to the diagnosis, as were 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 in a NRI 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 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 NRI for these three categories is 17%, i.e. 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).

Table 1. Clinical characteristics of the study population

	All n=721		Heart Failure Yes n=207 (28.7%)		No n=514		p-value
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 less than one flight of stairs	438	(60.8%)	150	(72.5%)	288	(56%)	<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/m2)	29.4	±6	29.1	±6.1	29.5	±5.9	0.48
- blood pressure systolic (mmHg)	156.0	±26.7	153.8	±30	157	±25.3	0.14
- blood pressure diastolic (mmHg)	87.0	±13.2	86.5	±15	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%)	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%)	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

(Continued)

Table 1. Continued

	All n=721		Heart Failure Yes n=207 (28.7%)		No n=514		p-value
physical examination (continued)							
- heart murmer suggesting mitral regurgitation	77	(10.7%)	46	(22.2%)	31	(6%)	<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%)	<0.01
lab values							
- haemoglobin (mmol/l)	8.6	±0.9	8.5	±1	8.6	±0.8	0.64
- CRP >6 mg/l (vs <= 6mg/l)	333	(49.3%)	109	(56.2%)	224	(46.6%)	0.03
- eGFR MDRD (ml/min/1.73m2)	65	±16.3	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%)	25	(13.2%)	13	(3%)	<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 or 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
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%)	9	(1.8%)	20	(10%)	<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

(Continued)

Table 1. Continued

	All n=721		Heart Failure Yes n=207 (28.7%)		No n=514		p-value
-- '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	0.11
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%)	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%)	121	(58.5%)	369	(71.8%)	
-- mild	144	(20%)	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.

All multivariable statistical models were found to have adequate calibration, with p-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 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.

Figure 1. Flow chart

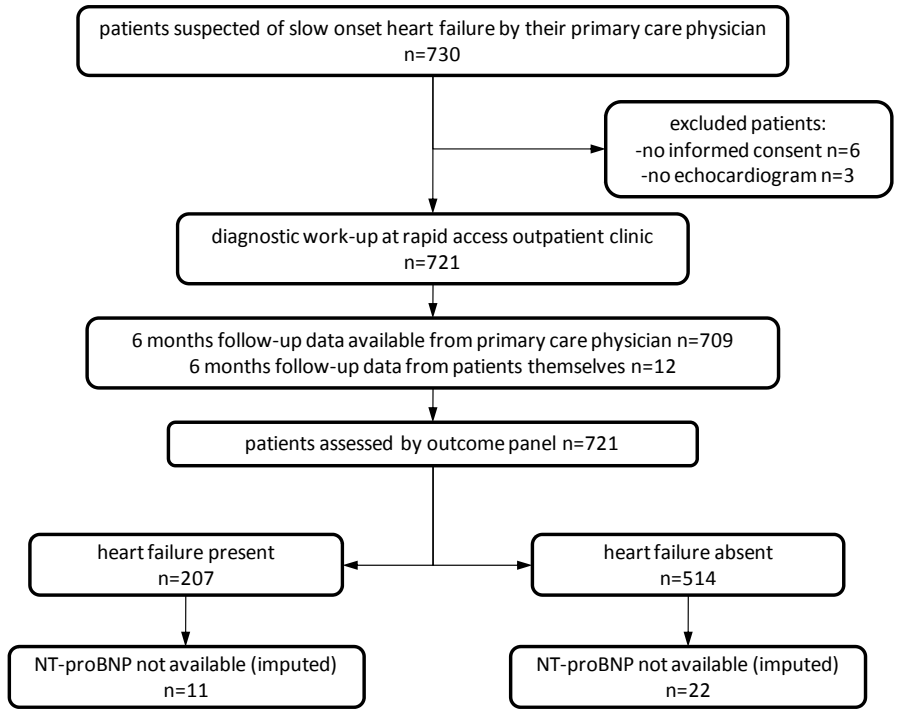


Table 2. Multivariable models predicting the presence of heart failure

	OR	[95% CI]	c-statistic	NRI	p-value LR test
model 1			0.83	reference	reference
age (years)	x 1.52	[1.25 - 1.86]	x = per 10 years		
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]			

(Continued)

Table 2. Continued

	OR	[95% CI]	c-statistic	NRI	p-value LR test
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.

Discussion

The findings among a large unselected group of patients suspected of new onset non-acute heart failure in primary care shows that a history taking and physical examination will markedly reduce the 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 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 mostly 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.

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 concords with a study by Cowie et al.[26] that reported a (statistically non-significant) multivariable odds ratio of 2.65. Badgett et al. reviewed 29 studies and concluded that chest X-ray cannot adequately exclude or confirm left ventricular dysfunction in clinical settings by itself.[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 carrying out 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 in similar patients and 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-pro)BNP were, however, performed in a hospital emergency room and not in an outpatient or primary care setting. It should be emphasised that univariable diagnostic characteristics such as sensitivity and specificity cannot be routinely generalised from one setting to another.[33]

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-pro)BNP are available, further facilitating the application of the diagnostic rule. As 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-pro)BNP measurement not be available. We believe our study population is representative of, and our study results generalisable to, the large population of patients suspected of heart failure, i.e. a population characterized by a mean age of around 70 years, predominantly white and female majority, often with considerable co-morbidity and on 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 it's 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 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 one diagnosis.[14] The outcome panel used all available information, excluding the NT-proBNP values. Incorporation bias cannot be excluded for other variables.

Finally, we analysed 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 two 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 less than 10% (13 points) is considered to rule out heart failure and a probability of more than 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 were 12.5% (12/96) and 4.7% (11/233), respectively.

In conclusion, the findings obtained in this study conducted in a large unselected group of patients suspected of 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 these, may be useful to quantify the probability of heart failure in patients with new symptoms suggestive of heart failure.

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

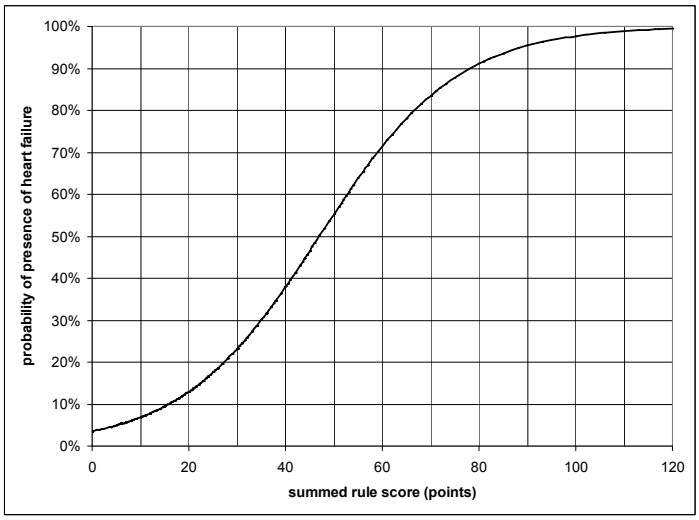
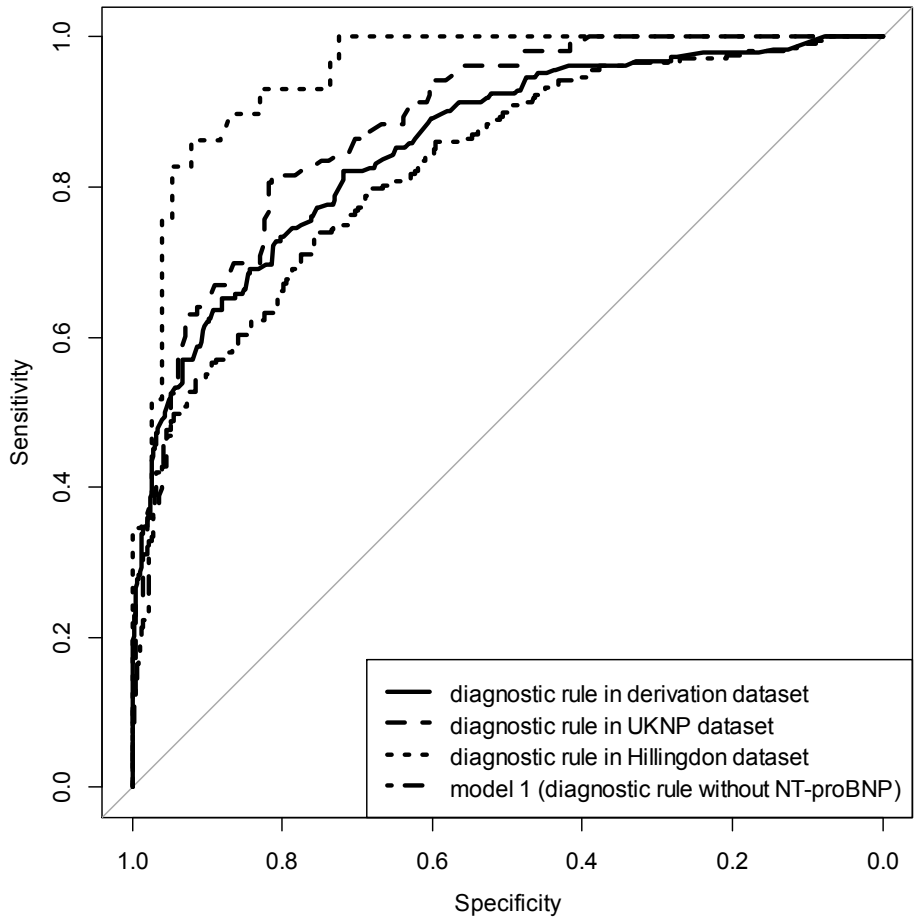


Table 3. Diagnostic Rule

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 over 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

Figure 3. Receiver operating characteristic curves



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.

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 the 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

Table 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.

Contributing Hospitals

Department of Cardiology, Deventer Hospital, Deventer
Department of Cardiology, St. Antonius Hospital, Nieuwegein and Utrecht
Department of Cardiology, University Medical Center, Utrecht
Department of Cardiology, Meander Medical Center, Amersfoort
Department of Cardiology, Isala Klinieken, Zwolle
Department of Cardiology, Catharina Hospital, Eindhoven
Department of Cardiology, Onze Lieve Vrouwe Gasthuis, Amsterdam

Ethical Approval

This study was approved by the medical ethical committee of the St Antonius Hospital, Nieuwegein, The Netherlands.

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. British Journal of General Practice. 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, Ismaila 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. Clinical Biochemistry. 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. European Heart Journal. 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. European Heart Journal. 1999;20:421-428.
16. Moons KGM, Grobbee DE. When should we remain blind and when should our eyes remain open in diagnostic studies? Journal of Clinical Epidemiology. 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). European Heart Journal. 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. Journal of Clinical Epidemiology. 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. Statistics in Medicine. 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. Statistics in Medicine. 1996;15:361-387.
23. Steyerberg EW. *Clinical Prediction Models*. New York: Springer; 2009; Page.
24. Donders AR, van der Heijden GJ, Stijnen T, Moons KGM. Review: a gentle introduction to imputation of missing values. Journal of Clinical Epidemiology. 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 AJS, 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. Biomedecine & Pharmacotherapy. 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. Journal of Clinical Epidemiology. 1992;45:1143-1154.

Chapter 4

Differentiating between reduced and preserved ejection fraction in patients suspected of heart failure in primary care with basic skills and simple tools

Kelder JC, Cramer MJ, Plokker HW-T, Mosterd A, Grobbee DE, Hoes AW.

(Submitted)

Abstract

Background

Distinguishing between heart failure with reduced ejection fraction (HFrEF) and heart failure with preserved left ventricular (LV) ejection fraction (DHF) in patients suspected of heart failure is crucial since different therapies are required. In settings without easy access to state-of-the-art echocardiography the key tools to categorize these patients are symptoms, signs and simple laboratory tests. The purpose of this diagnostic study was to identify diagnostic tools readily available to the primary care physician to help distinguish between both types of HF and no HF.

Methods and Results

Patients suspected of heart failure by their primary care physician were referred to a special diagnostic outpatient clinic for a full diagnostic work-up. The final diagnosis of heart failure was established by an expert panel and based on all available data (except NT-proBNP). When heart failure was judged to be present, the HFrEF or DHF phenotype was based on echocardiographic measurements. Multinomial logistic regression was used to compute odds ratios (ORs). Of all 717 patients, 66 (9.2%) had DHF, 140 (19.5%) had HFrEF and 511 (71.3%) no HF. Female gender is indicative for DHF, male for HFrEF (OR=0.37 [0.19-0.71]), whereas age had an OR of 1.00. Other statistically significant tools for the distinction DHF, HFrEF and no HF (jugular vein distension, cardio-thoracic ratio >0.50, an abnormal ECG and NT-proBNP) were all indicative of HFrEF more than DHF (ORs resp. 3.62, 4.49, 6.67, 1.63 (log(pg/ml))).

Conclusions

In this study no clear distinction could be made with readily available diagnostic tools. When the data point to severe symptoms it is probably HFrEF, whereas DHF is more likely when a sufficient number of mild signs and symptoms and laboratory tests compatible with heart failure are present. Notably age was not helpful in this setting.

Introduction

Chronic heart failure is a clinical syndrome that is currently often subdivided into two phenotypes: heart failure with reduced ejection fraction (HFrEF) and heart failure with preserved left ventricular (LV) ejection fraction (DHF). Following the European Society of Cardiology guideline patients with DHF have symptoms and/or signs of heart failure and a preserved left ventricular ejection fraction (EF >40–50%).[1] Ongoing scientific debate has proponents of the differences whereas others focus on the overlap between HFrEF and DHF[2]. Nevertheless for HFrEF potent evidence based treatments are available, for DHF not (yet)[1, 3-5]. It is therefore important to distinguish the two phenotypes in an early phase.

It has been shown that it is a challenge to diagnose DHF in a hospital setting with sophisticated diagnostic tools such as state-of-the-art echocardiographic measurements[6, 7], let alone in a setting such as primary care or out-patients clinics where these diagnostic tools are not (routinely and readily) available. In these settings, where most patients with suspected slow-onset heart failure are first assessed, attending physicians will primarily apply readily available diagnostic tools, such as signs and symptoms and ECG or (NT-pro)BNP measurements. Studies assessing whether these contribute to the distinction between HFrEF and DHF are, however, not available. We performed such a diagnostic study in patients suspected of non-acute heart failure in primary care.

Methods

Patients

For this diagnostic study we used the cohort of the Utrecht Heart Failure Organisation – Initial Assessment (UHFO-IA) study, a diagnostic study with the outcome heart failure yes or no.[8] In short the UHFO-IA study included 721 consecutive patients suspected of heart failure by their primary care physician. Patients were sent to one of eight heart failure outpatient diagnostic facilities where a standardized diagnostic work-up was performed, including ECG, chest X-ray, N-terminal pro-B-type Natriuretic Peptide (NT-proBNP) as well as echo-cardiography. Also information on the clinical history of the 6 months following the index visit was obtained. The final diagnosis of heart failure was established by an expert panel and based on all available data (except NT-proBNP). When heart failure was judged to be present, the HFrEF or DHF phenotype was subsequently ascertained based on echocardiographic measurements (see below). Four patients did not have an echocardiogram of sufficient quality and were excluded from the analysis. Thus, the study population consisted of 717 patients.

Echocardiography

Echocardiographic examinations (M-mode, 2D and Doppler-flow) were obtained according to the American Society of Echocardiography guidelines.[9] Left ventricular ejection fraction was assessed semi-quantitatively. No generally accepted criteria exist for the distinction between DHF and HFrEF.[10] We categorized the diastolic function as "normal", "impaired relaxation" or "restrictive filling" by a combination of information on left ventricular wall thickness, transmitral and pulmonary vein flow patterns, and left atrial volume, according to international guidelines[11], with the addition of mitral inflow pattern during Valsalva maneuver or tissue Doppler derived E' when available. All images were stored digitally for off-line assessments. Finally, potential non-cardiac causes of the signs and symptoms were identified, notably through standardized pulmonary function tests.

Data Analysis

To evaluate the diagnostic value of each variable, we used multinomial logistic regression, an extension of the binomial logistic regression in which three outcome categories were included: no heart failure (No-HF), DHF and HFrEF.[12] For each variable in the multinomial logistic regression the odds ratio (OR) for DHF versus No-HF and the OR for HFrEF versus No-HF were computed with corresponding 95% confidence intervals. We also calculated the ORs for the HFrEF versus DHF comparison. In order to enable the comparison of the magnitude of the ORs of dichotomous variables with the continuous variables, we rescaled all ORs to one unit: two standard deviations.[13] Furthermore, in case an OR was less than 1 we inversed (1/x) the OR. We choose to perform these calculations for the HFrEF versus DHF comparison only, since we have already shown how a reliable estimation of the probability of heart failure (either HFrEF or DHF), can be made in this population[8] and the primary aim in the present study was to distinguish between HFrEF and DHF. Multivariable multinomial logistic regression analysis was performed to assess the diagnostic value of the items, independent from age and gender. All analyses were performed with R version 2.13 (http://www.r-project.org).

Results

Of all 717 patients 66 (9.2%) had heart failure with preserved LVEF and 140 (19.5%) had heart failure with reduced LVEF whereas 511 (71.3%) patients had no heart failure. Age distinguished DHF from no HF to the same extent as HFrEF from no HF (OR 1.06 per year), consequently the OR of age for DHF versus HFrEF was 1.00. Female gender pointed at DHF rather than to HFrEF (OR 1.57 respectively 0.58 when compared to no HF; the OR for HFrEF versus DHF was 0.37). In addition, a history of hypertension and systolic blood pressure are more

predictive of the presence of DHF than HFrEF (OR for HFrEF versus DHF of 0.50 respectively 0.97 per mmHg).

Variables pointing statistically significantly to HFrEF compared to both no HF as well as DHF are: jugular vein distension (OR 7.67 versus no HF, 3.62 versus DHF), abnormal ECG (OR 1/0.06=16.7 resp. 1/0.15=6.7), cardio-thoracic ratio (CTR) >0.50 on chest x-ray (OR 8.24 resp. 4.49), pleural effusion on chest X-ray (OR 9.88 resp. 4.56) and NT-proBNP (OR on a log(pg/ml) scale 2.60 resp. 1.63).

Figure 1 is the result of all ORs for the HFrEF versus DHF comparison rescaled to the same unit and inversed when less than 1. The normalized ORs for the top seven variables were: systolic blood pressure (2.5), gender (2.7), jugular vein distension (3.6), CTR >0.50 (4.5), pleural effusion (4.6), NT-proBNP (5.4 log(pg/ml)) and (ab)normal ECG (6.6). The results of the multivariable analysis of these seven variables plus age are depicted in table 2 in the original units. For the distinction of the diagnostic trichotomy, the multivariable analysis would value the systolic blood pressure and pleural effusion redundant because no statistically significant OR was reached.

Discussion

For the primary care physician attempting to distinguish between no heart failure, HFrEF and DHF in patients suspected of slow onset heart failure, the task is difficult without the help of state-of-the-art echocardiography.

In this study cohort the prevalence of HFrEF and DHF was 9.2% respectively 19.5%. Taking into account the basic diagnostic skills and tools readily available to primary care physicians, this study shows that in the initial diagnostic phase HFrEF is easier to diagnose than DHF since the highest diagnostic ORs are found in the HFrEF/no-HF division (table 1).

Previous studies demonstrated that patients with DHF are older, more often female and hypertensive compared to those with HFrEF[14, 15]. In our study advancing age is a risk factor for both types of heart failure, with odds ratios for both of 1.06 per year and does not discriminate between DHF and HFrEF (figure 1). In line with previous studies female gender does have an OR > 1 for DHF versus no HF (1.57) and an OR < 1 for HFrEF versus no HF (0.58), but the cause of the gender difference is not clear. Associations are found with gender related to the evolution to eccentric LV remodeling in hypertensive heart disease.[16] Moreover males and females differ with respect to expression, diagnosis and also treatment of ischemic heart disease[17]. Lastly the diagnostic criteria are contested and different cut-off of LV mass index for males and females are proposed.[18]

Hypertension is compatible with diastolic dysfunction via left ventricular hypertrophy often interacting with coronary artery disease.[19] The present study is compatible with this viewpoint when history of hypertension and systolic blood pressure are concerned, but LVH on the ECG is quite a strong risk factor for both types of heart failure, without the ability to discriminate between the two phenotypes. The interplay of time between underlying disease onset and first presentation with symptoms could be a major factor.[19] Furthermore the differential under-diagnosis of coronary artery disease in patients with DHF could also be at play.[20] The metabolic syndrome has a high prevalence in patients with DHF, although the pathophysiological basis is an unresolved issue.[21] In this study some features of the metabolic syndrome, history of hypertension, history of diabetes and BMI, were evaluated, all associated with DHF more than HFrEF, where only history of hypertension was statistically significant; multivariable no clear picture emerged.

Consistent with our study Iwanaga et al.[22] have shown that patients with HFrEF have higher BNP levels than patients with DHF, despite equivalent filling pressures. Because BNP is highly correlated with end-diastolic wall stress, the differential effects of pressure and dilatation could explain the diagnostic ability of NT-proBNP for heart failure (HFrEF as well as DHF), whereas for distinguishing between HFrEF and DHF, NT-proBNP yields a normalized OR of 4.49. This phenomenon could also explain the similar diagnostic ability of CTR >0.50. Left and right sided filling pressures are probably not discriminative since they were found to be interrelated in HFrEF[23] and very recently a similar relation was found in DHF[24], which concurs with our finding that signs and symptoms pointing to these phenomena are associated with a higher probability for both DHF and HFrEF.

Overall, most variables associated with heart failure are more strongly related to HFrEF than DHF, indicating that HFrEF probably has more pronounced left and right sided increased filling pressures than DHF in patients with new, slow onset heart failure, i.e. patients in an early stage of the disease. Kitzman et al. came to a similar conclusion when comparing three selected groups of patients, i.e. healthy volunteers, patients with typical systolic heart failure and patients with isolated diastolic heart failure with LVEF of at least 50%, and no evidence of significant coronary, valvular, or pulmonary disease[25]. Moreover the three univariable most discriminating variables (NT-proBNP, CTR >50 and normal ECG) appear to be multivariable independent. Nowadays, for the primary care physician confronted with a patient suspected of heart failure the trichotomy of no heart failure, HFrEF and DHF lies in the subtle interpretation of the readily available evidence. When the data point to severe and overt symptoms it is probably HFrEF, while DHF is

more likely when a sufficient number of mild signs and symptoms compatible with heart failure are present.

Limitations

As mentioned we accepted the division of the two phenotypes of heart failure for pragmatic reasons and thus deliberately bypassing the hypothesis stating that many patients with DHF were misdiagnosed as for instance Paulus et al. argue.[4] Our study is more in line with Lee et al. stating "heart failure with reduced left ventricular systolic function and heart failure with preserved left ventricular systolic function are partially distinct entities, with potentially different approaches to early detection and prevention".[26]

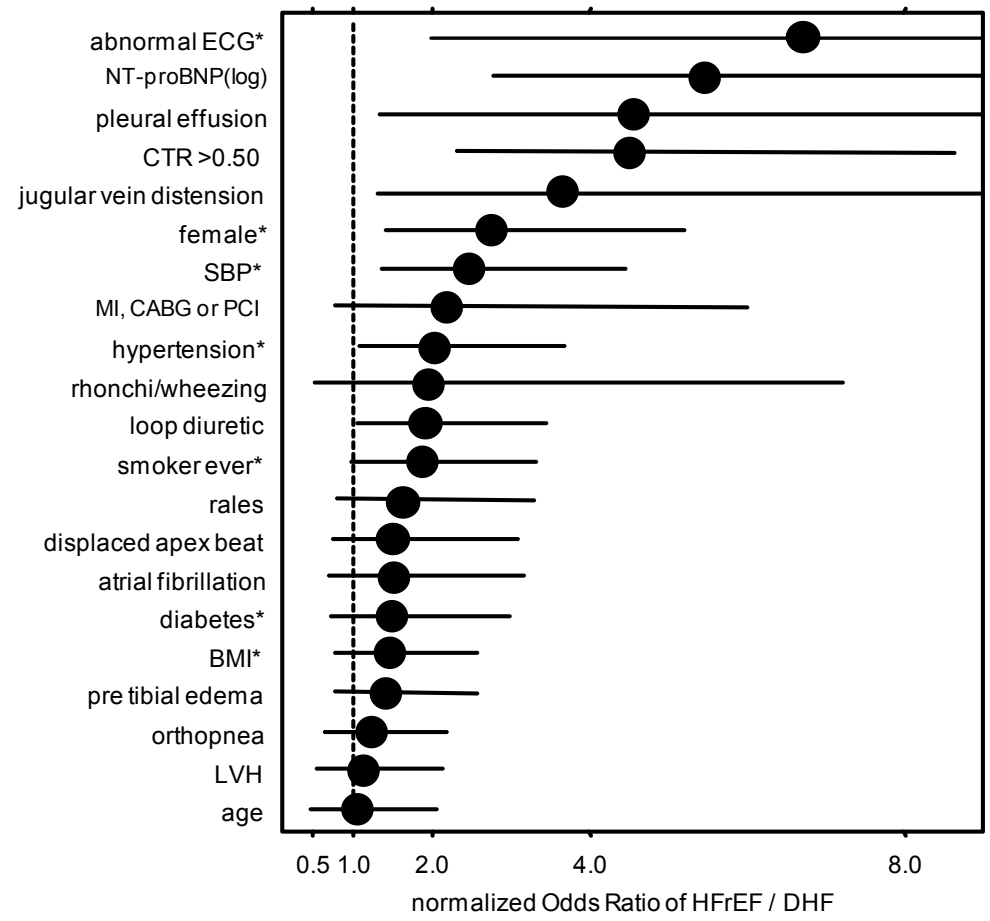
We did not measure E/E' systematically in all patients in the echocardiographic assessment in this cohort, therefore the measurement of the diastolic function of the LV was probably not consistent with the most recent standard.

This study was not designed to find a multivariable diagnostic rule to cover the trichotomy, for that more and larger studies are needed.

Conclusion

The distinction between heart failure and no heart failure in the initial assessment of patients suspected of new, slow onset heart failure can be made with sufficient confidence in a substantial proportion of suspected patients.[8] We conclude that in this study the distinction of no heart failure, HFrEF and DHF cannot be made clear with signs and symptoms and simple laboratory tests. However, the primary care physician confronted with a patient suspected of heart failure, the distinction lies in the subtle interpretation of the readily available evidence. When the data point to severe symptoms it is probably HFrEF, whereas DHF is more likely when a sufficient number of mild signs and symptoms compatible with heart failure are present.

Figure 1. Normalized Odds Ratios with 95% confidence intervals for the distinction HFrEF versus DHF (OR's marked * are inversed)



Caption for Figure 1 and Table 1: OR = odds ratio; DHF = heart failure preserved ejection fraction; HFrEF = heart failure reduced ejection fraction; no HF = no heart failure; MI = myocardial infarction; CABG = coronary artery bypass grafting surgery; PCI = percutaneous coronary intervention; BMI = body mass index. SBP = systolic blood pressure; LVH = left ventricular hypertrophy; ECG = electrocardiogram; CTR = cardio thoracic ratio; NT-proBNP = N-terminal pro-B-type natriuretic peptide;

Table 1. Univariable associations with three category outcome (no HF, DHF, HFrEF) (proportions and means)

	no HF (n=511)		DHF (n=66)		HFrEF (n=140)		p-value
female	340	66.5%	50	75.8%	75	53.6%	0.0029
age (year , mean (SD))	68.8	(12.1)	75.6	(9.1)	75.6	(9.8)	<0.0001
hypertension	256	50.1%	44	66.7%	70	50.0%	0.0382
diabetes	68	13.3%	16	24.2%	25	17.9%	0.0423
MI, CABG or PCI	22	4.3%	5	7.6%	21	15.0%	0.0001
smoker (ever)	311	60.9%	36	54.5%	96	68.6%	0.1090
orthopnea	141	27.6%	24	36.4%	57	40.7%	0.0077
BMI (kg/m2, mean (SD))	29.4	(5.9)	29.7	(6.2)	28.7	(5.7)	0.3416
SBP (mmHg, mean(SD))	157	(25)	162	(30)	150	(29)	0.0041
rales	43	8.4%	13	19.7%	40	28.6%	<0.0001
rhonchi / wheezing	41	8.0%	3	4.5%	12	8.6%	0.6661
displaced apex beat	19	3.7%	13	19.7%	38	27.1%	<0.0001
jugular vein distension	19	3.7%	5	7.6%	32	22.9%	<0.0001
pre tibial edema	111	21.7%	23	34.8%	60	42.9%	<0.0001
normal ECG	170	34.0%	11	17.2%	4	3.1%	<0.0001
atrial fibrillation	14	2.7%	11	16.7%	32	22.9%	<0.0001
LVH	27	5.3%	16	24.2%	36	25.7%	<0.0001
CTR >0.50	105	23.1%	22	35.5%	84	71.2%	<0.0001
pleural effusion	11	2.2%	3	4.5%	25	17.9%	<0.0001
NT-proBNP (log(pg/ml), mean(SD))	2.83	(1.28)	3.86	(1.77)	5.34	(1.71)	<0.0001

Table 1. Univariable associations with three category outcome (no HF, DHF, HFrEF) (odds ratios)

	OR [DHF/ no HF] (95% CI)			OR [HFrEF/ no HF] (95% CI)			OR [HFrEF/ DHF] (95%-ci)		
female	1.57	(0.87	2.84)	0.58	(0.40	0.85)	0.37	(0.19	0.71)
age (year , mean (SD))	1.06	(1.03	1.09)	1.06	(1.04	1.08)	1.00	(0.97	1.03)
hypertension	1.99	(1.16	3.42)	1.00	(0.69	1.45)	0.50	(0.27	0.92)
diabetes	2.09	(1.12	3.87)	1.42	(0.86	2.34)	0.68	(0.33	1.38)
MI, CABG or PCI	1.82	(0.67	4.99)	3.92	(2.09	7.37)	2.15	(0.78	6.01)
smoker (ever)	0.77	(0.46	1.29)	1.40	(0.94	2.09)	1.82	(1.00	3.32)
orthopnea	1.50	(0.88	2.57)	1.80	(1.22	2.66)	1.20	(0.66	2.20)
BMI (kg/m2, mean (SD))	1.01	(0.97	1.05)	0.98	(0.94	1.01)	0.97	(0.92	1.02)
SBP (mmHg, mean(SD))	1.01	(1.00	1.02)	0.99	(0.98	1.00)	0.98	(0.97	0.99)
rales	2.67	(1.35	5.28)	4.35	(2.69	7.05)	1.63	(0.80	3.31)
rhonchi / wheezing	0.55	(0.16	1.81)	1.08	(0.55	2.11)	1.98	(0.54	7.26)
displaced apex beat	6.35	(2.97	13.6)	9.65	(5.35	17.4)	1.52	(0.75	3.09)
jugular vein distension	2.12	(0.76	5.88)	7.67	(4.19	14.0)	3.62	(1.34	9.75)
pre tibial edema	1.93	(1.11	3.34)	2.70	(1.82	4.01)	1.40	(0.76	2.57)
normal ECG	0.40	(0.21	0.79)	0.06	(0.02	0.17)	0.15	(0.05	0.50)
atrial fibrillation	7.10	(3.07	16.4)	10.5	(5.43	20.4)	1.48	(0.69	3.16)
LVH	5.74	(2.90	11.4)	6.21	(3.61	10.7)	1.08	(0.55	2.13)
CTR >0.50	1.83	(1.04	3.22)	8.24	(5.23	13.0)	4.49	(2.33	8.65)
pleural effusion	2.16	(0.59	7.97)	9.88	(4.73	20.7)	4.56	(1.33	15.7)
NT-proBNP (log(pg/ml), mean(SD))	1.59	(1.34	1.90)	2.60	(2.21	3.04)	1.63	(1.34	1.97)

Table 2. Multivariable associations with three category outcome (no HF, DHF, HFrEF)

	OR [DHF / no HF]	(95% - ci)	OR [HFrEF / no HF]	(95% - ci)	OR [HFrEF / DHF]	(95% - ci)
age (year)	1.05	(1.01 - 1.08)	1.02	(0.99 - 1.05)	0.97	(0.94 - 1.01)
female gender	2.21	(1.03 - 4.72)	0.50	(0.27 - 0.92)	0.23	(0.10 - 0.54)
SBP (mmHg)	1.00	(0.99 - 1.02)	1.00	(0.99 - 1.01)	1.00	(0.98 - 1.01)
jugular vein distension	2.38	(0.69 - 8.14)	5.96	(2.24 - 15.9)	2.51	(0.75 - 8.41)
pleural effusion	1.35	(0.32 - 5.65)	2.33	(0.73 - 7.42)	1.73	(0.42 - 7.11)
CTR >0.50	1.01	(0.52 - 1.96)	2.84	(1.55 - 5.20)	2.80	(1.27 - 6.18)
normal ECG	0.65	(0.30 - 1.38)	0.26	(0.09 - 0.77)	0.41	(0.11 - 1.44)
NT-proBNP (log) (pg/ml)	1.52	(1.22 - 1.88)	2.01	(1.67 - 2.43)	1.33	(1.05 - 1.68)

OR = odds ratio; DHF = heart failure preserved ejection fraction; HFrEF = heart failure reduced ejection fraction; no-hf = no heart failure; SBP = systolic blood pressure; ECG = electrocardiogram; CTR = cardio thoracic ratio; NT-proBNP = N-terminal pro-B-type natriuretic peptide;

References

1. 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.
2. De Keulenaer GW, Brutsaert DL. The Heart Failure Spectrum: Time for a Phenotype-Oriented Approach. Circulation. 2009;119:3044-3046.
3. Hunt SA, Abraham WT, Chin MH, Feldman AM, Francis GS, Ganiats TG, Jessup M, Konstam MA, Mancini DM, Michl K, Oates JA, Rahko PS, Silver MA, Stevenson LW, Yancy CW. 2009 Focused Update Incorporated Into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Heart Failure in Adults: A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines Developed in Collaboration With the International Society for Heart and Lung Transplantation. Journal of the American College of Cardiology. 2009;53:e1-e90.
4. Paulus WJ, van Ballegoij JJM. Treatment of Heart Failure With Normal Ejection Fraction: An Inconvenient Truth! Journal of the American College of Cardiology. 2010;55:526-537.
5. Shah RV, Desai AS, Givertz MM. The Effect of Renin-Angiotensin System Inhibitors on Mortality and Heart Failure Hospitalization in Patients With Heart Failure and Preserved Ejection Fraction: A Systematic Review and Meta-Analysis. Journal of Cardiac Failure. 2010;16:260-267.
6. Zile MR, Brutsaert DL. New Concepts in Diastolic Dysfunction and Diastolic Heart Failure: Part I: Diagnosis, Prognosis, and Measurements of Diastolic Function. Circulation. 2002;105:1387-1393.
7. Thomas JT, Kelly RF, Thomas SJ, Stamos TD, Albasha K, Parrillo JE, Calvin JE. Utility of history, physical examination, electrocardiogram, and chest radiograph for differentiating normal from decreased systolic function in patients with heart failure. American Journal of Medicine. 2002;112:437-445.
8. 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.

9. 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.
10. Paulus WJ, Tschöpe C, Sanderson JE, Rusconi C, Flachskampf FA, Rademakers FE, Marino P, Smiseth OA, De Keulenaer G, Leite-Moreira AF, Borbély A, Édes I, Handoko ML, Heymans S, Pezzali N, Pieske B, Dickstein K, Fraser AG, Brutsaert DL. How to diagnose diastolic heart failure: a consensus statement on the diagnosis of heart failure with normal left ventricular ejection fraction by the Heart Failure and Echocardiography Associations of the European Society of Cardiology. European Heart Journal. 2007;28:2539-2550.
11. 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.
12. Biesheuvel CJ, Vergouwe Y, Steyerberg EW, Grobbee DE, Moons KGM. Polytomous logistic regression analysis could be applied more often in diagnostic research. Journal of Clinical Epidemiology. 2008;61:125-134.
13. Gelman A. Scaling regression inputs by dividing by two standard deviations. Statistics in Medicine. 2008;27:2865-2873.
14. Hogg K, Swedberg K, McMurray J. Heart failure with preserved left ventricular systolic function: epidemiology, clinical characteristics, and prognosis. Journal of the American College of Cardiology. 2004;43:317-327.
15. Owan TE, Hodge DO, Herges RM, Jacobsen SJ, Roger VL, Redfield MM. Trends in Prevalence and Outcome of Heart Failure with Preserved Ejection Fraction. N Engl J Med. 2006;355:251-259.
16. Rosen BD, Edvardsen T, Lai S, Castillo E, Pan L, Jerosch-Herold M, Sinha S, Kronmal R, Arnett D, Crouse JR, III, Heckbert SR, Bluemke DA, Lima JAC. Left Ventricular Concentric Remodeling Is Associated With Decreased Global and Regional Systolic Function: The Multi-Ethnic Study of Atherosclerosis. Circulation. 2005;112:984-991.
17. Shaw LJ, Bugiardini R, Merz CNB. Women and Ischemic Heart Disease: Evolving Knowledge. Journal of the American College of Cardiology. 2009;54:1561-1575.
18. Emery WT, Jadavji I, Choy JB, Lawrance RA. Investigating the European Society of Cardiology Diastology Guidelines in a practical scenario. European Journal of Echocardiography. 2008;9:685-691.

19. Vasan RS, Levy D. The Role of Hypertension in the Pathogenesis of Heart Failure: A Clinical Mechanistic Overview. Arch Intern Med. 1996;156:1789-1796.
20. Shah SJ, Gheorghiade M. Heart Failure With Preserved Ejection Fraction: Treat Now by Treating Comorbidities. JAMA. 2008;300:431-433.
21. Lam CSP, Donal E, Kraigher-Krainer E, Vasan RS. Epidemiology and clinical course of heart failure with preserved ejection fraction. Eur J Heart Fail. 2011;13:18-28.
22. Iwanaga Y, Nishi I, Furuichi S, Noguchi T, Sase K, Kihara Y, Goto Y, Nonogi H. B-Type Natriuretic Peptide Strongly Reflects Diastolic Wall Stress in Patients With Chronic Heart Failure: Comparison Between Systolic and Diastolic Heart Failure. J Am Coll Cardiol. 2006;47:742-748.
23. Drazner MH, Hamilton MA, Fonarow G, Creaser J, Flavell C, Warner Stevenson L. Relationship between right and left-sided filling pressures in 1000 patients with advanced heart failure. The Journal of Heart and Lung Transplantation. 1999;18:1126-1132.
24. Drazner MH, Prasad A, Ayers C, Markham DW, Hastings J, Bhella PS, Shibata S, Levine BD. The Relationship of Right- and Left-Sided Filling Pressures in Patients With Heart Failure and a Preserved Ejection Fraction. Circ Heart Fail. 2010;3:202-206.
25. Kitzman DW, Little WC, Brubaker PH, Anderson RT, Hundley WG, Marburger CT, Brosnihan B, Morgan TM, Stewart KP. Pathophysiological characterization of isolated diastolic heart failure in comparison to systolic heart failure. JAMA. 2002;288:2144-2150.
26. Lee DS, Gona P, Vasan RS, Larson MG, Benjamin EJ, Wang TJ, Tu JV, Levy D. Relation of Disease Pathogenesis and Risk Factors to Heart Failure With Preserved or Reduced Ejection Fraction: Insights From the Framingham Heart Study of the National Heart, Lung, and Blood Institute. Circulation. 2009;119:3070-3077.

Chapter 5

The furosemide diagnostic test in suspected slow-onset heart failure: popular but not useful

based on:

European Journal of Heart Failure (2011) 13, 513–517. doi:10.1093/eurjhf/hfr010

The furosemide diagnostic test in suspected slow-onset heart failure: popular but not useful

Kelder JC, Cramer MJ, Rutten FA, Plokker HW-T, Grobbee DE, Hoes AW.

Abstract

Aims

Early, slow-onset heart failure is difficult to diagnose from just signs and symptoms. The physician needs ancillary diagnostic tests. The 'loop-diuretic test' or 'furosemide test', characterized as weight loss and alleviation of symptoms after a short course of a loop-diuretic, could be a candidate. The furosemide test is not formally mentioned in the guidelines and no evidence could be found in the literature. We asked general practitioners (GPs) about their actual use of the furosemide test and studied the diagnostic accuracy in patients with suspected heart failure.

Methods and results

General practitioners completed a questionnaire about their use of the furosemide test. We then performed a diagnostic accuracy study among a representative and consecutive sample of patients suspected of new, slow-onset heart failure by the GP and who were referred to the rapid access heart failure diagnostic facility of one hospital. All patients underwent a standardized diagnostic work-up including echocardiography. The reference standard for the diagnosis of heart failure was the decision of an expert panel. Forty of the 54 GPs had actually used the furosemide test in the past year and 70% considered the test to be useful. Forty seven patients underwent the furosemide test and 12 (25%) were diagnosed with heart failure. None of the effects of the test (weight loss, alleviation of symptoms) was significantly associated with heart failure.

Conclusion

We cannot support the use of the furosemide test as an ancillary diagnostic test in patients suspected of new, slow-onset heart failure.

Introduction

Chronic heart failure is a common condition associated with high mortality and morbidity, and has a substantial impact on public health. Recognizing heart failure, especially in the early stages, is important since effective treatments are now available.[1, 2] Diagnosing early heart failure is difficult and imprecise when the physician only has symptoms and signs available for decision making, and the trend nowadays is for more patients to be diagnosed as outpatients.[3] In the initial assessment, symptoms and signs are often inadequate and general practitioners are aware of this problem.[4-6] A recent study from Sweden reported that patients in primary care are typically diagnosed solely on signs and symptoms (29% of cases) and in only 31% of cases was an echocardiogram included as part of the diagnostic work-up.[7]

The physician needs an ancillary diagnostic test to rule heart failure in or out. The guidelines recommend a range of ancillary diagnostic tests, notably electrocardiography (ECG), B-type natriuretic peptides (BNP or NT-proBNP), and chest X ray.[1,2] One of the oldest ancillary diagnostic tests was first described in 1971 in the Framingham Heart Study, where one of the major criteria for the diagnosis of heart failure was a ≥4.5 kg (10 lbs) weight loss within 5 days in response to loop diuretic treatment.[8] This 'loop-diuretic-test' or 'furosemide test' is not formally mentioned in the contemporary guidelines but the response to targeted heart failure treatment in case of volume overload typically diuretics is mentioned as an additional diagnostic tool in case of doubt.[1,9]

We set out to study the actual use of the loop diuretic test in daily practice by searching published reports and questioning a representative sample of primary care physicians. We also assessed the diagnostic value of the 'furosemide test' in a representative sample of patients with suspected heart failure in a primary care setting.

Methods

We assessed the opinions of a representative sample of GPs about the 'furosemide test' using an anonymous questionnaire which was handed out during a continuing medical education (CME) course on diabetes in November 2008. We asked the GPs to answer three questions: (i) How often do you apply the furosemide test? (ii) Which symptoms do you monitor and which criteria do you use for a 'positive test'? (iii) Are you convinced of the usefulness of the 'furosemide test' as a diagnostic test in heart failure?

To determine the diagnostic value of the 'furosemide test' we performed a diagnostic accuracy study among a representative, consecutive sample of patients

who were suspected of new, slow onset heart failure by the GP and were referred to the rapid access heart failure outpatient diagnostic facility of the St. Antonius Hospital in Nieuwegein, The Netherlands. All patients who consented to participate underwent a standardised diagnostic work-up programme as part of a larger study, the Utrecht Heart Failure Organisation - Initial Assessment (UHFO-IA) study. In short, all participants of the UHFO-IA study were asked about their medical history, signs and symptoms, underwent a 12-lead ECG, blood tests, including B-type natriuretic peptide measurements (NT-proBNP), chest X-ray, spirometry and echocardiography. The reference ('gold') standard for the diagnosis of heart failure was the decision of an expert panel consisting of a cardiologist, a pulmonologist and a general physician. They based their decision on the results of all diagnostic tests, including echocardiography, however, without the results of the B-type natriuretic peptide test or 'furosemide test'. The latter to prevent incorporation bias.[10-12] The criteria for heart failure specified in the European Society of Cardiology guidelines[13] were used for the final decision of the panel, which was made after 6 months of follow-up.

The GP was asked to provide an estimate of the probability of heart failure when the patient was referred to the rapid access facility.

Echocardiographic examinations (M-mode, 2D and Doppler-flow) were obtained in accordance with American Society of Echocardiography guidelines.[14] Left ventricular ejection fraction was assessed semi-quantitatively, where reduced systolic function was defined as ejection fraction <45-50%. Diastolic function was categorized as normal, impaired relaxation or restrictive filling by a combination of left ventricular wall thickness, transmitral and pulmonary flow patterns and left atrial volume.[15]

The inclusion criteria for the 'furosemide test' sub-study were: (i) referred by the general practitioner because of suspected heart failure, (ii) physically and mentally able to assess the effects of the 'furosemide test', and (iii) signed informed consent. Patients were excluded if they were already prescribed a loop or thiazide diuretic or if furosemide was deemed contra-indicated.

Before starting the furosemide test treatment the patient's weight was measured and their severity of fatigue, shortness of breath and oedema was assessed. Furthermore, they were given a diary to record their weight and symptoms. Patients were prescribed furosemide 40 or 80 mg once daily for 1 week at the discretion of the physician. Patients were invited to attend the outpatient clinic after this 1 week. Both patients and physician were blinded to the other results of the diagnostic assessment and thus did not know whether the patient had heart failure or not.

Table 1. Questionnaire on the current use in daily practice of the furosemide diagnostic treatment test in suspected heart failure; frequency of answers

1. How many times during the past year did you use the furosemide test treatment for suspected heart failure on a patient?	n=54
a. never:	26%
b. once or twice:	56%
c. three times or more:	19%
2. If you ever used the test, on which grounds did you consider the test positive for heart failure? (more than one answer could be given)	n=49
a. weight loss > 2 kg	39%
b. weight loss > 4 kg	10%
c. reduction of shortness of breath	57%
d. reduction of edema	14%
e. other	8%
3. What is your opinion about this statement: "The furosemide test treatment is useful when there is uncertainty about the diagnosis of heart failure"	n=54
a. fully disagree:	2%
b. disagree	6%
c. neutral	22%
d. agree	69%
e. fully agree	2%

The study was approved by the medical ethics committee of the St Antonius Hospital, Nieuwegein, The Netherlands.

Data analysis

The effects of the 'furosemide test' were measured from the absolute weight change (kg) in addition to the change in perception of fatigue, shortness of breath and oedema measured on a 4 point scale, ranging from worse (-1) to distinctly better (+2). A simple summation of the 3 effects (change in weight, symptomatic complaints, and oedema) was also calculated.

To quantify the association between the effects of the 'furosemide test' and the presence of heart failure, the odds ratio (OR) and p-values as computed by a logistic regression analysis were used. We also computed the area under the

receiver operator characteristic curve (c-statistic), a measure that combines the sensitivity and specificity of a diagnostic test, an assessment of discrimination between disease present versus disease absent.[16]

The OR and c-statistic were also computed for NT-proBNP (on a logarithmic scale) and the probability of heart failure being present as estimated by the referring GP upon referral is also presented.

Results

We performed PubMed searches for papers published from 1966 until September 2010, using the keywords: “loop diuretic”, “furosemide”, “frusemide”, “weight loss”, and “diuretic test” with “diagnosis”, “diagnostic test” and “heart failure” and failed to retrieve any relevant publications.

Table 2. Baseline characteristics of all 47 patients undergoing the ‘furosemide test’

	n	%
female gender	33	70.2
age (years, mean (sd))	70.8	(9.3)
medical history		
MI, CABG or PCI	5	10.6
hypertension	21	44.7
diabetes	5	10.6
COPD	14	29.8
items from history taking		
fatigue	35	74.5
shortness of breath	31	66.0
edema	30	63.8
chest pain	5	10.6
physical examination		
weight (kg, mean (sd))	81.3	(11.4)
BMI (kg/m2, mean (sd))	28.9	(3.9)
pulmonary rales	9	19.2
distended jugular veins	11	23.4
displaced apical beat	12	25.5
diagnosis of heart failure	12	25.5
reduced ejection fraction	6	
preserved ejection fraction	6	

SD, standard deviation; MI, myocardial infarction; CABG, coronary artery bypass graft surgery; PCI, percutaneous coronary intervention; COPD, chronic obstructive pulmonary disease.

In total, 54 Dutch GPs who attended a CME course on diabetes in November 2008 filled out the questionnaire about their current use in daily practice of the 'furosemide test'. The response rate was 85%. Furosemide treatment as a diagnostic test for heart failure had actually been used during the past year by 40 of the 54 (74%) GPs. Weight loss and reduction of shortness of breath were the most common measures of a positive result. More than 70% of GPs considered the furosemide test useful (table 1).

For the diagnostic accuracy study, 86 consecutive patients with suspected heart failure were considered. However, 39 patients were excluded for the following reasons: already on a diuretic (n=20), unable to assess the effects (wheelchair, hemi-paresis; n=4), or had diseases or symptoms interfering with the study (ie, tachycardia >130 bpm, systolic blood pressure >200 mmHg, angina pectoris; n=12) or refusal (n=3). Thus, 47 patients were included in the study.

Table 2 depicts the baseline characteristics of the participating patients. All patients had signs and symptoms compatible with heart failure, notably shortness of breath, fatigue or signs of fluid retention (i.e. pulmonary rales, elevated jugular venous pressure or peripheral oedema). One patient had atrial fibrillation at the time of investigation.

A dose of 40 mg furosemide once daily for 7 days was prescribed in 25 (53%) patients; the remaining 22 patients were prescribed 80mg once daily for 7 days. Ultimately, 12 (25.5%) patients were diagnosed with heart failure, of whom six had reduced (HFREF) and six had preserved left ventricular systolic function (HFPEF). The effects of the 'furosemide test' are given in table 3, along with the diagnostic value as quantified by the odds ratio and c-statistic.

None of the effect measures of the 'furosemide test' was statistically significantly associated with the presence of heart failure. Separate analysis of the patients who were eventually shown to have heart failure with reduced left ventricular ejection fraction resulted in similar figures (data not shown). None of the c-statistic values were statistically different from 0.5, the equivalence of flipping a coin.[10] Interestingly, five patients had a weight loss of 4 kg (8.8 lbs) or more in one week, of whom three did not have heart failure and two had HFREF. Conversely, 16 patients lost no more than 0.5 kg (1.1 lbs) weight in one week, of whom three had heart failure (2 HFREF and 1 HFPEF).

Table 3. Associations of the effects of the furosemide diagnostic treatment test and comparator tests with the presence of heart failure

	heart failure				p-value	OR	unit for OR	c-statistic
	yes, n=12		no, n=35					
furosemide test	mean change	sd	mean change	sd				
-weight (kg)	-1.8	1.5	-1.2	1.3	0.19	1.37	kg weight *loss*	0.65
-fatigue *	0.2	0.6	0.1	0.5	0.55	1.44	category	0.56
-shortness of breath *	0.5	0.5	0.2	0.6	0.10	2.59	category	0.67
-edema *	0.3	0.5	0.1	0.4	0.15	2.81	category	0.62
-summation of the above 4 items	1.0	1.4	0.3	1.3	0.17	1.40	summed category	0.65
comparator tests	**mean**	**sd**	**mean**	**sd**	**p-value**	**OR**	**unit for OR**	**c-statistic**
- NT-proBNP (pg/ml)	18	21	293	573	0.01	2.9	log(pg/ml)	0.78
-prior probability estimated by referring GP (%)	43	21	54	34	0.29	1.2	10% points	0.57

* –1 = worse, 0 = no difference, 1 = somewhat better, 2 = distinctly better; SD, standard deviation; OR, odds ratio; c-statistic, area under the receiver operator characteristic curve.

For the sake of comparison of diagnostic strength, the univariate ORs and c-statistics of the NT-proBNP measurement and the prior probability of heart failure estimated by the referring GP are also presented in table 3. Measurement of NT-proBNP was shown to have better diagnostic performance than the furosemide test. The estimated probability by the GPs had a diagnostic strength comparable to the furosemide test.

Discussion

A sample of 54 GPs attending a CME course on diabetes considered the 'furosemide test' a useful diagnostic test in the arsenal of the general practitioner for assessment of patients with suspected heart failure. To underscore the popularity of the 'furosemide test', 75% of the GPs mentioned that they had actually used the test in their practice in the preceding 12 months. From an evidence based point of view this is difficult to understand since no study has formally assessed the (added) diagnostic value of the furosemide test, but on the

other hand several guidelines do mention response to targeted treatment supporting the diagnosis. Furthermore, the Framingham criteria which include the furosemide test as a major criterion are still being applied.[17,18]

In our study of 47 representative, consecutive patients with suspected new slow onset heart failure from primary care, the 'furosemide test' appears to be an ineffective diagnostic tool. None of the effects of the test treatment (weight change, change in the perception of fatigue, shortness of breath or oedema) were statistically different for patients with versus patients without heart failure. When we assessed patients with HFREF, we did not find a significant result either. Given that our study did not show any relevant univariate associations, there is no rationale to pursue the 'furosemide test' any further, in patients with suspected new slow onset heart failure in primary care practice.

Although our study is relatively small, the results are clinically plausible. There is considerable diagnostic uncertainty in this population delineated by the general practitioners, as illustrated by the mean prior probability of heart failure of 40-50% as estimated by the referring GP (Table 3). As expected the diagnostic strength of the prior probability was small and comparable to the furosemide test, as this diagnostic uncertainty was the specific reason for referring the patient.

A small study could have overseen a clinically relevant diagnostic yield of a test when only p-values were taken into account. By considering also the ORs as a measure of association between the diagnostic test and the diagnosis, we found that none exceeded the value of 3. Suppose we have a diagnostic test with an odds ratio of 3, subsequently we set the positive predicted value (PPV) to be 80% (if the test is positive there is 80% probability of heart failure). That would imply that if the test is negative the probability of absence of heart failure (negative predicted value (NPV)) is only 43%, given that the OR can be written as a function of both predicted values as OR = (PPV/(1-PPV))/((1-NPV)/NPV).[19] Aiming for an PPV of 90% and a NPV of 80% would call for an OR of 36. Thus even an odds ratio of 3 would not constitute a useful test to classify an individual patient.

We do not know how the 'furosemide test' would perform in patients with suspected new acute heart failure, who are more often seen at the emergency department of the hospital. Due to the abrupt increase in preload and eventually within a few hours fluid overload, furosemide is a well established symptomatic initial treatment in patients with acute heart failure, and as such is implicitly used as a diagnostic tool.

As mentioned previously, we could not find any studies evaluating the use of loop diuretic test treatment as a diagnostic tool for heart failure. Circumstantial

evidence against the use of furosemide as a diagnostic test is given by Clarke et al.[20] in their study of 505 patients prescribed loop diuretics of whom only 56% fulfilled the diagnostic criteria for heart failure.

We would like to emphasize that the results of our study should not be generalised to patients with an established diagnosis of heart failure. To quote the ESC guidelines "It is critically important for healthcare providers to evaluate the fluid or volume status of patients with heart failure during the initial visit and each follow-up examination. This assessment plays a pivotal role in determining the need for diuretic therapy".[2]

Conclusion

Since the 1971 publication of the Framingham diagnostic criteria for heart failure, which include a 4.5 kg weight reduction on treatment as a major diagnostic criterion,[8] much has changed regarding the diagnosis and treatment of the underlying pathologies for heart failure as well as regarding the diagnosis and treatment of pathologies mimicking the signs and symptoms of heart failure. While the rationale for this diagnostic criterion is obvious and despite its apparent popularity, the present study does not support its use. We cannot support the use of the furosemide test in patients with suspected new slow onset heart failure in primary care, because our study shows that the test is of no additional value. General Practitioners should consider other ancillary diagnostic tests such as ECG, chest X-ray and natriuretic peptides to increase the accuracy of their heart failure diagnoses, as recommended in the guidelines.

References

1. 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(19):2388-2442.

2. Hunt SA, Abraham WTM, Chin MHM, Feldman AMM, Francis GS, Ganiats TGM, Jessup M, Konstam MAM, Mancini DM, Michl K, Oates JAM, Rahko PS, Silver MA, Stevenson LW, Yancy CW, Task Force MEMB, Antman EM, Smith SCJ, Adams CDM, Anderson JLM, Faxon DPM, Fuster V, Halperin JL, Hiratzka LFM, Jacobs AKM, Nishimura RA, Ornato JPM, Page RLM, Riegel B. ACC/AHA 2005 Guideline Update for the Diagnosis and Management of Chronic Heart Failure in the Adult-Summary Article: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Update the 2001 Guidelines for the Evaluation and Management of Heart Failure). Circulation 2005; 112(12):1825-1852.
3. Ezekowitz JA, Kaul P, Bakal JA, Quan H, McAlister FA. Trends in heart failure care: has the incident diagnosis of heart failure shifted from the hospital to the emergency department and outpatient clinics? Eur J Heart Fail 2011;13.
4. 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). European Heart Journal 2000; 21(22):1877-1887.
5. Fuat A, Hungin AP, Murphy JJ. Barriers to accurate diagnosis and effective management of heart failure in primary care: qualitative study. BMJ 2003; 326(7382):196.
6. Phillips SM, Marton RL, Tofler GH. Barriers to diagnosing and managing heart failure in primary care. Med J Aust 2004; 181(2):78-81.
7. 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(1):92-98.
8. McKee PA, Castelli WP, McNamara PM, Kannel WB. The natural history of congestive heart failure: the Framingham study. N Engl J Med 1971; 285(26):1441-1446.
9. Hoes AW, Voors AA, Rutten FH, Van Lieshout J, Janssen PGH, Walma EP. Multidisciplinaire richtlijn hartfalen. Huisarts en Wetenschap 2010; 53(7):368-389 (in Dutch).
10. Grobbee DE, Hoes AW. Clinical epidemiology: principles, methods, and applications for clinical research: Jones & Bartlett Publishers; 2008.
11. 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(7529):1379.
12. Cowie MR, Struthers AD, Wood DA, Coats AJS, 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(9088):1349-1353.

13. 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. European Heart Journal 2005; 26(11):1115-1140.
14. 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(5):358-367.
15. 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(3):246-270.
16. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983; 148(3):839-843.
17. Levy D, Kenchaiah S, Larson MG, Benjamin EJ, Kupka MJ, Ho KKL, Murabito JM, Vasan RS. Long-Term Trends in the Incidence of and Survival with Heart Failure. N Engl J Med 2002; 347(18):1397-1402.
18. Roger VL, Weston SA, Redfield MM, Hellermann-Homan JP, Killian J, Yawn BP, Jacobsen SJ. Trends in Heart Failure Incidence and Survival in a Community-Based Population. JAMA 2004; 292(3):344-350.
19. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. Journal of Clinical Epidemiology 2003; 56(11):1129-1135.
20. Clarke KW, Gray D, Hampton JR. Evidence of inadequate investigation and treatment of patients with heart failure. British Heart Journal 1994; 71(6):584-587.

Chapter 6

The Clinical utility of three B-type natriuretic peptide assays for the initial diagnostic assessment of new slow onset heart failure

based on:

Journal of Cardiac Failure. 2011 Sep;17(9):729-34. Epub 2011 May 28.

Clinical utility of three B-type natriuretic peptide assays for the initial diagnostic assessment of new slow-onset heart failure.

Kelder JC, Cramer MJ, Verweij WM, Grobbee DE, Hoes AW.

Abstract

Background

In patients suspected of new, slow onset heart failure data on the comparative diagnostic performance of automated B-type Natriuretic Peptide (BNP) assays is scarce.

Methods and results

Two hundred patients referred to a heart failure outpatient diagnostic facility underwent standard diagnostic work-up including echocardiography. The reference standard for the diagnosis of heart failure was an expert panel conclusion. N-terminal pro-BNP on Elecsys and BNP on Axsym and Centaur machines were measured in a single batch. Data were available for 172 patients; 51 had heart failure (29.7%). All 3 tests had high c-statistic values. An intermediate risk subset of 111 patients (34% with heart failure) was created by excluding patients with very high or very low probability based on history and physical examination, the subgroup most in need of an additional test. Applying different thresholds for ruling heart failure in or out, the positive predicted values in this 'gray zone' group were 75%, 76%, and 72%, respectively, and the negative predictive values 83%, 71%, and 85%, with the remaining 50% of patients having approximately 18% probability of heart failure.

Conclusion

In practice, a valid diagnosis in patients suspected of slow-onset heart failure remains elusive for many in the absence of echocardiographic imaging.

Introduction

Measurement of B-type natriuretic peptides (BNP) is useful in the diagnostic assessment of patients suspected of heart failure and recommended by the 2008 European Society of Cardiology (ESC) Heart Failure guideline.[1] They are particularly useful in settings with limited access to other diagnostic tests such as echocardiography, notably in primary care. BNP and the biologically inactive N-terminal counterpart (NT-proBNP) peptide assays are nowadays readily available with fully automated immunoassay as well as point-of-care testing methods.

Data on the comparative performance of 3 popular automated assays in patients suspected of new, slow onset heart failure is lacking. Our aim was to assess the comparative diagnostic accuracy and utility of NT-proBNP measured with the Roche assay, BNP measured with the Abbott assay and BNP with the Bayer assay in patients suspected of new, slow onset heart failure in primary care on top of the preceding tests (signs and symptoms, history, physical examination).[2] We were particularly interested, as measure of utility, in the predicted values of the 3 BNP tests in those patients where the clinical picture is unclear.

Methods

Patient Population

The first 200 patients included in a larger study, the UHFO-IA study[3], had their blood drawn for assessment of BNP by the 3 assays. The Utrecht Heart Failure Organisation – Initial Assessment (UHFO-IA) study recruited patients suspected of heart failure by their general practitioner (GP). The only exclusion criteria were a previous diagnosis of heart failure or acute signs and symptoms demanding immediate treatment. The objective was to include patients where the GP could neither immediately and safely rule out, nor diagnose heart failure; in other words patients in need of additional diagnostic work-up. Patients were referred to rapid access heart failure outpatient diagnostic facilities available in 8 hospitals. All patients underwent standard diagnostic work-up including ECG, laboratory measurements, chest X-ray, spirometry and echocardiography. The reference ('gold') standard for the diagnosis of heart failure was the decision of an expert panel consisting of a cardiologist, a pulmonologist and a GP. The panel based their decision on the results of all diagnostic tests: medical history, anamnesis, physical examination, laboratory values, ECG, spirometry, chest X-ray, echocardiography and six months clinical follow-up data (the latter e.g. to monitor the effect of targeted therapy). The panel did not receive the results of the B-type natriuretic peptides to prevent incorporation bias, since one of the aims of the original study was to assess the added diagnostic value of BNP in these patients. The final

decision of the panel was made following the criteria for heart failure of the 2008 European Society of Cardiology guideline[4] and more recently explicitly for patients suspected of heart failure by the HFSA 2010 heart failure guideline[5].

This study was approved by the medical ethical committee of the St Antonius Hospital, Nieuwegein, The Netherlands.

B-type natriuretic peptide measurements

Blood samples were taken and specimens of plasma were stored at −70 °C. At the end of the study, NT-proBNP and BNP levels were measured for all patients in a single batch after the frozen specimens were thawed at the SALTRO laboratories in Utrecht, The Netherlands. NT-proBNP was measured with an automated non-competitive immunoradiometric assay (Roche Inc., Mannheim, Germany) on an Elecsys 1010 analyser. For plasma BNP measurements the automated Abbott AxSYM BNP immuno-assay (Abbott Inc., Park Ill, USA) and ADVIA Centaur BNP immuno-assay (Siemens Healthcare Diagnostics Inc., Deerfield, IL, USA) were used. Total coefficients of variation are reported to be respectively 4.4%, 5.5% and 0.8%.[6] Results are given in pg/ml.

Data analysis

The ability of a BNP assay diagnostic test to discriminate between patients with and without heart failure was assessed by means of the c-statistic (area under the receiver operator characteristic (ROC) curve). The c-statistic is a rank order measure of discrimination combining sensitivity and specificity; specifically the c-statistic represents the probability of a random patient with the heart failure having a higher value of plasma BNP compared to a random patient without heart failure.[7] To assess the calibration we performed a goodness-of-fit test, the Hosmer-Le Cessie (HLC) test[8] in which a smaller p-value indicates larger difference between observed and expected probabilities of heart failure. Additionally, we computed predicted values, taking the 25th and 75th percentile of BNP values to respectively “rule out” and “rule in” heart failure as an, arbitrary, example, to enable to compare the 3 BNP assays. In daily practice, the cutoffs ruling out and diagnosing heart failure are chosen by the physician, taking into account the individual patient for whom the decision has to be made.

One other algorithm we assessed was published by the European Society of Cardiology (ESC) in the 2008 guideline: heart failure is unlikely when BNP <100 pg/ml (NT-proBNP <400 pg/ml) whereas heart failure is likely when BNP >400 pg/ml (NT-proBNP >2000 pg/ml).[1]

To gain more insight into the utility of the 3 tests, we selected those patients for whom the test would be considered indicated most appropriately, i.e. as

additional diagnostic test in patients where there is still diagnostic uncertainty after history taking and physical examination. To that end, we computed the predicted probability of heart failure for all patients based on a multivariate logistic regression model and selected the group of patients with less than 80% probability and more than 10% probability. The multivariate model used age, history of myocardial infarction, CABG or PCI, use of a loop diuretic, displaced ictus cordis, lung crackles, irregular pulse, pulse rate, heart murmur suggestive of mitral regurgitation and elevated jugular venous pressure to predict the presence of heart failure. This model was derived from all the 721 patients from the main study where the analyses took the natural hierarchy according to daily practice (starting with easily obtainable items from history taking) into account. Variables were allowed in the model only if they had additional value (based on the likelihood ratio test) conditional on items already included in the model.[3] More than 80% probability of heart failure present after history taking and physical examination would constitute the arbitrary threshold of not needing additional diagnostic tests and vice versa less than 10% is the arbitrary threshold of not needing additional diagnostic tests because heart failure is discarded from the differential diagnosis. In the 10-80% probability of heart failure group we computed the predicted values of the 3 BNP tests. All statistical calculations were performed with R, version 2.10 (http://www.r-project.org/).

Table 1. Characteristics of patients suspected of heart failure (HF) and according to the presence or absence of HF

	all participants (N=172)		**HF present (N=51)**		**HF absent N=121)**		**intermediate probability of heart failure (N=111; 38 HF)**	
	N	%	N	%	N	%	N	%
age, years, mean(sd)	70.2	(11.3)	75.4	(9.7)	68	(11.2)	74.4	(8.3)
female	113	65.7	30	58.8	83	68.6	71	64.0
Complaints								
shortness of breath	103	59.9	41	80.4	62	51.2	76	68.5
fatigue	121	70.3	40	78.4	81	66.9	79	71.2
ankle swelling	83	48.3	31	60.8	52	43.0	54	48.6
orthopnoea or paroxysmal nocturnal dyspnoea	58	33.7	23	45.1	35	28.9	35	31.5

Table 1. Continued

	all participants (N=172)		HF present (N=51)		HF absent N=121)		intermediate probability of heart failure (N=111; 38 HF)	
History								
never smoked	67	39.0	17	33.3	50	41.3	48	43.2
hypertension	88	51.2	28	54.9	60	49.6	62	55.9
diabetes	29	16.9	13	25.5	16	13.2	25	22.5
stroke or tia	15	8.7	9	17.6	6	5.0	11	9.9
atrial fibrillation	8	4.7	3	5.9	5	4.1	8	7.2
mi, pci or cabg	9	5.2	7	13.7	2	1.7	7	6.3
copd	47	27.3	16	31.4	31	25.6	33	29.7
Medication								
acei or at2blocker	52	30.2	22	43.1	30	24.9	40	36.0
loop diuretic	61	35.5	32	62.7	29	24.0	49	44.1
ß-blocker	49	28.5	15	29.4	34	28.1	30	27.0
Physical examination and other test results								
BMI kg/m^2, mean(sd)	29.5	(5.4)	28.5	(5.1)	29.9	(5.4)	28.9	(5.3)
pulmonary rales	26	15.1	14	27.5	12	9.9	19	17.1
elevated jugular venous pressure	15	8.7	11	21.6	4	3.3	9	8.1
laterally displaced apex beat	16	9.3	14	27.5	2	1.7	8	7.2
peripheral edema	48	27.9	20	39.2	28	23.1	31	27.9
eGRF MDRD, ml/min/m^2, mean(sd)	62.9	(15.0)	58.6	(15.4)	64.8	(14.4)	62	(14.6)
ejection fraction ≥0.45-0.50 on echocardiogram	130	75.6	20	39.2	110	90.1	81	73.0

HF = Heart Failure; tia = transient ischemic attack; mi = myocardial infarction; pci = percutaneous coronary intervention; cabg = coronary artery bypass grafting; copd = chronic obstructive pulmonary disease; acei = angiotensin converting enzyme inhibitor; at2blocker = angiotensin II blocking agent; BMI = body mass index; eGFR MDRD = estimated Glomerular Filtration Rate using the Modification of Diet in Renal Disease formula.

Results

Of the 200 samples 28 were lost due to technical or organizational reasons e.g. insufficient amount of blood, id label unreadable or lost. There were no relevant differences between patients with and without plasma samples (data not shown); consequently all analyses were performed on 172 patients. The mean age of the 172 patients was 70.2 years and 66% was female (Table 1). Heart failure was diagnosed by the panel in 51 (29.7%) patients of whom 41% had an ejection fraction as estimated by echocardiography of ≥45-50%. All patients had one or more complaints compatible with heart failure. Physical examination signs compatible with heart failure were more common among patients with heart failure. In table 2 and figure 1 the uni- and bi-variate measures of the 3 BNP tests are depicted. As expected, the distribution off all BNP assays was skewed to the right, therefore we used the log-transformed values in all computations. The Pearson correlation coefficient on the log scale was lowest for NT-proBNP with BNP on Axsym (0.84) and highest for NT-proBNP with BNP on Centaur (0.90). The largest c-statistic, the measure of discriminating ability between heart failure present versus heart failure absent, was found for the NT-proBNP test (0.86) but no statistically significant differences were found between the tests (table 2).

For the utility of the 3 tests we selected the patients of intermediate risk of heart failure after history taking and physical examination (>10% and <80%). This subset contained 111 patients of whom 38 (34.2%) had heart failure, the characteristics are given in table 1. Application of the 3 BNP tests in these patients resulted in loss of diagnostic ability as compared to the full cohort without relevant changes between the tests (table 2). The number of patients associated with application of the 'percentile' rule-in and rule-out cutoff levels is given in table 3. At best the prior probability of 34.% is increased to 75.9% when the test is positive (>75th percentile) and decreased to 18.0% when the test is negative (<25th percentile). Application of the ESC based cutoff levels resulted in lower negative predicted values and very few patients in the rule in category.

The only diagnosis other than heart failure that was systematically assessed was COPD, defined as GOLD (Global Initiative for Chronic Obstructive Lung Disease) stage ≥2. Fifty-four patients were diagnosed with COPD (20 in patients with heart failure, 34 in patients without heart failure).

Table 2. Univariate measures and diagnostic accuracy of 3 BNP assays for the diagnosis of heart failure (n=172; prevalence of heart failure 29.7%)

	NT-proBNP	Axsym-BNP	Centaur-BNP
all patients			
geometric mean	24.8	18.0	13.2
median (25th-75th percentile)	16 (7-55)	14 (8-34)	10 (5-26)
patients with heart failure			
geometric mean	129.5	61.3	50.4
median (25th-75th percentile)	185 (28-470)	85 (26-288)	58 (18-222)
patients without heart failure			
geometric mean	12.4	10.7	7.5
median (25th-75th percentile)	12 (6-22)	12 (6-18)	7 (4-14)
c-statistic (95%ci)	0.86 (0.80-0.92)	0.82 (0.73-0.90)	0.83 (0.76-0.91)
HLC p-value for calibration	0.051	<0.001	0.005
PPV > 75th percentile (95%CI)	0.81 (0.67-0.92)	0.74 (0.59-0.86)	0.77 (0.61-0.88)
NPV < 25th percentile (95%CI)	0.89 (0.75-0.97)	0.84 (0.69-0.93)	0.88 (0.75-0.96)
PPV BNP >400 pg/ml, NT-proBNP >2000 pg/ml (N/N)	1/1	5/5	3/3
NPV BNP <100 pg/ml, NT-proBNP <400 pg/m (95%CI)	0.76 (0.69-0.82)	0.81 (0.73-0.87)	0.80 (0.73-0.86)
applied in 111 patients of intermediate risk of heart failure, prevalence 34.2%			
c-statistic (95%ci)	0.86 (0.73-0.89)	0.79 (0.68-0.90)	0.81 (0.71-0.91)
PPV >75th percentile (95%CI)	0.75 (0.57-0.89)	0.76 (0.56-0.90)	0.72 (0.53-0.86)
NPV <25th percentile (95%CI)	0.83 (0.59-0.96)	0.71 (0.48-0.89)	0.85 (0.62-0.97)
PPV BNP >400 pg/ml, NT-proBNP >2000 pg/ml (N/N)	na (N=0)	3/3	2/2
NPV BNP <100 pg/ml, NT-proBNP <400 pg/m (95%CI)	0.71 (0.61-0.79)	0.76 (0.66-0.84)	0.76 (0.66-84)

geometric mean = exponential of the mean of log transformed mean; c-statistic = area under the Receiver Operator Characteristic curve, a measure of discrimination, a value of 1 is perfect discrimination, 0.5 is equivalent to flipping a coin; CI = confidence interval; PPV = positive predicted value; NPV = negative predicted value.

Discussion

BNP, whether it is measured with the NT-proBNP on the Roche Elecsys, BNP on the Bayer Centaur and BNP on the Abbott AxSYM, is a helpful diagnostic instrument in the assessment of new, slow onset heart failure. When comparing these 3 assays, the inference could be that most information regarding the diagnosis of heart failure is given by NT-proBNP, followed by BNP on the Centaur and least by BNP on the Axsym but differences are small. The NT-proBNP showed marginally better discriminatory power for detecting heart failure as

demonstrated by the c-statistic of 0.86, the BNP on the Centaur resp. Axsym scored 0.83 and 0.82. There were no statistically significant differences on the non-parametric c-statistic scale.

Table 3. Utility of 3 BNP assays for the diagnosis of heart failure applied in 111 patients at intermediate risk of heart failure (34.2% prevalence)

	patients with disease (N)	patients without disease (N)	prevalence of heart failure	% of all patients
'percentile' guided				
>75th percentile				
- NT-proBNP	24	8	75.0%	28.8%
- Axsym-BNP	22	7	75.9%	26.1%
- Centaur-BNP	23	9	71.9%	28.8%
<25th percentile				
- NT-proBNP	3	15	16.7%	16.2%
- Axsym-BNP	6	15	19.7%	18.9%
- Centaur-BNP	3	17	15.0%	18.0%
between 25th and 75th percentile				
- NT-proBNP	11	50	18.0%	55.0%
- Axsym-BNP	10	51	16.4%	55.0%
- Centaur-BNP	12	47	20.3%	53.2%
ESC guideline[(1)]				
BNP >400 pg/ml (NT-proBNP >2000 pg/ml)				
- NT-proBNP	0	0	na	0%
- Axsym-BNP	3	0	100%	2.7%
- Centaur-BNP	2	0	100%	1.8%
BNP <100 pg/ml (NT-proBNP <400 pg/ml)				
- NT-proBNP	30	73	29.1%	92.8%
- Axsym-BNP	23	71	24.5%	84.7%
- Centaur-BNP	23	73	24.0%	86.5%
BNP 100-400 pg/ml (NT-proBNP 400-2000 pg/ml)				
- NT-proBNP	8	0	100%	7.2%
- Axsym-BNP	12	2	85.7%	12.6%
- Centaur-BNP	13	0	100%	11.7%

The c-statistic can be interpreted as the probability that a test will rank a randomly chosen patient with heart failure higher than a randomly chosen patient without heart failure[7] (even if this situation will never emerge in practice).

However, the predictive values have more practical use, because ultimately, a non-perfect diagnostic test should provide an indication of presence of disease expressed as a probability. Moreover we selected from our cohort only patients with diagnostic uncertainty after taking history taking and physical examination into account. Thus we excluded patients with very high (>80%) or very low (<10%) probability of heart failure, these patients do not need an additional diagnostic test. We used different cutoff values for ruling in heart failure and ruling out heart failure as is proposed for instance in the ESC guideline.[4] The chosen cutoff levels were the 75th respectively 25th percentile since no evidence based levels exist for patients suspected of new, slow onset heart failure. For each test the actual values were 7 and 55 for NT-proBNP, 8 and 34 for BNP on the Axsym and 5 and 26 for BNP on the Centaur. In doing so we created an intermediate group for whom no diagnosis could be made, this amounts to 55% of the patients, predominantly due to the choice of cutoff levels. Add to this amount the false positives and false negatives and the NT-proBNP, BNP on Axsym and BNP on Centaur have respectively 64.9% (n=72), 66.7% (n=74) and 68.0% (n=71) of patients not adequately categorized. One could argue that these figures are not very effective to reach at a conclusion in a fair number of patients, but fortunately a practising physician has more diagnostic tests at his or her disposal, notably (among others): signs and symptoms, course in time, ECG and echocardiography.

The proposed cutoff levels from the ESC guideline appear to be too high, resulting in the majority of patients ending up in the rule out category, reducing the negative predicted value.

Our results are in line with previous reports. Recently, the analytical performance and diagnostic accuracy of immunoassays for BNP, among which the 3 assays we report, were compared.[6,9] The general conclusion was that all automated natriuretic peptide methods showed acceptable analytic performance and clearly differentiated between healthy individuals and heart failure patients. Because, however, results of one assay cannot be substituted for another, it is important to know which specific assay is being used in a specific institution. As mentioned before, this 'test research' is not directly applicable in daily practice.[10] A study in patients with selection criteria comparable to our study, i.e. patients suspected of new, slow onset heart failure, gave similar results comparing a point-of-care BNP test with the NT-proBNP assay.[11] The utility of the tests was assessed by their ability to prevent unnecessary referrals for echocardiography. The point-of-care test prevented 24% while NT-proBNP prevented 25% unnecessary referrals, using the 'single cutoff level' method. These results emphasize the ability of BNP to discriminate between heart failure present versus absent. The calibration on the other hand is not often reported. Where calibration is not an issue in a 'one cutoff level' binary test, the interpretation in daily practice of a diagnostic test such as

BNP will be ordinal by nature, e.g. a BNP level twice the upper level of normal will have another interpretation compared to a BNP level of 10 times the upper level of normal. In our study all 3 BNP assays suffered from lack of calibration as single tests. Figure 2 illustrates the lack of differentiation in the lower 3 quintiles. It should be interesting to study whether re-calibration can be accomplished by adding other tests, notably physical examination.

In our study consecutive patients were referred to the participating hospitals from primary care to reduce selection bias. Patients in need of urgent care were excluded, consequently our study does not address the diagnosis of acute heart failure.

A point of interest is the absence of a 'gold' standard diagnostic test for heart failure. The presence of heart failure in our study was established by consensus evaluation, using all available diagnostic information. This is an established method as reference standard.[12]

In patients suspected of new, slow onset heart failure 3 common BNP assays (Roche NT-proBNP, Abbott AxSYM BNP and Bayer Centaur BNP) have comparable and satisfactory diagnostic power when used as a single test. However, a valid diagnosis remains difficult in a substantial proportion of patients and there is room for improvement in the early diagnosis of heart failure in the absence of full echocardiographic imaging.

Funding/Support

This study was funded by the Dutch Ministry of Health, ZON-MW grant 945-02-014. Roche Diagnostics Inc. (Mannheim, Germany), Abbott Inc. (Illinois, USA), and Bayer Inc. (Leverkusen, Germany, now Siemens Healthcare Diagnostics Inc., Deerfield, IL) supplied, unrestrictively, the assays for analysis of NT-proBNP, BNP AxSYM and BNP Centaur, respectively.

Figure 1. Kernel density plot (violin plot) plus boxplot for patients with and without heart failure (HF) on a log scale. Top left: NT-proBNP; top right: BNP on Axsym; bottom left: BNP on Centaur

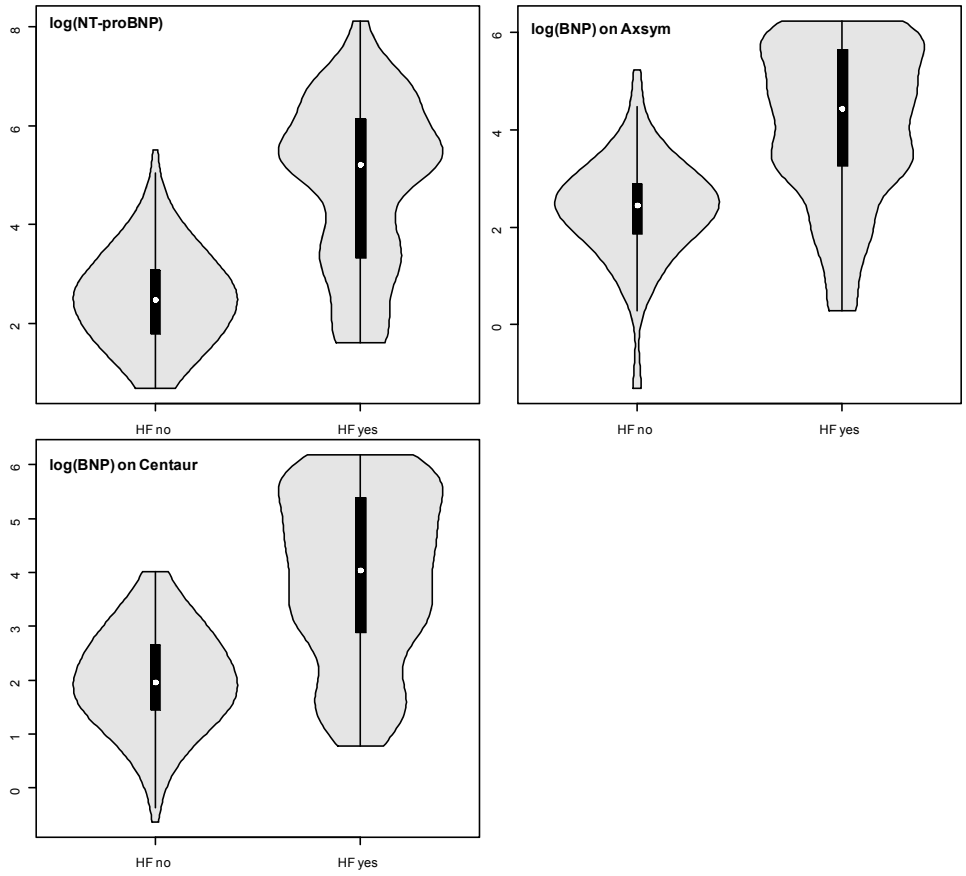
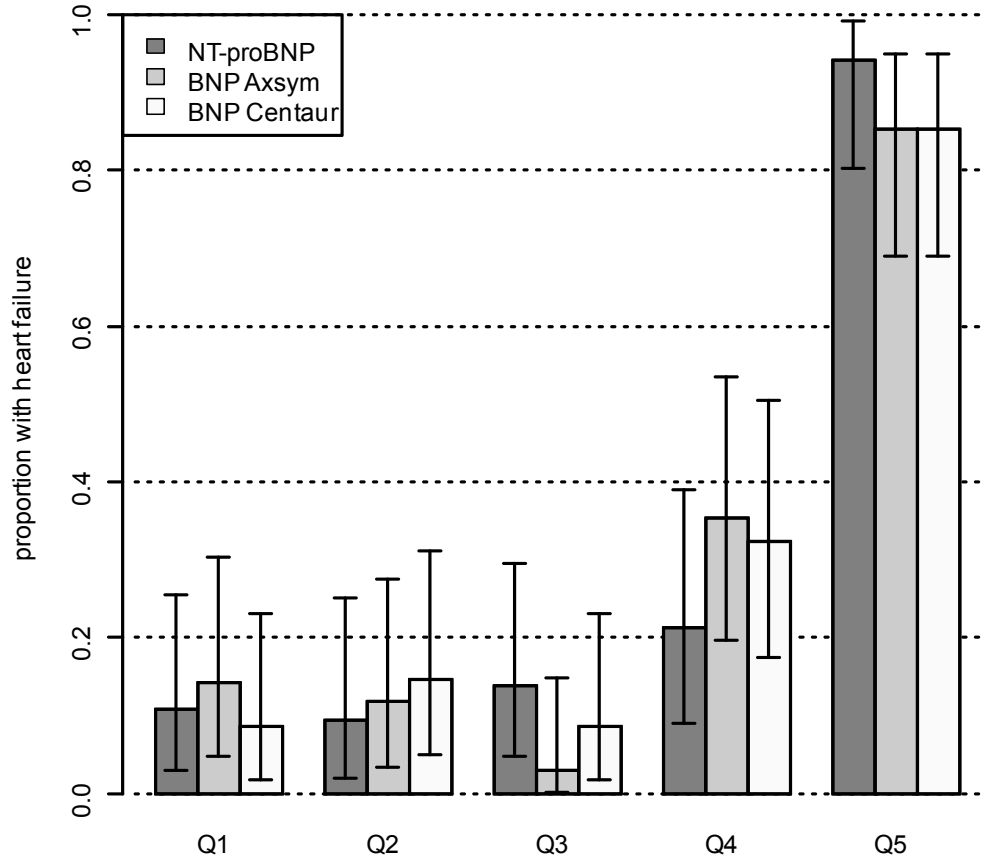


Figure 2. Calibration of probability of heart failure by quintiles (Q) of 3 BNP tests, with 95% confidence intervals.



References

1. 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.
2. Kelder JC, Rutten FH, Hoes AW. Clinically relevant diagnostic research in primary care: the example of B-type natriuretic peptides in the detection of heart failure. Fam Pract. 2009;26:69-74.
3. Kelder JC, Hoes AW, Cramer MJ, Wijngaarden JV, Moons KG, Grobbee DE. Diagnostic rule for the initial assessment of suspected heart failure. a practical tool for clinicians (abstract). Eur Heart J. 2009;30 (suppl 1):829.
4. Authors/Task Force M, 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, Guidelines ESCCfP, Vahanian A, Camm J, De Caterina R, Dean V, Funck-Brentano C, Hellemans I, Kristensen SD, McGregor K, Sechtem U, Silber S, Tendera M, Widimsky P, Zamorano JL, Document R, Auricchio A, Bax J, Bohm M, Corra U, della Bella P, Elliott PM, Follath F, Gheorghiade M, Hasin Y, Hernborg A, Jaarsma T, Komajda M, Kornowski R, Piepoli M, Prendergast B, Tavazzi L, Vachiery J-L, Verheugt FWA, Zannad F. 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.
5. Lindenfeld J, Albert NM, Boehmer JP, Collins SP, Ezekowitz JA, Givertz MM, Katz SD, Klapholz M, Moser DK, Rogers JG, Starling RC, Stevenson WG, Tang WH, Teerlink JR, Walsh MN. HFSA 2010 Comprehensive Heart Failure Practice Guideline. J Card Fail. 2010;16:e1-194.
6. Rawlins ML, Owen WE, Roberts WL. Performance characteristics of four automated natriuretic peptide assays. American Journal of Clinical Pathology. 2005;123:439-445.
7. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29-36.

8. Hosmer DW, Hosmer T, Cessie SL, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. Statistics in Medicine. 1997;16:965-980.
9. Clerico A, Prontera C, Emdin M, Passino C, Storti S, Poletti R, Zyw L, Zucchelli GC. Analytical performance and diagnostic accuracy of immunometric assays for the measurement of plasma B-Type Natriuretic Peptide (BNP) and N-Terminal proBNP. Clin Chem. 2005;51:445-447.
10. Moons KGM, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. Clin Chem. 2004;50:473-476.
11. Fuat A, Murphy JJ, Hungin APS, Curry J, Mehrzad AA, Hetherington A, Johnston JI, Smellie WSA, Duffy V, Cawley P. The diagnostic accuracy and utility of a B-type natriuretic peptide test in a community population of patients with suspected heart failure. British Journal of General Practice. 2006;56:327-333.
12. Moons KGM, Grobbee DE. When should we remain blind and when should our eyes remain open in diagnostic studies? Journal of Clinical Epidemiology. 2002;55:633-636.

Chapter 7

Economic analysis of diagnostic strategies for patients suspected of non-acute heart failure in primary care

Kelder JC, van Hout BA, Cramer MJ, Grobbee DE, Hoes AW.

(Submitted)

Abstract

Aims

To provide evidence based treatment promptly to patients with heart failure requires early diagnosis. When signs and symptoms suggest heart failure the general practitioner (GP) may follow a variety of diagnostic strategies. Strategies may be dynamic and while a GP may think that the present signs and symptoms may not be sufficient to justify referral to a cardiologist, this may change when the patient presents his/herself with more advanced symptoms. Here, these strategies are assessed with emphasis on the balance between costs and effects.

Methods and Results

In this economic analysis of the UHFO-IA study (721 patients with suspected heart failure) strategies were used to distinguish whom to treat immediately, whom to refer to the cardiologist and in whom to refrain from heart failure treatment. Seven strategies were compared: the diagnostic rule derived in this study (with four different sets of criteria to rule heart failure in or out), the GP decision, refer all patients to the cardiologist and do nothing. A model was built to estimate effects, taking into account estimates on the condition of the patient (accuracy of the diagnosis, survival, indices on quality of life) and transition probabilities between conditions, as well as to estimate the costs, for which two methods were used. The average UHFO-IA patient had an estimated 5.92 years to live (3.11 QALYs) costing €2688 (method 1) or €2489 (method 2) when the (hypothetically) 'do nothing' strategy was to be followed. Other strategies resulted in:

	Life years gained	QALYs gained	Method 1	Method 2
Refer all	6.76	3.57	€ 2,598	€ 2,172
<20% = do not treat >70% = do treat in between = refer	6.48	3.45	€ 2,584	€ 2,185
<10%; >70%	6.72	3.56	€ 2,581	€ 2,163
<20%; >80%	6.48	3.45	€ 2,583	€ 2,184
<10%; >80%	6.72	3.62	€ 2,537	€ 2,126
GP-guided	6.44	3.41	€ 2,728	€ 2,366

Conclusion

It is cost effective to apply the UHFO-IA derived rule to arrive at one of three outcomes: 1) heart failure probably not present (probability less than 10% or 20%); do not start treatment for heart failure, 2) heart failure probably present (probability more than 70% or 80%); start treatment for heart failure and 3) diagnostic uncertainty remains then refer for the cardiology consultation.

Introduction

Heart failure is defined as a complex clinical syndrome that can result from any structural or functional cardiac disorder that impairs the ability of the ventricle to fill with or eject blood.[1] Heart failure is associated with impaired quality of life and decreased life expectancy; an estimated 40% of patients die within the first year, thereafter the annual mortality can be less than 10%.[2, 3] When a patient has heart failure, there are a number of treatment options. Among these are diuretics, aldosteron antagonists, ACE-inhibitors, beta-blockers, angiotensin II receptor antagonists and digoxin. Additionally, a number of devices can help change the prognosis of patients with heart failure such as pacemakers, defibrillators and as a last resort a left ventricular assist device. For some selected cases heart transplantation may be an option.[1, 4]

Given the availability of effective treatment, acknowledging the presence of heart failure and treating the condition, may increase one's life expectancy. However, the signs and symptoms are diffuse and available additional tests to determine whether a patient has heart failure may be inaccurate. The most objective test to assess the presence of heart failure, i.e. an echocardiogram read by a cardiologist, implies that patients have to be seen by a cardiologist. This may be more costly than is strictly necessary and alternative strategies may be worthwhile to consider.

When patients present themselves at a general practitioner's (GP) office and there are signs and symptoms which suggest heart failure, the GP may follow a variety of diagnostic strategies. At any point in time within those strategies, two decisions are possible, treatment or no treatment. As such, strategies may be dynamic and while a GP may think that the present signs and symptoms may not be sufficient to justify referral to a cardiologist, this may change when the patient presents his/herself with more advanced symptoms.

In the main study, of which this economic analysis is a part, data were collected of 721 patients with suspected heart failure, and different strategies have been defined whom to treat immediately, whom to refer to the cardiologist and in whom to refrain from heart failure treatment (for the time being).[5] Here, these strategies are assessed with emphasis on the balance between costs and effects.

Methods

Data were collected concerning 721 patients with suspected heart failure. After referral to the cardiologist 207 (28.7%) were categorized as heart failure patients and 514 as not (or not yet) suffering from heart failure.

A diagnostic rule, based on signs and symptoms, the patient's medical history and the result of a plasma BNP-test, was derived to help GPs decide whether the patient may suffer from heart failure: the rule included in total 10 items. Each patient is given a number of points on the basis these variables. It starts with 0 points if one is under 60, 4 points if one is between 60-70, 7 points when between 70-80 and 10 points if over 80. 15 points are added in case of a history of myocardial infarction (MI), coronary artery bypass grafting (CABG) or percutaneous coronary intervention (PCI). When already using a loop diuretic, 10 points are added. If there is a displaced apex beat, 20 points are added, in case of rales 14 points, in case of an irregular pulse 11 points, when there is a heart murmur suggestive of mitral regurgitation 10 points and in case of an elevated jugular venous pressure 12 points are added. For pulse rate every 3 beats per minute over 60 adds 1 point. Additional points are given in case of an increased N-terminal pro-B-type Natriuretic Peptide (NT-proBNP) (pg/ml). When this value is between 100-200, 8 points are added, when between 200-400 16, when between 400-800 24, when between 800-1600 32, between 1600-3200 40 and when this value is over 3200, 48 points.

On the basis of the number of points that each patient is assigned a predicted probability that the patient suffers from heart failure is derived. Now four different strategies are defined on the basis of that prediction.

1. **20/70 rule:** If the patient has a predicted probability of heart failure of <20%: wait and see; if the patient has a predicted probability of >70%: start treatment. In both cases the patient is not referred to the cardiologist. Contrarily, the patient is referred if the predicted probability lies between 20% and 70%.
2. **10/70 rule:** as 1. but with 10% and 70%
3. **20/80 rule:** as 1. but with 20% and 80%
4. **10/80 rule:** as 1. but with 10% and 80%
5. **GP-decision:** an alternative is to follow the GPs intuition. In the clinical part of the study GPs were asked what they would have done if no diagnostic rapid-access outpatient clinic would have been available. Answers were available for 354 patients.

Only for the sake of comparison we defined the non-real world strategies 6 and 7:

6. **Do nothing:** no treatment and perfect diagnosis at no costs.
7. **Refer all:** refer all patients to the cardiologist, having perfect diagnosis and treatment.

All patients were assessed by an outcome panel to judge during consensus meetings whether the patient indeed suffered from heart failure. This assessment is defined as the reference (or 'gold') standard and thus for each strategy the

percentage of correct positive, false positive, correct negative and false negative diagnoses can be estimated. Table 1 presents the results of the strategies.

It should be noted that the 38 patients diagnosed 'false negative' according to the 20/70 rule are the same patients diagnosed 'false negative' according the 20/80 rule. Similarly, the 11 patients diagnosed 'false negative' according to the 10/70 rule are the same patients diagnosed 'false negative' according the 10/80 rule. Logically, the ones diagnosed 'true positive' are also the same.

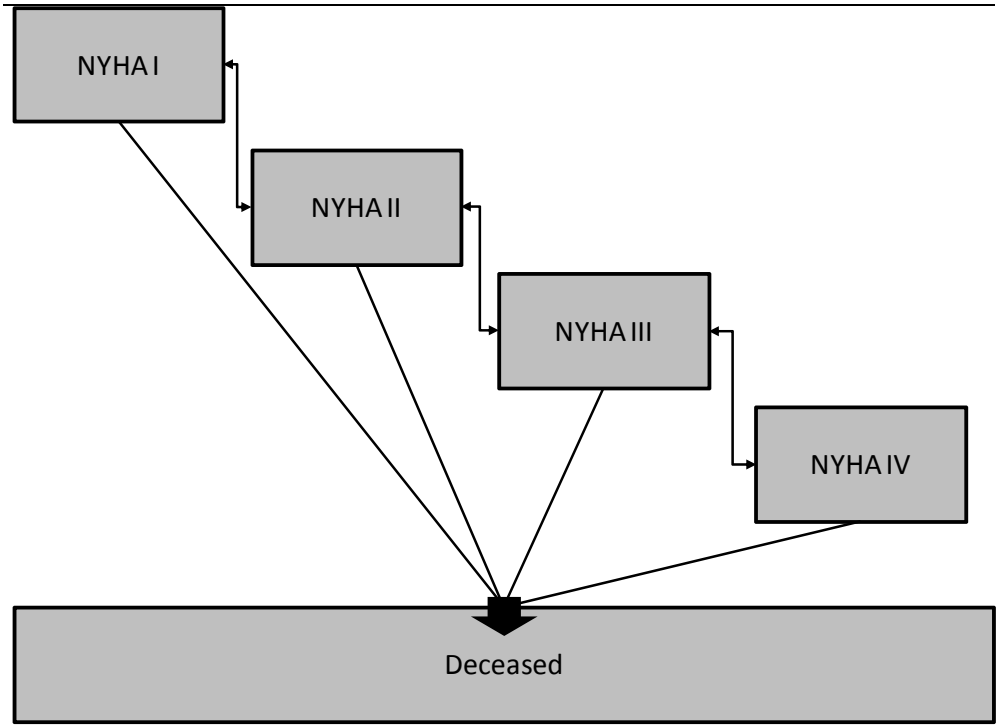
Table 1. Results diagnostic strategies

	heart failure: No	heart failure: Yes
20/70 rule		
heart failure considered absent: no treatment	365	38
uncertain: referral to cardiologist	137	85
heart failure considered present: treat	12	84
10/70 rule		
heart failure considered absent: no treatment	222	11
uncertain: referral to cardiologist	280	112
heart failure considered present: treat	12	84
20/80 rule		
heart failure considered absent: no treatment	365	38
uncertain: referral to cardiologist	143	106
heart failure considered present: treat	6	63
10/80 rule		
heart failure considered absent: no treatment	222	11
uncertain: referral to cardiologist	286	133
heart failure considered present: treat	6	63
GP-decision		
heart failure considered absent: no treatment	133	53
uncertain: referral to cardiologist	104	42
heart failure considered present: treat	12	10

A model

As an aid to analyzing what the best strategy is, a small but simple model is used. Figure 1 presents the model. It categorizes patients according to their NYHA class, distinguishing between NYHA class I to IV. At each point in time patients may move one category up or down.

Figure 1.



NYHA Class I: patients with no limitation of activities; they suffer no symptoms from ordinary activities.
Class II: patients with slight, mild limitation of activity; they are comfortable with rest or with mild exertion.
Class III: patients with marked limitation of activity; they are comfortable only at rest.
Class IV: patients who should be at complete rest, confined to bed or chair; any physical activity brings on discomfort and symptoms occur at rest.

The model may also be written in matrix algebra where *x* measures the numbers of patients in each category and $a_{I\text{-}d}(t)$ the probability to move from state *I* to state *d* within the time interval [*t-1,t*].

$$\begin{bmatrix} x_I(t) \\ x_{II}(t) \\ x_{III}(t) \\ x_{IV}(t) \\ d(t) \end{bmatrix} = \begin{bmatrix} a_{I-I} & a_{II-I} & 0 & 0 & 0 \\ a_{I-II} & a_{II-II} & a_{III-II} & 0 & 0 \\ 0 & a_{II-III} & a_{III-III} & a_{IV-III} & 0 \\ 0 & 0 & a_{III-IV} & a_{IV-IV} & 0 \\ a_{I-d} & a_{II-d} & a_{III-d} & a_{IV-d} & 1 \end{bmatrix} \begin{bmatrix} x_I(t-1) \\ x_{II}(t-1) \\ x_{III}(t-1) \\ x_{IV}(t-1) \\ d(t-1) \end{bmatrix}$$

This is the model in its most simple form. For various reasons, a number of states have been added.

- To distinguish between patients in NYHA II and III with and without a history of observed decrements in health, both health states are broken down in two: with and without such former decrements. This is needed for the wait and see strategies in which patients are treated after a decrement; i.e. treatment naive versus treated patients. This increases the number of health states with two.
- All health states (referring to patients who are alive) are repeated to take account of patients who are and who are not taking their medication. This increases the number of health states with six.
- To enable distinguishing between death due to heart failure and other causes of death, two types of "death" are distinguished: death due to heart failure, and death due to other causes. This increases the number of health states with one.

Thus, there is a total of 14 health states. Additionally to adding health states, all transition probabilities are made time dependent, to reflect the effect of age.[6]

The core of the model is formed by the transition probabilities. Two relatively recent articles are used as the main sources. With respect to the survival probabilities usage is made of the Seattle Heart Failure model.[7] With respect to the transition probabilities to go from one NYHA state to another the estimates are used from Yao et al.[6]

Model estimates

Mortality; the Seattle Heart Failure equation

The Seattle Heart Failure model (SHFM) is a multivariate risk equation which aims to predict expected survival after 1, 2 and 3 years. It was derived from a cohort of 1125 heart failure patients and was validated using data from a number of trials. The equation can be written as: $S(t) = \exp(-\lambda t.\exp(x'\beta))$

The shape parameter λ is estimated at 0.0405 on the basis of the PRAISE1 data. The β coefficients from the so called SHFM score ($\exp(x'\beta)$), also based on PRAISE1 data, are tabulated in table 2. On the basis of table 2 one may estimate survival for a patient taking, for example, treatment with statins or diuretics into account. The same study reports additional estimates reflecting the effects of various treatments expressed as additional decreases in the SHFM score. In case of an ACE-inhibitor one may subtract a value 0.261, 0.415 in case of a β-blocker, 0.139 for an angiotensin receptor blocker and 0.301 for an aldosteron antagonist.

Table 2. The Seattle Heart Failure model coefficients

Variable	Unit of measurement	β	95% confidence interval
Age	Decade	0.086	(-0.015 - 0.186)
Gender	Male=1; female=0	0.085	(-0.176 - 0.346)
NYHA	Class	0.470	(0.019 - 0.921)
EF	100/EF	0.030	(0.010 - 0.049)
Ischaemic etiology	yes=1, no=0	0.303	(0.071 - 0.535)
SBP(if<160)	10mm HG	-0.131	(-0.195 - -0.067)
Diuretics	mg/kg/day	0.164	(0.093 - 0.236)
Allopurinol	yes=1, no=0	0.452	(0.157 - 0.746)
Statin	yes=1, no=0	-0.462	(-0.892 - -0.022)
Sodium if <138	138-Sodium	0.049	(0.005 - 0.093)
Total cholesterol	100/(mg/dl)	0.791	(0.044 - 1.538)
Hemoglobin <16	16- hemoglobin	0.117	(0.052 - 0.182)
Hemoglobin >16	Haemoglobin-16	0.290	(0.010 - 0.569)
% Lymphocytes (<47%)	per 5%	-0.109	(-0.167 - -0.050)
Uric acid (>3.4)	mg dl	0.062	(0.022 - 0.103)

EF = left ventricular ejection fraction (%); SBP = systolic blood pressure

When analyzing the coefficients, it may be noted that age has to increase with 5.5 years to get the same effect as one step up in NYHA. To be of use in the model it is

assumed that death due to other causes is not included, which is the main reason why two types of death are included in our analyses.

The UHFO-IA study cohort

Using the model and distinguishing the NYHA classes one may estimate the expected one year survival of the UHFO-IA study (Utrecht Heart Failure Organisation - Initial Assessment) cohort. NYHA class was derived from the patient data. In the absence of exact data, total cholesterol is set at 193 mg/dl, lymphocytes at 19% and uric acid at 3.4 mg/dl. Table 3 summarizes the medication use as relevant in the mortality risk equation. For the distributions of variables and the comparison for heart failure and non heart failure patients we refer to the main study. [5]

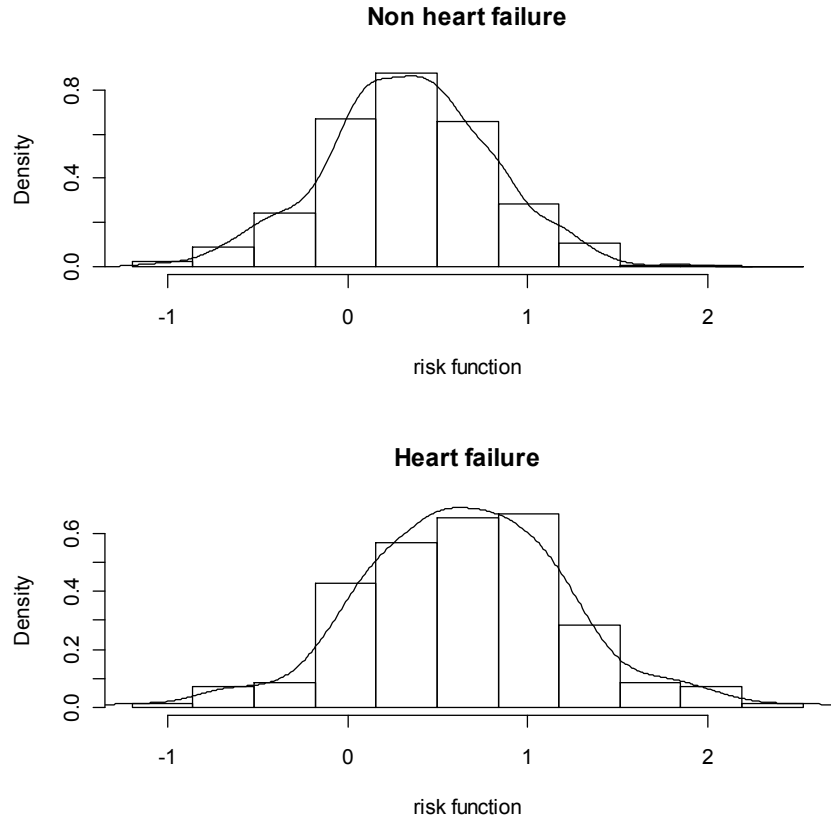
When combining all data and selecting only the heart failure patients, we find that the average patient in NYHA II is estimated to have a 1-year probability to die of 5.56%; while the average patient in NYHA III has a 9.73% 1-year probability to die. Figure 2 presents the distributions of the risk function (depicted as linear predictors $(x'\beta)$).

Before being able to use the estimated probabilities of mortality it should be noted that the model uses probabilities to die while being in a certain NYHA class *within* a time interval, while the risk equation predicts the probability to die given that one is in certain NYHA class *at the beginning* of a time interval. So, it may well be that no-one dies in NYHA II or III, but that all patients die in NYHA IV. Additionally it may be that explanatory variables change in time and that the risk profile of a patient changes completely when going from one NYHA stage to another. To be able to understand this better, it was analyzed to what extent part of the differences in survival per NYHA class was explained by differences in age, treatment and other variables. When these components are distinguished in the Seattle Heart Failure model score we derive the last four columns in table 4.

The difference in $x'\beta_{age}$ between patients in NYHA II en NYHA III (patients with heart failure) reflects that patients in NYHA III are, on average, 1.88 years older. Surprisingly, we find a positive difference in the risk due to treatment.

Table 5 shows the percentages of patients on treatment and we find that in the group with heart failure there are fewer patients in NYHA III who receive statins, angiotensin receptor blockers and aldosteron antagonists; this while there are small differences in the use of ACE-inhibitors, diuretics and beta-blockers.

Figure 2. Estimated risk (depicted are the linear predictors) by heart failure



When all is taken into account we find an additional difference of 0.023 in $x'\beta_{other}$ between patients with heart failure in NYHA II en NYHA III. This is very small and given that the model takes into account the effects of the medication use and of the increasing age, no additional corrections seem needed in the risk factors when moving from one state to another. This does not imply that no further correction would be needed for the potential overestimation of the mortality in the lower NYHA stages and the potential underestimation of mortality in the higher stages. Indeed patients in class II are likely to die in III or IV and the bias may increase with the time horizon. It is neglected here as the overestimation is compensated by an underestimation and given the fact that the time horizon of the estimates was relatively short.

Table 3. Medication data

	heart failure: no				heart failure: yes			
MEDICATION	NO		YES		NO		YES	
ACE-inhibitors	425	82.7%	89	17.3%	131	63.3%	76	36.7%
beta-blockers	402	78.2%	112	21.8%	145	70.0%	62	30.0%
statins	449	87.4%	65	12.6%	182	87.9%	25	12.1%
aldosteron antagonists	505	98.2%	9	1.8%	198	95.7%	9	4.3%
AT II receptor blockers	457	88.9%	57	11.1%	190	91.8%	17	8.2%

The probability to die from other causes is estimated to be equal to the age and gender specific probability to die as estimated by the Dutch Bureau of Statistics. No correction is made for the fact that this also includes a probability to die of heart failure. Figure 3 shows the hazard rates. These rates define part of the survival of the patients with heart failure and all of the survival of patients without heart failure. The average age of the patients without heart failure is 68.8 years and the average life expectancy 16.4 years. When discounted at an annual rate of 5% this is 9.9 years.

Table 4. Estimated probabilities to die (according to the Seattle Heart Failure model) in the UHFO-IA cohort

Heart failure	NYHA class	N	Estimated probability to die	$x'\beta_{age}$	$x'\beta_{NYHA}$	$x'\beta_{Rx}$	$x'\beta_{other}$
NO	II	226	4.6%	0.566	0.940	-0.159	-1.280
	III	288	7.2%	0.614	1.410	-0.213	-1.270
YES	II	57	5.6%	0.639	0.940	-0.287	-1.067
	III	150	9.7%	0.655	1.410	-0.211	-1.044

Although it is assumed that survival of non-heart failure patients is not affected by the diagnostic strategy, treatment is. Patients who are diagnosed false positive

are assumed to be treated indefinitely with beta-blockers and ACE-inhibitors. As such, the costs of treatment of non-heart failure patients do depend on the diagnostic strategy and are obviously minimal for the gold standard.

Figure 3. Annual probability to die by age and gender for the average Dutch population in 2008.

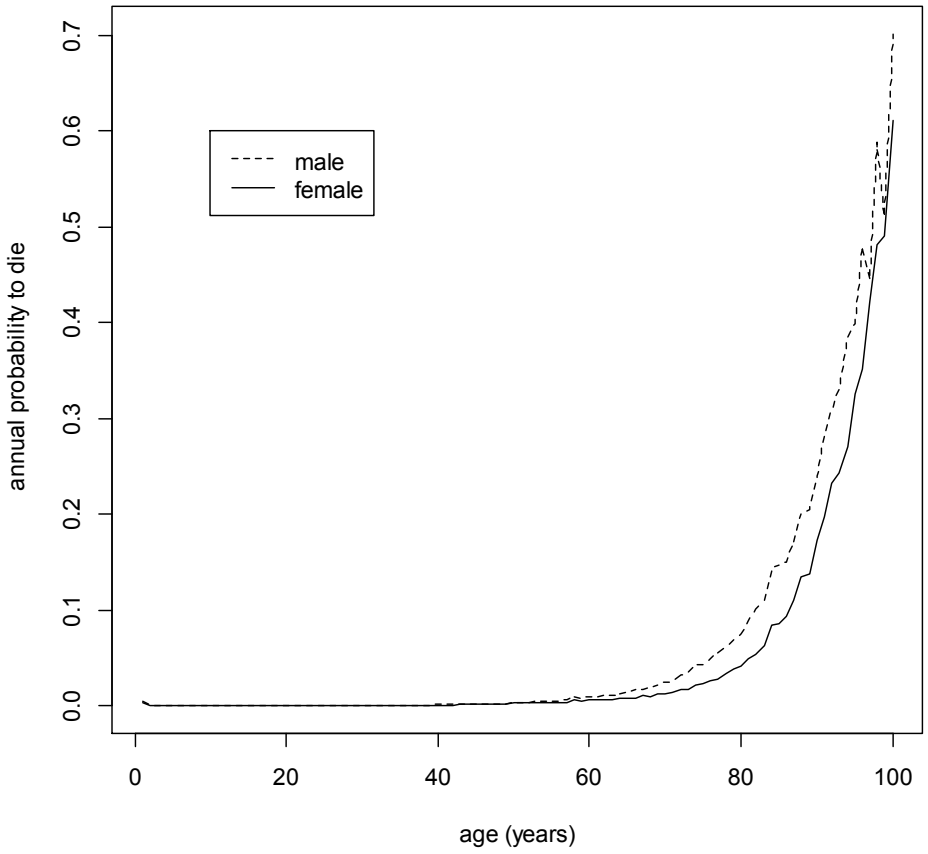


Table 5. Medication use in the UHFO-IA cohort

	no heart failure		heart failure	
	NYHA II	NYHA III	NYHA II	NYHA III
Diuretic	16.4%	28.5%	50.9%	56.7%
ACE-inhibitor	10.6%	22.6%	36.8%	36.7%
Beta blocker	19.0%	24.0%	28.1%	30.7%
Statin	11.5%	13.5%	22.8%	8.0%
Angiotensin receptor blocker	9.7%	12.2%	10.5%	7.3%
Aldosteron antagonist	0.9%	2.4%	7.0%	3.3%

Transition probabilities from NYHA class to NYHA class

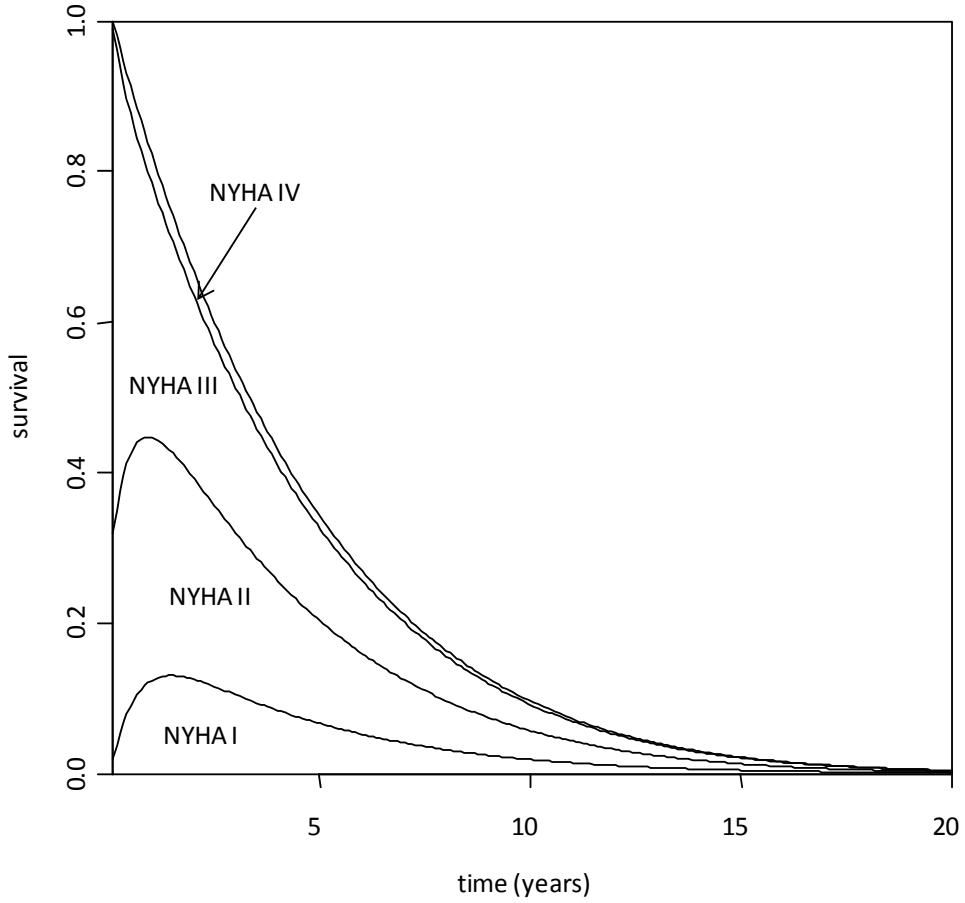
As a first estimate of the probability to go from one state to another we use the estimates from Yao et al.[6] concerning medical therapy of heart failure patients. For a 65-year old patient they use the monthly probabilities reported in table 6. The transition probabilities are time independent. Here, we estimate that the transition probabilities concerning an increase in NYHA (worsening) slightly increase with age and the transition probabilities to get better (a decrease in NYHA) slightly decrease with age. We estimate that this change is in the same range as the age specific increase in the death rate and set it at 1% per year.

Table 6. NYHA transition probabilities (for a 65 years old person)

To\From	I	II	III	IV
I	0.7956	0.0710	0.0047	0.0000
II	0.1245	0.8448	0.0893	0.1064
III	0.0738	0.0765	0.8845	0.1064
IV	0.0061	0.0077	0.0216	0.7872

Naturally, account is taken of the fact that all estimates concern probabilities and that there is the possibility, when extrapolating these independently, they sum up to a figure exceeding 1. So, all probabilities to transfer are transformed into rates, subsequently, the probability to stay in a stage is calculated, and then all rates are transformed back into probabilities. When combining the various estimates concerning mortality and transition probabilities (and disregarding the effects of diagnosis and treatment) estimates can be obtained of the time that each patient is expected to spend in each stage (and of their life expectancy) taking account of their baseline characteristics. Figure 4 presents the sum over all patients with heart failure.

Figure 4. Expected survival of the study cohort patients with heart failure without extra treatment (distribution stacked according to NYHA)



Strategies

When it is known that patients have heart failure, they can be treated. A variety of agents can be prescribed. The Seattle Heart Failure Model includes ACE-inhibitors, β-blockers, angiotensin receptor blockers and aldosteron antagonists. Here, we limit ourselves to the two most important ones: ACE-inhibitors and β-blockers. When patients are not yet on these drugs, their use decreases their SHFM-score with a value of 0.261 (ACE-inhibitor) and 0.416 (β-blocker). The associated hazard ratios are 0.77 and 0.66. We use the same hazard ratios to express that treatment does not only decrease mortality but also slows the disease and multiply the probabilities to go from a lower NYHA class to a higher with the same ratios. We assume that treatment does not affect the probabilities to improve.

Now we compare multiple strategies all differing in the sensitivity to pick up patients as being heart failure patients. When testing there are four different outcomes possible:

- Correct Positive. We estimate that patients who are diagnosed correct positive will get treated with a β-blocker and an ACE-inhibitor and we adapt the transition-probabilities.
- False Positive. We estimate that patients who are false positive will get treatment but that their transition-probabilities are not affected.
- Correct Negative. We estimate that patients who are diagnosed correct negative will not get treatment and that their transition-probabilities are not affected.
- False Negative. We estimate that patients who are diagnosed false negative will not get treatment and that their transition-probabilities are not affected. That is until they worsen. Then, they will be sent to the cardiologist for further testing and the decision will be taken that the patient does suffer from heart failure and the transition probabilities are affected.

As indicated earlier, to enable to have patients start their medication later onwards, two stages have been added to the model: NYHA II, treatment naive and NYHA III, treatment naive. For patients who do not have heart failure, the probabilities to die from heart failure are set to zero.

When none of the patients receives treatment, the model predicts a life expectancy of 5.61 years. Of these years, on average 0.62 years is spent in NYHA I, 2.04 in NYHA II, 2.53 in III and 0.42 in IV. Moreover, the model predicts that patients go from NYHA I to II on average 0.93 times, from II to III on average 0.93 times and from III to IV on average 1.04 times. Tables 7 and 8 hold the same numbers and compares them with the estimates when all patients are treated with ACE-inhibitors and β-blockers.

Table 7. Life expectancy (years) and distribution over time in NYHA classes

	NYHA I	NYHA II	NYHA III	NYHA IV	Total
Do nothing	0.62	2.04	2.53	0.42	**5.61**
Refer all	0.99	2.45	2.59	0.36	**6.38**

In the estimates account is taken of the notion that when patients receive treatment, they may stop taking their medication. It is estimated that during the first year 10% of patients per year stop taking their medication. This percentage is estimated to decrease slightly with 1% (relative) per year.

Table 8. Estimated number of transitions between NYHA classes

	No medication		ACE-inhibitors and β-blockers	
	improve	Worsen	improve	Worsen
I	1.58	NA	1.89	NA
II	2.85	0.93	2.86	1.10
III	0.46	0.93	0.40	0.99
IV	NA	1.04	NA	0.90
death due to heart failure	NA	0.70	NA	0.64
death due to other causes	NA	0.30	NA	0.36

Costs

For the year 2003, the Dutch National Institute for Public Health and the Environment (RIVM) estimated the costs of heart failure at €375 million. For the same year, it was estimated that 178,900 people had heart failure (77,200 men and 101,700 women; 9.6 per 1000 men and 12.4 per 1000 women). At the same time, the incidence was estimated at 19,200 females and 16,200 males. As such the total costs per incident case can be estimated at €10,593. The total costs related to hospitals and nursing homes are €8,475. Twenty three percent of the costs are related to the costs of long term nursing homes and 57% due to hospitalisations. The other 20% are related to drugs, overhead, GP care and others.

Method 1

A first estimate of the costs per stage can be obtained by risk-ratios. Estimates of such risk-ratios have been published by Yao et al. [6] They estimate that patients in NYHA II costs 1.18 as much as patients in NYHA I, that patients in NYHA III costs 1.83 as much as patients in NYHA I and patients in NYHA IV 4.99 as much. When limiting this to the costs of hospitalisations and nursing homes (the costs of medication are accounted for) we model 80% of €375 million. Then when using the average time in the various states and assuming that ¾ of the patients are on ACE-inhibitors, one obtains estimates as in table 9. Logically, when multiplying those average costs per year with the average time in each state, an estimate is obtained of €8,475, the average institutional costs per incident case.

Table 9 also holds the estimates of the quality adjusted life years (QALYs) indices from Yao et al.[6]

Table 9. Estimates of costs per year and QALYs

	Costs per year	QALY
NYHA I	€ 1,011	0.815
NYHA II	€ 1,197	0.720
NYHA III	€ 1,854	0.590
NYHA IV	€ 5,044	0.508

Method 2

An alternative is to acknowledge that most hospital costs are made when patients deteriorate. So, hospital costs may best be linked to the changes from one state to another. Assuming that ¾ of the patients are on ACE-inhibitors and beta-blockers, one may estimate the number of changes to NYHA III at 0.93 per incident case and the number of changes to IV at 1.04 per incident case. When assuming that the hospital costs are associated with these movements one may estimate the costs per move at €3,065. Because the nursing home costs per incident case are €2,436 and the average time in NYHA III is 2.10 year (assuming 75% treated with ACE-inhibitors and beta-blockers and starting with an average duration of 5.05 years on the basis of the national incidence and prevalence figures) and in NYHA IV 0.31 year. We then further assume that the costs of nursing homes are associated with being in either NYHA III or IV (and applying the Yao ratios), we estimate the average annual costs of nursing homes in those states at €830 and €2,258. The costs of treatment with ACE-inhibitors and β-blockers are estimated at €12.50 each per month. The costs of an assessment by a cardiologist are estimated at €200 and the costs of the BNP test are estimated at €20.

Table 10. Expected life years, mortality and durations in NYHA classes

	N	NYHA I	NYHA II	NYHA III	NYHA IV	Alive	Death due to HF	Other death
Do nothing	207	0.62	2.04	2.53	0.42	5.61	69.5%	30.5%
Refer all	207	0.99	2.45	2.59	0.36	6.38	64.2%	35.8%
<20; >70 and <20; >80								
correct positive	169	0.91	2.30	2.47	0.35	6.03	64.0%	36.0%
false negative	38	0.96	3.08	3.17	0.46	7.68	57.5%	42.5%
total		0.90	2.47	2.70	0.40	6.48	62.8%	37.2%
<10; >70 and <10; >80								
correct positive	196	0.98	2.55	2.80	0.41	6.75	65.1%	34.9%
false negative	11	0.85	2.52	2.46	0.33	6.15	50.3%	49.7%
total		0.97	2.55	2.78	0.41	6.72	64.3%	35.7%
GP-guided								
correct positive	52	1.00	2.47	2.59	0.37	6.42	61.6%	38.4%
false negative	53	0.70	2.50	2.82	0.44	6.46	63.5%	36.5%
total		**0.85**	**2.49**	**2.71**	**0.41**	**6.44**	**62.6%**	**37.4%**

Results

Table 10 presents the estimates of the numbers of life years in the various health states according to the various strategies. When none of the patients receives additional therapy, the life expectancy of the patients with heart failure is 5.92 years. When all patients get therapy the life expectancy increases to 6.76 years. Using the 20/70 rule or the 20/80 rule does not make any difference in the decisions about patients *with* heart failure. When following those rules the average life expectancy increases to 6.48 years. Similarly, using the 10/70 rule or the 10/80 rule does not make any difference in the decisions about patients *with* heart failure. Now, more patients get treated and the average life expectancy increases to 6.72 years. When following the GP-guidance strategy the average life expectancy is 6.44 years. This is considerably less, but here account has to be taken of the fact that the patient characteristics of the 105 patients whom are included here are different from the 207 from the other strategies. A similar argument explains the higher life expectancy of patients who are found false negative in some of the decision-rule strategies. Table 11 presents the average ages and estimates of the numbers of life years lost in comparison to similar individuals (regarding age and sex) without heart failure.

Table 11. Average age and life years lost due to heart failure

	average age		average number of life years lost due to heart failure	
	correct positive	false negative	correct positive	false negative
Refer all	75.5	NA	5.4	0.0
<20; >70 and <20; >80	76.4	71.4	4.3	7.3
<10; >70 and <10; >80	76.0	66.3	4.3	11.3
GP-guided	73.1	76.4	6.3	4.2

Table 12. Expected number of transitions by state

	Do nothing		**Refer all**		**<20; >70 and <20; >80**		**<10; >70 and <10; >80**		**GP-guided**	
Correct positive	207		207		169		196		52	
Towards	**improve**	**worsen**	**improve**	**worsen**	**improve**	**worsen**	**improve**	**worsen**	**improve**	**worsen**
NYHA I	1.58	0.00	1.90	0.00	1.77	0.00	1.88	0.00	1.88	0.00
NYHA II	2.85	0.93	2.86	1.10	2.71	1.03	2.86	1.09	2.70	1.08
NYHA III	0.46	0.93	0.40	0.99	0.37	0.95	0.40	0.99	0.36	0.97
NYHA IV	0.00	1.04	0.00	0.90	0.00	0.85	0.00	0.90	0.00	0.82
dead due to HF	0.00	0.70	0.00	0.64	0.00	0.64	0.00	0.65	0.00	0.62
other dead	0.00	0.30	0.00	0.36	0.00	0.36	0.00	0.35	0.00	0.38
False Negative	0		0		38		11		53	
Towards					**improve**	**worsen**	**improve**	**worsen**	**improve**	**worsen**
NYHA I					2.09	0.00	2.04	0.00	1.83	0.00
NYHA II					3.03	1.22	2.72	1.18	2.90	1.07
NYHA III					0.42	1.12	0.34	1.07	0.42	1.02
NYHA IV					0.00	0.94	0.00	0.77	0.00	0.95
dead due to HF					0.00	0.57	0.00	0.50	0.00	0.64
other dead					0.00	0.43	0.00	0.50	0.00	0.36

Table 12 depicts the permutations of transition probabilities. The movements towards NYHA III en NYHA IV are involved in costs method 2. Now, we arrive at table 13 presenting the average institutional cost per patient as calculated by the two methods for the assorted strategies. Besides the institutional costs, account has to be taken of the costs of the BNP tests and the costs of referral to the cardiologist. Table 14 presents the results. Next, account has to be taken of the costs of medication and in doing so account has to be taken of the fact that some of the patients are already using beta blockers and/or ACE-inhibitors.

Table 13. Expected costs (hospital and nursing home)

Strategy	Life years	Costs Method 1	Costs Method 2
Do nothing	5.92	€ 8,361	€ 7,742
Refer all	6.76	€ 8,815	€ 7,295
<20; >70 and <20; >80	6.48	€ 8,560	€ 7,183
<10; >70 and <10; >80	6.72	€ 8,780	€ 7,292
GP-guided	6.44	€ 8,503	€ 7,271

Table 14. Costs of diagnostic testing

	BNP-test	referral to cardiologist	costs of diagnosis
Do nothing	0%	0%	€ 0
Refer all	0%	100%	€ 200
<20; >70	100%	30.8%	€ 92
<10; >70	100%	54.4%	€ 132
<20; >80	100%	34.5%	€ 100
<10; >80	100%	58.1%	€ 139
GP-guided	0%	41.2%	€ 112

Table 15 presents the results in terms of the numbers of years that the various groups of patients are estimated to use medication. The last column presents the costs of the medication taking account of the numbers of patients in each category.

Table 15. Years of medication use after diagnosis (discounted) and associated costs

	Correct positive		False negative		False positive		Costs
	B-blocker	ace inhibitor	B-blocker	ace inhibitor	B-blocker	ace inhibitor	
Do nothing	0.00	0.00	0.00	0.00	0.00	0.00	€ 0
Refer all	3.11	2.90	0.00	0.00	0.00	0.00	€ 259
<20; >70	3.15	2.62	2.14	2.64	6.61	6.76	€ 264
<10; >70	3.18	2.83	1.89	3.36	6.61	6.76	€ 277
<20; >80	3.15	2.62	2.14	2.64	6.64	6.43	€ 252
<10; >80	3.18	2.83	1.89	3.36	6.64	6.43	€ 266
GP-guided	3.06	2.74	2.04	1.89	8.36	8.86	€ 686

Finally, table 16 presents the results in terms of total costs and life years. It is noted that the costs do not include the institutional costs of patients without heart failure. These are not expected to differ between any of the strategies.

Table 16. Total cost and effects

	Effects		Costs per QALY gained	
	Life years gained	QALYs gained	Method 1	Method 2
Do nothing	5.92	3.11	€ 2,688	€ 2,489
Refer all	6.76	3.57	€ 2,598	€ 2,172
<20; >70	6.48	3.45	€ 2,584	€ 2,185
<10; >70	6.72	3.56	€ 2,581	€ 2,163
<20; >80	6.48	3.45	€ 2,583	€ 2,184
<10; >80	6.72	3.62	€ 2,537	€ 2,126
GP-guided	6.44	3.41	€ 2,728	€ 2,366

Figure 5 a, b c and d present the costs, life years and QALYs graphically. Each time we magnified the area around the 10/70, 20/70, 20/80 and 10/80. We conclude that the 10/70 and 20/70 strategies as well as the GP-guided strategy are not on the border of the concave cost efficacy plane. This is formed by 1) doing nothing 2) 20/80 3) 10/80 and 4) refer all. So, the relevant costs effectiveness ratios concern those comparing 1) 20/80 versus doing nothing; 2) 10/80 versus 20/80 and 3) refer all versus 10/80. Table 17 presents the results.

Figure 5a to d. Costs (method 1 and 2) and effects (life years gained and QALYs)

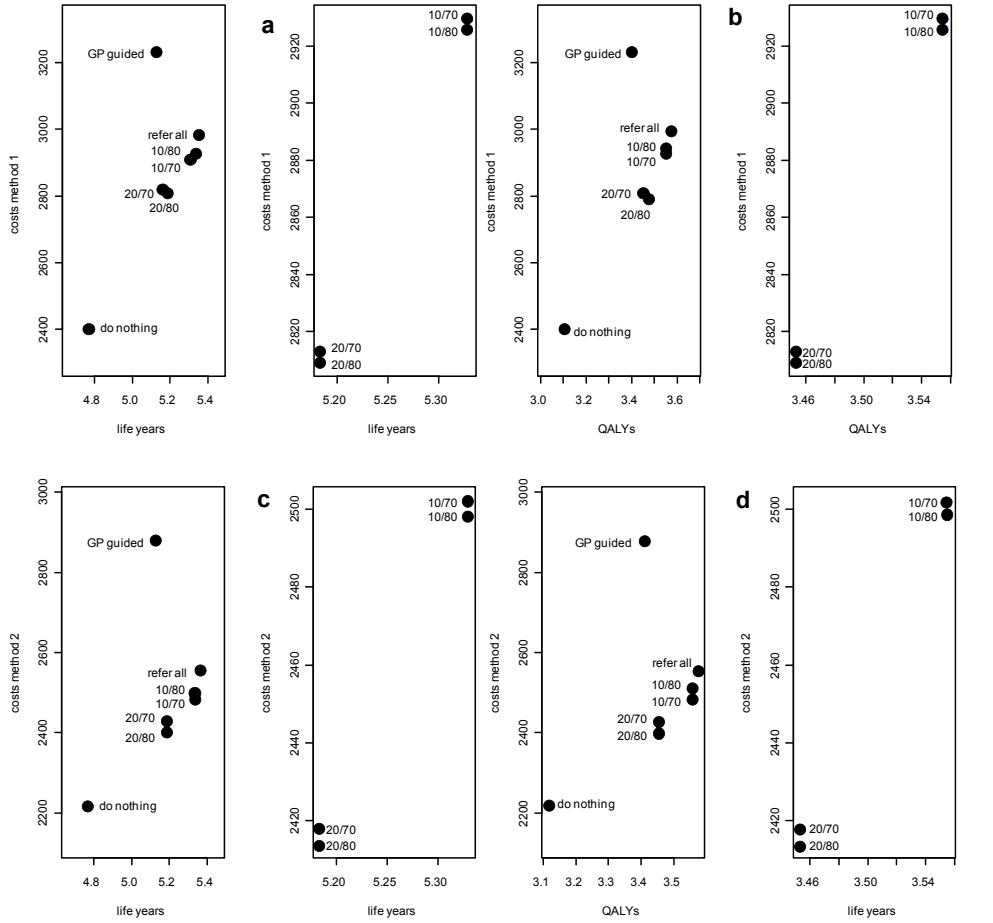


Table 17. Cost effectiveness ratios

	cost method 1		cost method 2	
	life years	QALYs	life years	QALYs
20/80 vs do nothing	€ 979	€ 1,210	€ 458	€ 566
10/80 vs 20/80	€ 786	€ 1,118	€ 571	€ 812
refer all vs 10/80	€ 2,544	€ 3,500	€2,188	€ 3,010

Discussion

The UHFO-IA study set out to enhance diagnostic capabilities for the general practitioner with respect to heart failure. Heart failure is difficult to diagnose in early stages, and even at later stages, because other concomitant diseases are also able to express the same symptomatology.

The diagnostic simple rule from the UHFO-IA study is a major enhancement of the diagnostic accuracy as indicated by the area under the ROC-curve as a measure of discriminatory power of 0.86. When this benefit is gauged by prognostic impact immediately the deadly nature of heart failure is revealed, whereby the gain in diagnostic power by means of the simple diagnostic rule, will extend life on average for about 6 months.

We restricted our cost calculation model pertaining to diagnosis to one NT-proBNP measurement and one cardiology consultation including one echocardiogram; pertaining to medical therapy decisions ensuing diagnosis we only took ACE-inhibitors and beta-blockers into account. For that reason the life time costs as depicted in table 16 are underestimated, albeit for all strategies. The discussion will focus on comparing strategies. From the UHFO-IA diagnostic rule which outputs the probability of heart failure, we selected 4 plausible strategies leading to rejection of the diagnosis heart failure (when less than 10% or 20% probability of heart failure), accepting the diagnosis of heart failure without further ado and begin treatment (more than 70% or 80% probability) or when neither acceptation nor rejection can be concluded and cardiologic consultation is sought.

There is a near linear relation between life years left and QALYs left. The period based cost calculation method 1 consequently amounts to higher figures compared to the transition based method 2. With the exception of the “do nothing” strategy, method 1 was 15% more expensive per crude life year or QALY left. When comparing the most noteworthy real life contrast 10/80 versus 20/80 strategy, the cost per QALY gained using cost calculation method 1 amounts to €1,118 and for cost calculation method 2 to €812 (table 17). This is the reflection of the false negative rate of 5.3% for the 10/80 strategy versus 18.4% for the 20/80 strategy.

Costs calculation method 2 provides a framework which is conceptually more modifiable and intuitive: preventing transitions for the worse is an instrument to work on as compared to postponing death which makes things only more costly.

We also must address the fact that about 50% of patients were not given an alternative route by their GP in case the rapid access diagnostic hospital based

facility would not have been available. We cannot rule out that 'confounding factors' were at play, i.e. that the non-response of the GPs may have been influenced by patient characteristics that may influence the cost and effects of the diagnostic strategies.

The 'do nothing' strategy leading to no treatment and perfect diagnosis at no costs is of course purely hypothetical. The 'refer all' strategy in which all patients are referred to the cardiologist, have perfect diagnosis and treatment, also has the hypothetical status as well as serious side effects in the form of extra diagnostic measures (ranging from extra consultations to coronary angiography) and vast extensive logistic consequences not included in the model.

We conclude that in patients consulting their GP for symptoms suggestive of non-acute heart failure it is cost effective (pertaining to the extension of lifetime), to have a plasma NT-proBNP measurement as the only additional diagnostic procedure, after history taking and physical examination, subsequently applying the UHFO-IA derived rule to arrive at one of three outcomes:

1. heart failure probably not present (probability less than 10% or 20%); do not start treatment for heart failure.
2. heart failure probably present (probability more than 70% or 80%); start treatment for heart failure.
3. diagnostic uncertainty remains; go for the cardiology consultation.

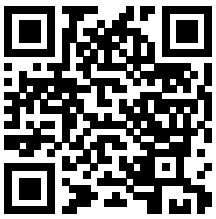
Limitations

The estimates of costs drawn from statistics dating from 2003. Unfortunately, at the time of writing (March 2012) no such detailed statistics are available from a more recent time.

References

1. Hunt SA, Abraham WT, Chin MH, Feldman AM, Francis GS, Ganiats TG, Jessup M, Konstam MA, Mancini DM, Michl K, Oates JA, Rahko PS, Silver MA, Stevenson LW, Yancy CW. 2009 Focused Update Incorporated Into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Heart Failure in Adults: A Report of the American College of Cardiology Foundation/ American Heart Association Task Force on Practice Guidelines Developed in Collaboration With the International Society for Heart and Lung Transplantation. Journal of the American College of Cardiology. 2009;53:e1-e90.
2. Cowie MR, Wood DA, Coats AJS, Thompson SG, Suresh V, Poole-Wilson PA, Sutton GC. Survival of patients with a new diagnosis of heart failure: a population based study. Heart. 2000;83:505-510.
3. Hobbs FDR, Roalfe AK, Davis RC, Davies MK, Hare R, and the Midlands Research Practices Consortium Prognosis of all-cause heart failure and borderline left ventricular systolic dysfunction: 5 year mortality follow-up of the Echocardiographic Heart of England Screening Study (ECHOES). Eur Heart J. 2007;28:1128-1134.
4. 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.
5. 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.
6. Yao G, Freemantle N, Calvert MJ, Bryan S, Daubert JC, Cleland JGF. The long-term cost-effectiveness of cardiac resynchronization therapy with or without an implantable cardioverter-defibrillator. European Heart Journal. 2007;28:42-51.
7. Levy WCM, Mozaffarian D, Linker DT, Sutradhar SCP, Anker SDM, Cropp ABP, Anand IS, Maggioni A, Burton P, Sullivan MD, Pitt B, Poole-Wilson PA, Mann DLM, Packer M. The Seattle Heart Failure Model: Prediction of Survival in Heart Failure. Circulation. 2006;113:1424-1433.

General discussion



General discussion

Heart failure is on the increase. The latest incidence and prevalence figures for the Netherlands date from 2007: approximately 120,200 patients were known with heart failure while another 39,400 patients were newly diagnosed with heart failure. The good news is that evidence based treatments are available for patients with heart failure and a greater benefit is to be expected when the disease is diagnosed early so that treatment can start promptly.[1] The bad news is that heart failure is a common, costly, disabling, and potentially deadly condition.[2] In the Netherlands for the year 2010 heart failure was the principal discharge diagnosis in almost 30,000 hospital admissions (1.8 per 1,000 inhabitants) and for 6,424 patients it was the principal cause of death.[3] An individual of 40 years of age has a 20% risk of developing heart failure in the remainder of his or her life.[4] Once diagnosed with heart failure the outlook is depressing with an overall death rate of more than 10% each year. Once hospitalized for heart failure, within one year 40% of patients die or are readmitted.[1] Diagnosing heart failure in primary care is a challenging task.[5] Patients usually present themselves to the primary care physician in an early phase of the disease when signs and symptoms are 'non-specific' (i.e. signs and symptoms are still mild and not characteristic for heart failure). For a typical Dutch general practitioner (GP), who provides care for 2,000 patients, their practice will have around 30 patients with established heart failure and each year 10 new cases of heart failure will be diagnosed. Not all of these 10 new cases will be patients presenting themselves initially to the GP with signs and symptoms of early phase slow-onset heart failure; a minority will have acute heart failure with an initial presentation at the hospital emergency room. On the other hand, the number of patients suspected of heart failure is likely to be at least 4 times as high as the number of patients diagnosed with incident heart failure.[6]

For diagnosing heart failure in primary care "Simple and reliable diagnostic procedures are very important for primary care physicians, who are responsible for the early diagnosis of heart failure and implementation of adequate treatment in most healthcare systems".[5] Unfortunately, however, signs and symptoms are often unreliable because many other conditions or treatments can exert heart failure like symptoms, such as dyspnea, fatigue and peripheral oedema, with other signs much less common and harder to detect, such as a third heart sound.[7] Indeed, very low univariable positive predictive values of solitary elements of the physical examination have been published.[8] As a consequence, earlier studies have reported that primary care physicians often over-diagnose heart failure (34% false-positive diagnoses) and also under-diagnosis is believed to be common, although adequate estimates are missing.[9, 10] This has led to discussions whether GPs can actually diagnose or exclude heart failure and whether perhaps all

suspected patients should be referred for to a cardiologist or specialist for diagnostic work-up.[11-13]

diagnostic research

Given all these circumstances, diagnostic research for heart failure should be on par with intervention research assessing the effect of therapies on prognosis. Unfortunately, the available evidence from diagnostic research is scarce compared to the number of randomized trials on treatment effects. For example, three key clinical trials assessing the effect of beta-blocker treatment randomized nearly 9,000 heart failure patients, followed by trials in specific sub-areas such as elderly patients. This resulted in a class I recommendation with level of evidence A in the European Society of Cardiology (ESC) 2008 guideline.[1] Also, the relative expensive device treatment by means of cardiac resynchronization therapy with defibrillator function holds the 'Class I - Level A' recommendation. Indicative for the subordinate position of diagnostic research in this field is that neither class of recommendation nor level of evidence is given for diagnostic strategies. This is partly attributable to the difficulty of grading the quality of diagnostic studies, but also the relatively low quality of many diagnostic studies plays a role.

Diagnostic research is intended to advance the scientific knowledge of the diagnostic process in clinical practice. The structure of a diagnostic study is straightforward in that it should follow clinical practice. First, the domain is typically patients suspected of the disease, in whom there is diagnostic uncertainty. Second, the process of diagnosis involves usually multiple tests (multivariable) that have a clear hierarchical order, since simple non-burdening tests are performed before costly burdening tests. Third, the goal is to predict the presence of the disease, therefore the 'end-product' of the diagnostic process is a probability of presence of the disease. Finally the decision making process will judge the predicted probability to be of sufficient certainty to consider the disease present or absent, or insufficient certainty is still present requiring a subsequent diagnostic step (including referral or watchful waiting). When the interest lies in one specific diagnostic test, the added value of this diagnostic tests on top of already available information is the relevant measure.[14]

Contrary to common believe, diagnostic research does not suffer from confounding since it does not address the causal explanation of the disease, therefore the multivariable approach is used for the sake of precision and the search for independent predictors.

Also of note is that, once the diagnostic test result is available, there is no need for the physician to know how to calculate the post-test probability (with archetypal terms as sensitivity and specificity), since now there is a new and better estimate of the probability of the disease being present, based on application of the regression formula using multiple diagnostic items.[15]

measure of diagnostic value

In this thesis diagnostic strategies, including the use of biomarkers such as (NT-pro)BNP, is addressed in the domain of patients suspected of slow-onset heart failure by their primary care physician. One of the measures of 'diagnostic value' we use is the area under the receiver operating characteristic (ROC) curve, also known as concordance (or c-) statistic. The c-statistic is a non-parametric measure of discrimination, defined as the probability that a test result will be higher in a patient with the disease than a patient without the disease.[16] Without any information supplied by the test this probability is 0.5, and a perfect diagnostic test will have perfect discrimination and thus a c-statistic of 1. It is obvious the c-statistic has no easy interpretation in clinical practice, since it is unlikely that two patients have a consultation of whom the doctor already knows that one patient has the disease and the other does not. On the other hand a better measure of diagnostic value proves hard to find.[17] New measures of appraising the diagnostic value of a test introduce decision making into the equation by means of classifying a patient as diseased or not or intermediate. Much debate is going on about these so called reclassification indices.[18 (with comments)] In the main analysis of the Utrecht Heart Failure Organisation – Initial Assessment (UHFO-IA) study we calculated both. Taking the value of a test one step further, thereby leaving the realm of diagnostic research, the cost-effectiveness can be assessed. Then even more elements are introduced into the equation: the costs and utilities of the effects of available treatments.

diagnosis of heart failure

As to the diagnostic value of BNP, there are other considerations to be made not directly related to the study design. In the ESC guideline the definition of heart failure consists of three elements, the first two elements refer to symptoms and signs. In the third element, "Objective evidence of a structural or functional abnormality of the heart at rest", explicit reference is made to BNP as an example of objective evidence.[1] Consequently, one can theoretically diagnose heart failure with history taking, physical examination and the blood sample test. Daily practice is different because of several reasons. First, the generally accepted 'gold' standard for structural or functional abnormality of the heart at rest is the echocardiographic examination. Historically the echocardiogram is used to define heart failure as inclusion criterion for the trials for treatment of heart failure.

As a result evidence based treatment is applicable only to the subgroup of heart failure patients with certain echocardiographic abnormalities of the heart, notably a reduced ejection fraction of the left ventricle. Also, the echocardiogram has proved to be useful in finding the underlying pathologies leading to heart failure, of which some are treatable.[1] Most primary care physicians do not have direct access to an echocardiographic laboratory (yet). Moreover no clinical trials on treatment of heart failure have been performed with a BNP level as inclusion criterion. Therefore diagnosing heart failure in primary care without an echocardiogram cannot directly serve the purpose of finding indications for evidence based treatment. It should be noted that in order to exclude heart failure the ESC guideline does not recommend echocardiography when both the ECG and BNP values are normal. Sometimes situations occur where an echocardiogram will not be obtained because additional information provided by the echocardiogram is believed not to outweigh the extra effort for a typically elderly and frail patient. Another situation is when an echocardiographic laboratory is simply not available as is often the case in low and middle income countries. Under these circumstances practitioners still need to make decisions, to treat is as much as a decision as to refrain from therapy. The studies in this thesis show that the added diagnostic value of NT-proBNP in patients suspected of heart failure in primary care is considerable; e.g. the c-statistic increased from 0.79 to 0.92 (chapter 2). In Chapter 3 similar results were found, the c-statistic increased from 0.83 to 0.86 and independent datasets confirmed these findings. Following the view that diagnostic testing is about predicting the presence of the disease, we developed a prediction rule (predicting the presence of heart failure) taking into account the age of the patient, a history of objective coronary artery disease, the use of a loop diuretic, and on physical examination a displaced apex beat, lung rales, irregular pulse, pulse rate, heart murmur suggesting mitral regurgitation and elevated jugular venous pressure and finally the blood sample NT-proBNP result. Applying the prediction rule makes the setting of thresholds explicit and consequently makes the uncertainty about the diagnosis quantifiable. For instance, setting the threshold of the rule at 10% or below indicating no heart failure and at 80% or above indicating heart failure present, leaves the physician with almost 60% of his/her patients in the ‘gray’ zone. If the purpose is efficient referral, this means that 40% of all patients suspected of heart failure will be referred for a primary diagnostic echocardiogram while the patients with more than 80% probability of heart failure (10% of all patients suspected) could then be referred for a secondary diagnostic echocardiogram, i.e. looking for (treatable) underlying pathologies and setting the indication for specific treatments. At this stage 32.3% of patients suspected would not be referred for an echocardiogram, while 4.7% of these patients will have heart failure.

It would be interesting to know when these 'false negative' categorized patients will eventually receive the diagnosis of heart failure, because the personal relationship between patient and clinician and the longitudinal continuity of care are core competencies of general practice.[19]

Physical examination

The central part of this thesis is the UHFO-IA study as presented in chapter 3. The a priori goal of this study was to determine the value of diagnostic strategies available to the primary care physician, encompassing not only history taking and physical examination and NT-proBNP, but also regular blood sample analyses, ECG, chest X-ray and spirometry. As shown in Chapter 3 the addition of chest X-ray or ECG to the history taking and physical examination improved predictive ability. Therefore a diagnostic strategy including chest X-ray and or ECG will lead to more diagnostic certainty. One must realize that all this knowledge is conditional on the availability of the results from taking a history and performing a physical examination, which in itself holds a c-statistic of 0.83, according to our findings. Paradoxically, in a patient suspected of heart failure no items from the symptoms section yielded additional diagnostic value because almost all patients were breathless on exercise, fatigued, tired and/or complained of ankle swelling. Performing a physical examination was shown to be of crucial importance for diagnosing heart failure. Evidently, many doctors would argue that physical examination is a versatile asset in the arsenal of the diagnostician in any patient suspected of a disease, at least doctors are still being taught so, but it is hard to find empirical evidence for specific diseases as outlined for heart failure above. Also in other fields of medicine scarce evidence of the diagnostic value of physical examination is available, e.g. cruciate ligament lesions of the knee[20], pulmonary embolism[21], lumbosacral radiculopathy[22], deep vein thrombosis[23] and herpes zoster[24]. In the study on herpes zoster, the physical examination led to an adequate diagnosis without the need of additional (laboratory) tests. Possibly physical examination is taken for granted as being an important ritual (anthropologically it is about transformation[25]) or is considered too self-evident to study, like the utility of using a parachute when jumping out of an aeroplane[26]. In the UHFO-IA study we showed, however, that one cannot only clearly show that physical examination makes a difference, its contribution can also be explicitly quantified. Apparently, relatively large studies are required to identify physical examination items that, in combination, are instrumental in identifying or ruling out heart failure suspected patients, since earlier and (much) smaller studies than ours often failed to identify physical examination items that were indicative of heart failure. In our study specifically the pulse rate and regularity, lung rales, apex beat, distension of the jugular vein and heart murmur all proved to be of importance. In demonstrating the value of physical examination we are standing

on the shoulder of giants, e.g. those who invented the stethoscope, introduced percussion and discovered patterns in the physical examination to uncover diseases. Physical examination is rightly so an intricate part of medical practice, it must be taught in medical school and skills should be maintained afterwards. In this day and age we should continue studying the physical examination with the contribution of modern tools to acquire more scientific knowledge. However, as this thesis illustrates, we need large descriptive studies to document the important diagnostic value of physical examination.

References

1. 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.
2. McMurray JJV, Pfeffer MA. Heart failure. Lancet. 2005;365:1877-1889.
3. Rutten FJC, Poos MR, Engelfriet PR, eds. *Hoe vaak komt hartfalen voor en hoeveel mensen sterven eraan?* Bilthoven: RIVM; 30 januari 2012. Volksgezondheid Toekomst Verkenning, Nationaal Kompas Volksgezondheid. http://www.nationaalkompas.nl
4. Lloyd-Jones DM, Larson MG, Leip EP, Beiser A, D'Agostino RB, Kannel WB, Murabito JM, Vasan RS, Benjamin EJ, Levy D. Lifetime Risk for Developing Congestive Heart Failure: The Framingham Heart Study. Circulation. 2002;106:3068-3072.
5. Hobbs FDR, Doust J, Mant J, Cowie MR. Heart failure: Diagnosis of heart failure in primary care. Heart. 2010;96:1773-1777.
6. Cowie MR, Struthers AD, Wood DA, Coats AJS, 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.
7. Watson RD, Gibbs CR, Lip GYH. ABC of heart failure. Clinical features and complications. BMJ. 2000;320:236-239.
8. Davie AP, Francis CM, Caruana L, Sutherland GR, McMurray JJ. Assessing diagnosis in heart failure: which features are any use? QJM. 1997;90:335-339.

9. Murphy JJ, Bossingham CM. Open access echocardiography. General practitioners use echocardiography approximately. BMJ. 1995;311:325.
10. Remes J, Miettinen H, Reunanen A, Pyorala K. Validity of clinical diagnosis of heart failure in primary health care. European Heart Journal. 1991;12:315-321.
11. Berkin KE. Does everyone in heart failure need echocardiography? Age Ageing. 1999;28:421-422.
12. McIntyre HF. Heart failure in need of a diagnosis. Ontological fallacy in heart failure. BMJ. 2009;338:b1304.
13. Hobbs R. Can heart failure be diagnosed in primary care? BMJ. 2000;321:188-189.
14. Grobbee DE, Hoes AW. *Clinical epidemiology: principles, methods, and applications for clinical research*. Boston: Jones & Bartlett Publishers; 2008; Page 58-102.
15. Miettinen OS. *Epidemiological Research: Terms and Concepts*. Dordrecht: Springer; 2011; Page 10.
16. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29-36.
17. Cook NR. Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction. Circulation. 2007;115:928-935.
18. Pencina MJ, D'Agostino Sr RB, D'Agostino Jr RB, Vasan RS. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. Statistics in Medicine. 2008;27:157-172.
19. WONCA Europe. The European definition of general practice / family medicine. 2011. http://www.woncaeurope.org/Web documents/European Definition of family medicine/Version 2011/Definition 3rd ed 2011 with revised wonca tree.pdf
20. Wagemakers HP, Luijsterburg PA, Boks SS, Heintjes EM, Berger MY, Verhaar JA, Koes BW, Bierma-Zeinstra SM. Diagnostic accuracy of history taking and physical examination for assessing anterior cruciate ligament lesions of the knee in primary care. Arch Phys Med Rehabil. 2010;91:1452-1459.
21. Courtney DM, Kline JA, Kabrhel C, Moore CL, Smithline HA, Nordenholz KE, Richman PB, Plewa MC. Clinical features from the history and physical examination that predict the presence or absence of pulmonary embolism in symptomatic emergency department patients: results of a prospective, multicenter study. Ann Emerg Med. 2010;55:307-315 e301.
22. Coster S, de Bruijn SF, Tavy DL. Diagnostic value of history, physical examination and needle electromyography in diagnosing lumbosacral radiculopathy. J Neurol. 2010;257:332-337.

23. Toll DB, Oudega R, Vergouwe Y, Moons KG, Hoes AW. A new diagnostic rule for deep vein thrombosis: safety and efficiency in clinically relevant subgroups. Fam Pract. 2008;25:3-8.
24. Opstelten W, van Loon AM, Schuller M, van Wijck AJ, van Essen GA, Moons KG, Verheij TJ. Clinical diagnosis of herpes zoster in family practice. Ann Fam Med. 2007;5:305-309.
25. Verghese A. A doctor's touch. TED talk. Accessed: 1-3-2012. http://www.ted.com/talks/abraham_verghese_a_doctor_s_touch.html
26. Smith GCS, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. BMJ. 2003;327:1459-1461.

Summary

Summary

The beginning of the diagnostic phase is a patient suspected of the disease (which can be a disease, syndrome, as well as 'not otherwise specified'). The next part of the diagnostic phase is selecting tests (including additional items from history taking and physical examination) and interpreting the results of these tests conditional on the knowledge thus far. The diagnostic process involves sequential tests, which have a natural order. Starting with the information obtained from medical history taking (demographics, symptoms, past medical history, medication use), followed by physical examination. In the next step the diagnostician has to consider readily available laboratory test, such as blood sample tests , chest X-ray, ECG and pulmonary function tests. The end of the diagnostic phase is a decision based on the estimated probability of the disease, with one of three outcomes: sufficient certainty that the disease is present leading to targeted interventions (which can be no treatment), insufficient certainty leading to the next diagnostic phase (often implying referral) or sufficient certainty the disease is absent.[1]

The disease of interest in this thesis is heart failure, which is defined as a syndrome in which patients have the following features[2]:

- Symptoms typical of heart failure (breathlessness at rest or on exercise, fatigue, tiredness, ankle swelling) and
- Signs typical of heart failure (tachycardia, tachypnoea, pulmonary rales, pleural effusion, raised jugular venous pressure, peripheral oedema, hepatomegaly) and
- Objective evidence of a structural or functional abnormality of the heart at rest (cardiomegaly, third heart sound, cardiac murmurs, abnormality on the echocardiogram, raised natriuretic peptide concentration).

Specifically the interest lies in the slow onset (or non-acute) manifestation of the syndrome as opposed to acute heart failure, where urgent therapeutic interventions are warranted. Typically, patients with slow onset heart failure present themselves to the GP first, thus the initial diagnostic assessment is performed by the GP.

The aim of this thesis is to assess diagnostic strategies in patients suspected of heart failure in primary care by means of empirical diagnostic research. In the literature much attention has been paid to B-type Natriuretic Peptide (BNP, a blood sample measurement) as a diagnostic test for heart failure. It is nowadays an intricate part of the definition of heart failure (see above). In this thesis, BNP is used as a leverage to assess diagnostic strategies encompassing not only BNP but other diagnostic tests as well, with special emphasis on physical examination.

In **Chapter 1** the focus is on design issues of clinically relevant diagnostic research. With the emergence of novel diagnostic tests, e.g. point-of-care tests, clinically relevant empirical evidence is needed to assess whether such a test should be used in daily practice. With the example of the value of B-type natriuretic peptides (BNP) in the diagnostic assessment of suspected heart failure, we will discuss the major methodological issues crucial in diagnostic research; most notably the choice of the study population and the data analysis with a multivariable approach.

BNP have been studied extensively in the emergency care setting, but also several studies in the primary care are available. The usefulness of this test when applied in combination with other readily available tests is still not adequately addressed in the relevant patient domain, i.e. those who are clinically suspected of heart failure by their GP. Future diagnostic research in primary care should be targeted much more at answering the clinically relevant question 'Is it useful to add this (new) test to the other tests I usually perform, including history taking and physical examination, in patients I suspect of having a certain disease'.

In **Chapter 2** the example of the BNP test from chapter 1 is further explored in an individual patient data (IPD) meta-analysis making use of three previously published diagnostic studies. The IPD meta analysis was followed by external validation. The additional diagnostic yield of BNP above standard clinical information was compared to ECG and chest X-ray. Derivation was performed on two existing datasets from Hillingdon (n=127) and Rotterdam (n=149) while the UK Natriuretic Peptide Study (n=306) served as validation dataset. Included were patients with suspected heart failure referred to a rapid-access diagnostic outpatient clinic. Case definition was according to the ESC guideline. Logistic regression was used to assess discrimination (with the c-statistic) and calibration.

Of the 276 patients in the derivation set 30.8% had heart failure. The clinical model (encompassing age, gender, known coronary artery disease, diabetes, orthopnea, elevated jugular venous pressure, pulmonary crackles, pitting edema and S3 gallop) had a c-statistic of 0.79. Adding respectively chest X-ray, ECG or BNP to the clinical model increased the c-statistic to 0.84, 0.85 and 0.92. Neither ECG nor chest-X ray added significantly to the 'clinical plus BNP' model. All models had adequate calibration. The 'clinical plus BNP' diagnostic model performed well in the validation cohort with comparable inclusion criteria (c-statistic=0.91 and adequate calibration). Using separate cut off values for 'ruling in' (typically implying referral for echocardiography) and for 'ruling out' heart failure, creating a 'grey zone' was created resulting in insufficient proportions of patients with a correct diagnosis.

BNP has considerable diagnostic value in addition to signs and symptoms in patients suspected of heart failure in primary care. However, using BNP alone with the currently recommended cut off levels is not sufficient to make a reliable diagnosis of heart failure in individual patients.

Chapter 3 introduces the design and presents the major findings of the Utrecht Heart Failure Organisation - Initial Assessment (UHFO-IA) study. This is a cross-sectional diagnostic accuracy study with external validation. 721 consecutive patients suspected of new onset heart failure underwent standardized diagnostic work-up including chest X-ray, spirometry, ECG, 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. 207 (28.7%) patients 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 a 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 (see table).

This study estimated the quantitative diagnostic contribution of elements of the history and physical examination in the diagnosis of heart failure in primary care outpatients, and 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.

UHFO Diagnostic Rule for patients in primary care suspected of slow-onset heart failure

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 over 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

Instruction manual: summate all points gathered from the 10 parameters. The sum equals the probability (%) of the presence of heart failure.

In **Chapter 4** we explore the question whether it is feasible in the initial diagnostic assessment to further specify the diagnosis of heart failure by phenotype, i.e. heart failure with reduced left ventricular (LV) ejection fraction (HFrEF) and heart failure with preserved ejection fraction, sometimes called diastolic heart failure (DHF). Distinguishing between HFrEF and DHF in patients suspected of heart failure is crucial since different therapies are required. In settings without easy access to state-of-the-art echocardiography, the key tools to categorize these patients are symptoms, signs and simple laboratory tests. The purpose of this study was to identify diagnostic tools readily available to the primary care physician to help distinguish between both types of heart failure and no heart failure (no HF). Patients included in the UHFO-IA study had their final diagnosis of heart failure established by an expert panel, based on all available data (except NT-proBNP). When heart failure was judged to be present, the HFrEF or DHF phenotype was based on echocardiographic measurements. Multinomial logistic

regression was used to compute odds ratios (ORs). Of all 717 patients, 66 (9.2%) had DHF, 140 (19.5%) had HFrEF and 511 (71.3%) no HF. Female gender is indicative for DHF, male for HFrEF (OR=0.37 [0.19-0.71]), as expected, whereas age had an OR of 1.00, opposing to what was to be expected from the literature. Other statistically significant tools for the distinction DHF, HFrEF and no HF (jugular vein distension, cardio-thoracic ratio >0.50, an abnormal ECG and NT-proBNP) were all indicative of HFrEF more than DHF (ORs resp. 3.62, 4.49, 6.67, 1.63 (log(pg/ml))).

In this study no clear distinction could be made with readily available diagnostic tools. When the data point to severe symptoms it is probably HFrEF, whereas DHF is more likely when a sufficient number of mild signs and symptoms and laboratory tests compatible with heart failure are present. Notably age was not helpful in this setting.

The use of the furosemide treatment test as a diagnostic test in suspected heart failure seems popular among Dutch GPs, but the diagnostic value of this treatment test has not been assessed. In **Chapter 5**, in a sub-study of the UHFO-IA, the diagnostic value of the use of a short course of administering furosemide in trying to establish a diagnosis of slow-onset heart failure is determined. The furosemide test is not formally mentioned in the guidelines and no evidence could be found in the literature. We asked general practitioners (GPs) about their actual use of the furosemide test and studied the diagnostic accuracy in patients with suspected heart failure. General practitioners completed a questionnaire about their use of the furosemide test. We then performed a diagnostic accuracy study among a representative and consecutive sample of patients from the UHFO-IA study (see above). Forty of the 54 GPs had actually used the furosemide test in the past year and 70% considered the test to be useful. Forty seven patients underwent the furosemide test and 12 (25%) were diagnosed with heart failure. No effect of the test (weight loss, alleviation of symptoms) was significantly associated with heart failure.

We cannot support the use of the furosemide test as an ancillary diagnostic test in patients suspected of new, slow-onset heart failure.

In another sub-study of the UHFO-IA, presented in **Chapter 6**, three different automated assays which measure BNP or NT-proBNP in a blood sample are compared and the implications for clinical practice of the differences are assessed. The first 200 patients included in the UHFO-IA study (see above) had their NT-proBNP (on Elecsys®) and BNP (on Axsym® and Centaur® machines) measured in a single batch. Data were available for 172 patients; 51 had heart failure (29.7%). All 3 tests had high c-statistic values. An intermediate risk subset

of 111 patients (34% with heart failure) was created by excluding patients with very high or very low probability based on history and physical examination, the subgroup most in need of an additional test. Applying different thresholds for ruling heart failure in or out, the positive predicted values in this intermediate risk group were 75%, 76%, and 72%, respectively, and the negative predictive values 83%, 71%, and 85%. Applying the thresholds left 50% of patients 'in between'. In practice, a valid diagnosis in patients suspected of slow-onset heart failure remains elusive for many in the absence of echocardiographic imaging. No statistically significant difference was found between the 3 automated assays which measure BNP or NT-proBNP.

In **Chapter 7** the economic implications of several diagnostic strategies in suspected heart failure are explored. The full course of the patient is taken into account, starting from being suspected of the disease, being diagnosed timely or later, being treated and experiencing the vicious nature of heart failure until an often early death.

To provide evidence based treatment promptly to patients with heart failure requires early diagnosis. When signs and symptoms suggest heart failure the general practitioner (GP) may follow a variety of diagnostic strategies. Strategies may be dynamic and while a GP may think that the present signs and symptoms may not be sufficient to justify referral to a cardiologist, this may change when the patient presents his/herself with more advanced symptoms. Here, these strategies are assessed with emphasis on the balance between costs and effects.

In this economic analysis of the UHFO-IA study (721 patients with suspected heart failure) strategies were used to distinguish whom to treat immediately, whom to refer to the cardiologist and in whom to refrain from heart failure treatment. Seven strategies were compared: the diagnostic rule derived in this study (with four different sets of criteria to rule heart failure in or out), the GP decision, refer all patients to the cardiologist and do nothing and still obtain perfect diagnosis. A model was built to estimate effects, taking into account estimates on the condition of the patient (accuracy of the diagnosis, survival, indices on quality of life) and transition probabilities between conditions, as well as to estimate the costs, for which two methods were used. Cost estimation method 1 used risk ratios, e.g. patient in NYHA class II cost 1.83 times as much as patients in NYHA class I. The cost calculation used the risk ratio and time spent in each NYHA class. Cost estimation method 2 used the notion that costs are generated when patients deteriorate to NYHA class III or IV, when hospital admission is often the consequence, therefore the number of deteriorations determines the costs. The average UHFO-IA patient had an estimated 5.92 years to live (or 3.11 quality

adjusted life years (QALYs)) costing €2688 (method 1) or €2489 (method 2) when the (hypothetically) 'do nothing' strategy was to be followed.

Other strategies resulted in:

	Life years gained	QALYs gained	Method 1	Method 2
Refer all	6.76	3.57	€ 2,598	€ 2,172
<20% = do not treat >70% = do treat in between = refer	6.48	3.45	€ 2,584	€ 2,185
<10%; >70%	6.72	3.56	€ 2,581	€ 2,163
<20%; >80%	6.48	3.45	€ 2,583	€ 2,184
<10%; >80%	6.72	3.62	€ 2,537	€ 2,126
GP-guided	6.44	3.41	€ 2,728	€ 2,366

Conclusion: it is cost effective to apply the UHFO diagnostic rule to arrive at one of three outcomes: 1) heart failure probably not present (probability <10% or 20%); do not start treatment for heart failure, 2) heart failure probably present (probability >70% or 80%); start treatment for heart failure and 3) diagnostic uncertainty remains then refer for the cardiology consultation.

In the **general discussion** it is argued that diagnostic studies assessing the value of items from physical examination in suspected heart failure, but also in other suspected diseases, have been too small and often lacked a valid study design. As a consequence, these studies tend to discredit solitary elements of the physical examination. The plea is made to perform large descriptive studies in the relevant patient domain, optimally designed to study the value of diagnostic tests, including findings from physical examination.

References

1. Grobbee DE, Hoes AW. *Clinical epidemiology: principles, methods, and applications for clinical research*. Boston: Jones & Bartlett Publishers; 2008; Page 58-102.
2. 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.

Samenvatting

Samenvatting

Uit de **algemene inleiding**: mevrouw Ten Cate is een weduwe van 76 jaar oud; zij is bekend met hoge bloeddruk en diabetes, waarvoor zij respectievelijk chloortalidon en metformine voorgeschreven heeft gekregen, alsmede een dieet. Kortgeleden bezocht zij het spreekuur van haar huisarts omdat het steeds moeilijker voor haar werd om de trap in haar huis naar boven te lopen of boodschappen te doen zonder behoorlijk kortademig te worden. In de spreekkamer was ze nog steeds kortademig van de korte reis naar de praktijk. Ze loopt al geruime tijd met een stok vanwege artrose van haar linker heup en vanwege de pijn die dat geeft neemt ze regelmatig een NSAID. In het volgende gesprek (anamnese) wordt duidelijk dat ze geen andere pijn lijdt, geen koorts heeft en er waren geen medicatieveranderingen. Het was haar opgevallen dat ze minder eetlust had en toch de afgelopen paar weken 3 kilo was aangekomen. Er was ook een periode van hoesten, maar deze is bijna voorbij. Welhaast retorisch stelt mevrouw Ten Cate de vraag of haar leeftijd de schuld van dit ongemak is.

Als diagnosticus heeft haar huisarts reeds veel informatie ingewonnen, deels bewust, deels minder bewust, maar is dit voldoende om een diagnose te stellen? Daarbij kan worden aangetekend dat in feite niet de patiënt de diagnose heeft, maar de dokter.[1] In ieder geval zal deze vroeg diagnostische fase moeten eindigen met een conclusie met daarop gebaseerd een actie, ook al kan deze actie inhouden dat de patiënt wordt gerustgesteld omdat de ziekte niet aanwezig is of de actie kan zijn gecontroleerd afwachten. De conclusie is zelden een diagnose in de aristotelische zin van het woord, waar de determinatie van de ziekte bestaat uit het kiezen van een element uit een bestaande lijst. Op deze wijze zou diagnostiek kunnen worden onderzocht los van het domein van klinisch redeneren en handelen. In plaats daarvan wordt een diagnose gesteld omdat het een functie heeft bij de informatieverstrekking aan etiologie en prognose, alsmede de handelingen bedoeld om de prognose te veranderen: therapie en preventie.[2]

De start van de diagnostische fase is bij een patiënt die wordt verdacht van een ziekte (dat kan een specifieke ziekte zijn, een syndroom of zelfs een ‘niet nader gespecificeerd’ ziektebeeld). Vervolgens moeten de diagnostische testen worden gekozen (daaronder vallen ook aanvullende onderdelen uit de anamnese en lichamelijk onderzoek) en zullen de uitslagen moeten worden geïnterpreteerd in het licht van de reeds beschikbare informatie. Het diagnostisch proces bestaat uit opeenvolgende tests, die een natuurlijke orde te hebben. Het begint met informatie verkregen uit de voorgeschiedenis en anamnese (zoals leeftijd, geslacht, medicatiegebruik en klachten), gevolgd door lichamelijk onderzoek. In de volgende stap neemt de diagnosticus testen in overweging die makkelijk

beschikbaar zijn en weinig belastend, zoals bloedonderzoek, röntgenfoto van de thorax, ECG en longfunctieonderzoek. Het eindpunt van de diagnostische fase is een beslissing gebaseerd op de geschatte kans dat de gezochte ziekte aanwezig is en daaruit volgen drie uitkomsten: voldoende zekerheid dat de ziekte aanwezig is, dan volgt gerichte behandeling (voor sommige ziekten is dat niets doen), onvoldoende zekerheid over de aanwezigheid van de ziekte, dan volgt een volgende diagnostische fase (vaak verwijzing) of er is voldoende zekerheid dat de ziekte niet aanwezig is.[3]

In dit proefschrift is hartfalen de ziekte waarnaar wordt gezocht. Hartfalen wordt gedefinieerd als een syndroom waarbij patiënten de volgende kenmerken vertonen:[4]

- Symptomen passend bij hartfalen (bijvoorbeeld verminderde inspannings-tolerantie, zich veelal uitend in klachten van kortademigheid en vermoeidheid of perifeer oedeem) en
- Onderzoeksbevindingen passend bij hartfalen (bijvoorbeeld crepiteren van de longen, verhoogde centraal veneuze druk, perifeer oedeem, vergrote lever, heffende/verbrede ictus, hartgeruis, tachycardie, tachypnoe, 3^{e} harttoon) en
- Objectief bewijs voor een structurele of functionele afwijking van het hart in rust.

Specifiek ligt de interesse in dit proefschrift in de vorm van hartfalen die geleidelijk ontstaat, in tegenstelling tot de acute vorm, waar dringend behandeling is geïndiceerd. Patiënten met geleidelijk ontstaan hartfalen zullen zich doorgaans bij de huisarts presenteren en deze voert dus de initiële diagnostische beoordeling uit. Het doel van dit proefschrift is het beoordelen van diagnostische strategieën bij patiënten die door de huisarts worden verdacht van geleidelijk ontstaan hartfalen door middel van empirisch diagnostisch onderzoek.

De laatste tijd is in de literatuur veel aandacht besteed aan het B-type natriuretisch peptide (BNP, een eiwit meetbaar in het bloed) als een diagnostische test voor hartfalen. Recentelijk is in de Europese richtlijn een verhoogd BNP opgenomen als mogelijk bewijs voor structurele of functionele afwijking van het hart.[5]

In dit proefschrift is BNP het knooppunt om de diagnostische strategieën te beoordelen, waar naast BNP ook andere testen worden beoordeeld, met nadruk op het lichamelijk onderzoek.

Hoofdstuk 1 is gericht op vraagstukken die samenhangen met het ontwerpen van klinisch relevante diagnostische studies. Vier onderwerpen komen aan de orde die ook in het verdere verloop van het proefschrift een belangrijke rol spelen:

i) inclusie van de relevante patiënten (domein). Het is het meest relevant voor de dagelijkse klinische praktijk een test te gebruiken in die patiënten waar twijfel is over de diagnose. In diagnostische studies wordt vaak gebruik gemaakt van twee groepen patiënten, de ene groep heeft overduidelijk de ziekte en de andere groep heeft de ziekte overduidelijk niet. De nieuwe diagnostische test wordt beoordeeld naar het vermogen beide groepen uit elkaar te houden. In de praktijk echter is er geen diagnostische twijfel in beide groepen. Om Milton Packer te parafraseren: "het doel is niet om patiënten met overduidelijk hartfalen te onderscheiden van gezonde jonge vrijwilligers, maar om patiënten met hartfalen te onderscheiden van oudere patiënten met meerdere cardiovasculaire risicofactoren en aandoeningen, die echter geen hartfalen hebben".[6]
ii) het toepassen van de multivariabele aanpak. Diagnostiek gaat nooit over één test; immers leeftijd, geslacht, voorgeschiedenis, medicatiegebruik zijn meestal zelfs al bekend voor de patiënt binnenkomt.
iii) de natuurlijke hiërarchie van klinische diagnostische testen met de nadruk op de toegevoegde waarde van de nieuwe test. Een diagnostische test moet een toegevoegde waarde hebben op dat wat reeds bekend is voordat de test zou worden uitgevoerd. Een MRI onderzoek bijvoorbeeld wordt immers gedaan als anamnese, lichamelijk onderzoek, bloedonderzoek en andere minder ingrijpende diagnostische testen zijn gedaan en er nog steeds diagnostische twijfel is.
iv) maakt de studie het mogelijk dat kans op ziekte wordt voorspeld? Het heeft weinig zin om relatieve waarden te verkrijgen uit de studie, bijvoorbeeld mannen hebben een twee keer grotere kans op hartfalen dan vrouwen; net zo min heeft het zin om waarden te verkrijgen die een omrekening vereisen, zoals de zogenaamde predictieve waarden. Bijvoorbeeld een positief predictieve waarde van 90% wil zeggen dat als de diagnostische test positief is, dan is de kans op ziekte 90%. De omrekening zit in het feit dat deze maat direct afhankelijk is van de 'voorafkans' op ziekte. Ook maten als sensitiviteit, specificiteit en likelihood ratio hebben omrekeningen nodig om te komen tot de kans op ziekte voor deze patiënt. Deze redenering komt min of meer automatisch uit op een multivariabele formule, die rekening kan houden met de vele factoren die van invloed zijn op de kans op ziekte.

In dit hoofdstuk wordt BNP als test voor hartfalen gebruikt en dient als voorbeeld om de diagnostische waarde van een nieuwe test te beoordelen. Vanuit het oogpunt van de huisarts in de hoedanigheid van diagnosticus practicus wordt nagegaan hoe, in het licht van bovengenoemde punten, de toegevoegde diagnostische waarde van de nieuwe test kan worden beoordeeld met behulp van gepubliceerd onderzoek.

In **hoofdstuk 2** wordt het voorbeeld van de BNP verder uitgewerkt in een 'individual patient data' (IPD) meta-analyse. Daartoe wordt gebruik gemaakt van drie reeds gepubliceerde diagnostische studies: de Hillingdon studie (n=127) en de Rotterdam studie (n=149), terwijl de UK Natriuretic Peptide Study (n=306) diende als validatie. Het doel van deze studie was de toegevoegde diagnostische waarde van BNP te kwantificeren. Afgemeten aan de c-statistic (een maat voor discriminatie waarbij 0.5 gelijk het opgooien van een munt is en 1.0 perfect scheiding van ziek en niet ziek is) had het model met alleen direct beschikbare klinische parameters reeds een waarde van 0.79. Dit klinische model omvatte leeftijd, geslacht, reeds bekend met coronaire hartziekte, diabetes, orthopnoe, verhoogde centraal veneuze druk, crepiteren van de longen, oedemen en een 3^{e} harttoon. Het toevoegen aan dit model van een röntgen thoraxfoto, ECG en BNP leidde tot een c-statistic van respectievelijk 0.84, 0.85 en 0.92. Het 'klinisch plus BNP' model had ook in de validatie dataset een c-statistic van 0.91. Dit model werd toegepast om klinische beslissingen te nemen (wel behandelen, niet behandelen of nog steeds twijfel) met gebruik van aparte drempelwaarden, zoals aangeraden in de richtlijn, voor 'zeker ziek' en 'zeker niet ziek' en dus een tussengroep met nog steeds twijfel. Daarbij kwam een onaanvaardbaar hoog percentage patiënten in de tussengroep terecht of kreeg een verkeerde diagnose.

Hoofdstuk 3 introduceert de Utrecht Heart Failure Organisation - Initial Assessment (UHFO-IA) studie, het hoofdonderdeel van dit proefschrift. 721 opeenvolgende patiënten konden worden geïncludeerd in deze studie. Alle patiënten ondergingen dezelfde diagnostische testen (anamnese, lichamelijk onderzoek, röntgen thoraxfoto, spirometrie, ECG, NT-proBNP meting in bloed en echocardiografie) in een van de acht speciaal voor deze studie geformeerde sneldiagnostiek poliklinieken. De gouden standaard voor de diagnose hartfalen was de consensus van een expertpanel dat gebruik maakte van alle beschikbare informatie, inclusief echocardiografie en de ziektegeschiedenis van de zes maanden volgend op het bezoek aan de diagnostiekpolikliniek; wel was het expertpanel geblindeerd voor de NT-proBNP waarde. 207 (28.7%) patiënten bleken hartfalen te hebben. De combinatie van drie onderdelen uit de voorgeschiedenis (leeftijd, reeds bekend met coronaire hartziekte en gebruik van een lisdiureticum) plus zes onderdelen van het lichamelijk onderzoek (polsfrequentie en regelmaat, crepiteren van de longen, verhoogde centraal veneuze druk, heffende/verbrede ictus en hartgeruis passend bij mitralisinsufficientie) gaf een diagnostische waarde, afgemeten aan de c-statistic, van 0.83. NT-proBNP was de test met de meeste toegevoegde waarde in het volgende hiërarchische niveau, met een bereikte c-statistic van 0.86. Afgemeten aan de 'net reclassification improvement' (NRI; hoeveel patiënten gaan richting de gewenste 0% kans op ziekte als ze de ziekte niet hebben en hoeveel patiënten

gaan richting 100% als ze de ziekte wel hebben, minus het aantal patiënten dat de niet gewenste tegenovergestelde weg ging) gaf de toevoeging van NT-proBNP een NRI van 69%. Vervolgens werd een eenvoudige regel afgeleid uit het model en die werd gevalideerd in twee externe onderzoekspopulaties, resulterende in een c-statistic van 0.95 respectievelijk 0.88). De eenvoudige diagnostische regel met negen parameters kan een goed hulpmiddel voor de huisarts zijn in de dagelijkse praktijk om de kans in te schatten dat een patiënt hartfalen heeft.

UHFO diagnostische regel voor patiënten die door de huisarts worden verdacht van geleidelijk ontstaan hartfalen

score: som van de punten		**punten**
leeftijd (jaren)	<60	0
	60-70	4
	70-80	7
	>80	10
voorgeschiedenis van myocardinfarct, PCI of CABG	ja	15
gebruik lisdiureticum	ja	10
heffende/verbrede ictus	ja	20
crepiteren van de longen	ja	14
onregelmatig onregelmatige pols	ja	11
hartgeruis passend bij mitralisinsufficientie	ja	10
pols frequentie (min^{-1})		(boven 60 min^{-1}) / 3
verhoogde centraal veneuze druk	ja	12
NT-proBNP (pg/ml)	<100	0
	100-200	8
	200-400	16
	400-800	24
	800-1600	32
	1600-3200	40
	>3200	48

Gebruiksaanwijzing: tel de punten verkregen uit de 10 parameters op; deze som geeft de kans op hartfalen in procenten.

In **hoofdstuk 4** wordt verder ingegaan op de vraag of bij de initiële diagnostische beoordeling hartfalen reeds kan worden onderscheiden in de twee bestaande fenotypes: hartfalen met verminderde ejectiefractie (HFrEF) en hartfalen met intacte ejectiefractie, ook wel diastolisch hartfalen genoemd (DHF). Dit onderscheid is van groot belang, omdat voor HFrEF bewezen levensverlengende

en levenskwaliteit verbeterende therapieën bestaan en voor DHF (nog) niet. De onderzoeksvraag in deze secundaire analyse van de UHFO-IA studie was: “Is het mogelijk om patiënten verdacht voor hartfalen te onderscheiden in patiënten zonder hartfalen (no HF), DHF en HFrEF, tijdens de initiële diagnostische fase met behulp van diagnostische testen direct beschikbaar voor de huisarts?” Dus zonder gebruik te maken van een specialistisch echocardiografie laboratorium.

In de UHFO-IA studie werden, zoals boven beschreven, alle patiënten beoordeeld door een expertpanel op aanwezigheid van hartfalen, maar daarnaast werd ook bij iedere patiënt die de diagnose hartfalen toegewezen kreeg, het onderscheid HFrEF en DHF gemaakt op basis van alle beschikbare informatie, inclusief (en voornamelijk op basis van) echocardiogram, maar zonder de wetenschap van de NT-proBNP. Bij 4 van de 721 patiënten kon het echocardiogram geen informatie geven over de aanwezigheid van DHF; deze patiënten werden uitgesloten. De analyse bestond uit een multinomiale logistische regressie, met no HF, DHF of HFrEF als uitkomst. Met behulp van deze analysemethode werden odds ratios (ORs) berekend. De verdeling van diagnosen was dat 66 (9.2%) patiënten DHF hadden, 140 (19.5%) HFrEF en 511 (71.3%) no HF. Betreffende het onderscheid HFrEF/DHF wijst vrouwelijk geslacht naar DHF en dus mannelijk naar HFrEF (OR=0.37 [0.19-0.71]), dit was volgens verwachting uit de beschikbare literatuur; terwijl leeftijd in de literatuur een risicofactor is voor DHF, kwam dat in deze patiëntgroep niet uit: OR = 1.00. De statistisch significante parameters in het onderscheid no HF, DHF en HFrEF waren verhoogde centraal veneuze druk, cor-thorax ratio >0.50 op de X-thorax, een afwijkend ECG (t.o.v. een niet afwijkend ECG) en NT-proBNP. Al deze parameters wezen naar HFrEF, met ORs voor HFrEF/DHF respectievelijk 3.62, 4.49, 6.67 en 1.63 (de laatste per log(pg/ml). De conclusie was dat het gezochte onderscheid tijdens de initiële diagnostische fase niet goed te maken is zonder informatie uit het echocardiogram; het lijkt erop dat HFrEF gepaard gaat met ernstige symptomen en laboratorium utslagen en dat DHF gepaard gaat met meerdere milde symptomen en laboratorium uitslagen, waarbij leeftijd in dit stadium niet onderscheidend werkt.

Het gebruik van de furosemide behandeling als diagnostische test voor hartfalen lijkt populair onder Nederlandse huisartsen, maar de diagnostische waarde van deze test was nooit onderzocht. In **hoofdstuk 5** wordt de furosemide behandeling test onderzocht in 47 opeenvolgende patiënten uit de UHFO-IA studie, waarvan 12 met hartfalen. Overvulling is een belangrijk symptoom van hartfalen en de furosemide behandeling test sluit daarbij aan en dat vond ook een meerderheid van 54 huisartsen die werden gevraagd naar hun gebruik van de furosemide test. Zevenenzeventig procent van de ondervraagde huisartsen had de test gebruikt in het afgelopen jaar en 70% vond de test waardevol. Echter was met behulp van de furosemide test, afgemeten aan gewichtsverlies en/of symptoomverbetering geen

onderscheid te maken tussen patiënten met en zonder hartfalen. De furosemide test is dus niet bruikbaar als het patiënten betreft met de verdenking op langzaam ontstaan hartfalen.

In **hoofdstuk 6** worden bij de eerste 200 patiënten van de UHFO-IA studie drie veelgebruikte BNP laboratorium meetmethoden vergeleken. In één laboratorium-meetronde werden het NT-proBNP op de Elecsys® en BNP op AxSYM® en Centaur® apparaten gemeten in plasma dat was ingevroren. Gegevens waren bruikbaar van 172 patiënten waarvan 51 (26.7%) hartfalen hadden volgens het expertpanel. Alle drie meetmethoden hadden hoge c-statistic waarden. In een volgende analyse werden patiënten geselecteerd met een intermediaire kans op hartfalen waarbij de twijfel waarschijnlijk het grootst zou zijn geweest. Een subgroep van 111 patiënten werd geformeerd (34% met hartfalen) door patiënten met een lage kans of hoge kans op hartfalen uit te sluiten op basis van diagnostische testen vóór laboratoriumtesten worden ingeschakeld en daarmee de patiënten te identificeren waarvoor laboratorium onderzoek het meest toepasselijk is. Door gebruik te maken van verschillende drempelwaarden voor het aantonen en uitsluiten van hartfalen werden in deze intermediaire groep positief voorspellende waarden gevonden van 75%, 76%, and 72%, respectievelijk en negatief voorspellende waarden van 83%, 71%, and 85%. Echter door de verschillende drempelwaarden voor aantonen en uitsluiten, werd een 'grijze' tussenzone gecreëerd ter grootte van 55 patiënten met 18% kans op hartfalen en daarmee is de kans op een goede diagnose voor een individuele patiënt te klein om op deze wijze in de dagelijkse praktijk te worden gebruikt. Tussen de drie meetmethoden werd geen significant verschil gevonden.

In **hoofdstuk 7** wordt de kosteneffectiviteit onderzocht van verschillende diagnostische strategieën die de huisarts kan volgen om te onderscheiden wie direct te behandelen, wie wordt verwezen voor nadere diagnostiek naar de cardioloog en wie de behandeling voor hartfalen wordt onthouden. Zeven strategieën werden vergeleken: iedereen verwijzen en aldus perfecte diagnostiek; vier strategieën werden direct afgeleid van de UHFO diagnostische regel (zie boven) waarbij de variatie komt van het stellen van verschillende drempelwaarden voor de kans op hartfalen voor direct behandelen en behandeling onthouden; de door de insturend huisarts voorgestelde strategie (die zij/hij waarschijnlijk zou volgen als de patiënt niet ingesloten was in de UHFO-IA studie); de zevende strategie is onwerkelijk en behelst geen diagnostiek uitvoeren en toch perfecte diagnostiek verkrijgen en is alleen ter vergelijking. Om de effecten te schatten werd gebruik gemaakt van een model dat rekening kon houden met gegevens als juistheid van de diagnose, overlevingsduur, indicatoren van kwaliteit van leven, alsmede de kans dat wordt overgegaan van de ene stadium van de ziekte (afgemeten aan NYHA klasse) naar de andere met alle

gevolgen van dien. Voor het schatten van de kosten werden twee methoden gevolgd: methode 1 gebruikte relatieve risico's zoals bv. de schatting dat een patiënt in NYHA klasse II 1.83 keer zoveel kost als een patiënt in NYHA klasse I. Bij het berekenen van de kosten werd de tijd betrokken die een patiënt doorbracht in een NYHA klasse. Methode 2 maakte gebruik van het besef dat kosten worden gegenereerd als de patiënt verslechtert naar NYHA klasse III of IV, de klassen die vaak gepaard gaan met ziekenhuisopname. Het aantal verplaatsingen is in dit model kostenbepalend. De gemiddelde UHFO-IA patiënt had volgens de berekeningen gemiddeld 5.92 jaren te leven (3.11 jaren als werd gecorrigeerd voor kwaliteit van leven (QALY)) hetgeen volgens methode 1 €2,688 kostte en volgens methode 2 €2,489 wanneer de 'niets doen met perfecte diagnose' strategie werd gevolgd. De andere strategieën resulteerden in het volgende:

	jaren te leven	QALYs te leven	methode 1	methode 2
iedereen verwijzen	6.76	3.57	€ 2,598	€ 2,172
<20%* = niet behandelen				
>70% = behandelen	6.48	3.45	€ 2,584	€ 2,185
ertussen = verwijzen				
<10%; >70%	6.72	3.56	€ 2,581	€ 2,163
<20%; >80%	6.48	3.45	€ 2,583	€ 2,184
<10%; >80%	6.72	3.62	€ 2,537	€ 2,126
voorstel huisarts	6.44	3.41	€ 2,728	€ 2,366

* percentages betekenen kans op hartfalen en zijn berekend uit de UHFO diagnostische regel als boven beschreven

Concluderend kan worden gezegd dat het kosteneffectief is om diagnostiek volgens de UHFO diagnostische regel te verrichten om tot één van drie uitkomsten te geraken: 1) hartfalen is waarschijnlijk niet (kans <10% of 20%) aanwezig, dan niet behandelen 2) hartfalen is waarschijnlijk wel (kans >70% of 80%) aanwezig, dan starten met gerichte therapie en 3) diagnostische twijfel blijft bestaan, dan verwijzen naar cardioloog.

In de **algemene discussie** wordt beargumenteerd dat studies die de diagnostische waarde van geïsoleerde onderdelen van het lichamelijk onderzoek beoordelen bij patiënten verdacht van hartfalen (maar ook andere ziekten) te klein zijn geweest en vaak niet goed zijn opgezet. Daarom brengen deze studies lichamelijk onderzoek in diskrediet. Het pleidooi wordt gehouden om grote descriptieve studies uit te voeren in het relevante patiëntendomein, met een optimaal ontwerp om de waarde te kunnen beoordelen van diagnostische testen, inclusief lichamelijk onderzoek.

Literatuurreferenties

1. Miettinen OS. *Epidemiological Research: Terms and Concepts*. Dordrecht: Springer; 2011; Pag 10.
2. Whitbeck C. What is diagnosis? Some critical reflections. Theoretical Medicine and Bioethics. 1981;2:319-329.
3. Grobbee DE, Hoes AW. *Clinical epidemiology: principles, methods, and applications for clinical research*. Boston: Jones & Bartlett Publishers; 2008; Pag 58-102.
4. Hoes AW, Voors AA, Rutten FH, Van Lieshout J, Janssen PGH, Walma EP. Multidisciplinaire richtlijn hartfalen. Huisarts en Wetenschap. 2010;53:368-389.
5. 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.
6. Packer M. Should B-Type Natriuretic Peptide Be Measured Routinely to Guide the Diagnosis and Management of Chronic Heart Failure? Circulation. 2003;108:2950-2953.

Dankwoord

Dankwoord

Als je, zoals ik, tien jaar bezig bent geweest met je promotie en daarvóór ook al lang bezig bent geweest een promotieonderzoek te vinden met het volgende eisenpakket: mooi aansluiten bij vorige en huidige werkzaamheden, tussen de eerste en tweede lijn, tussen huisarts en specialist en methodologisch interessant, dan kan ik u melden dat je een hoop dank bent verschuldigd.

De promotie heb ik gevonden bij Arno Hoes, die samen met o.a. Carl Moons, Rick Grobbee en Hans Grundmeijer (AMC Amsterdam, huisartsgeneeskunde), de ZON-MW Doelmatigheid subsidie heeft verkregen. Het hele promotietraject is uitgezet en gevolgd door Arno Hoes, Rick Grobbee en de immer motiverende Maarten-Jan Cramer,

mijn opleiders epidemiologie, vanuit Amsterdam (VU EMGO instituut) en Rotterdam (Thoraxcentrum Eramus, de KLEP afdeling) die de begeestering voor de klinisch wetenschappelijke methode aan mij overbrachten, enkele namen: Jan Tijssen, Jaap Deckers, Karin Meeter, Ale Algra (Hans, je komt hier om onderzoek te doen, niet de wereld te verbeteren),

huisartsopleiders die de begeestering voor huisartsgeneeskunde overbrachten: Hans Drost (Brunssum), Max Dirckx (Geleen),

afdeling R&D cardiologie van het Sint Antonius Ziekenhuis in Nieuwegein: Herre Kingma, Boudewijn Uppelschoten, Egbert Koomen, Mike Bosschaert,

medewerkers afdeling medische opleidingen St Antonius Ziekenhuis, locatie Nieuwegein,

maatschap cardiologie St Antonius Ziekenhuis Nieuwegein: Norbert van Hemel, Carl Ascoop, Thijs Plokker, Sjef Ernst, Egbert Bal, Jurriën ten Berg, Lucas Boersma, Frank Eefting, Jan van der Heyden, Wybren Jaarsma, Gijs Mast, Benno Rensing, Maarten-Jan Suttorp, Eric Wever, Maurits Wijffels, Marco Post,

Julius Centrum UMC Utrecht: Monique den Hartog, Curt Brugman, Inge Sikking, Karin Nijssen, Ben van Hout, Frans Rutten,

medewerkers polikliniek cardiologie / longziekten UMC Utrecht,

de longartsen en cardiologen van de UHFO-IA studie expertpanel bijeenkomsten, met o.a. Jan Lammers,

laboratoria (klinisch chemie, echocardiografie/hartfunctie en longfunctie) van de participerende centra van de UHFO-IA studie en ook het huisartsenlaboratorium SALTRO voor het belangeloos opslaan en analyseren van alle monsters,

boden St Antonius Ziekenhuis locatie Nieuwegein, die geheel belangeloos bloed vervoerden van het UMC Utrecht naar St Antonius Ziekenhuis,

deelnemende centra UHFO-IA studie: naast St Antonius Ziekenhuis Nieuwegein en UMC Utrecht, Deventer Ziekenhuis: Jan van Wijngaarden, St Antonius Ziekenhuis locatie Utrecht: Rob van Tooren, Meander Medisch Centrum Amersfoort: Arend Mosterd, Onze Lieve Vrouwe Gasthuis Amsterdam: Peter Landsaat†, Catharina Ziekenhuis Eindhoven: Hans Bonnier, Isala Klinieken Zwolle: Ed de Kluiver,

Martin Cowie and collegues from the United Kingdom, for sharing their datasets and co-authoring,

de patiënten die hebben geparticipeerd aan de UHFO-IA studie, alsmede hun verwijzende huisartsen,

hartfalenpolikliniek: Esther van der Perk, Ria de Vries, Tonny van Geffen, Aliënde van Goor, Esther de Haan, Joke van den Berg, Anita van Oostrom, Linda Corsten,

uit de buitencategorie: Jacob Six, Freddy Vermeulen, Frank Cox, Pascal van Dessel, Monique Stofmeel, Leo Muller, Nicolien Breet,

tezamen met vele andere niet bij naam genoemde personen hebben zij buitengewoon bijgedragen aan dit proefschrift,

speciaal wil ik noemen mijn schoonouders Hanny† en Karel Loth, mijn paranimfen Kees-Jan Kelder, kleine broer en Joke Helwig, collega van het eerste uur, mijn ouders, Puck en (ouwe) Hans Kelder, Emma (superdochter) en Chris (voor het leven)

DANK •

Curriculum Vitae

Curriculum Vitae

Johannes (Hans) Christiaan Kelder was born on 16 October 1958 in Bussum, the Netherlands. In 1977 he graduated from the Gooisch Lyceum (secondary school, VWO) in Bussum. He did his medical studies at the University of Amsterdam and obtained his medical degree in 1985. His training in epidemiologic research was conducted at the EMGO institute, Free University, Amsterdam and he is board certified as epidemiologist since 1993. Later on he was trained general practice / family practice at the State University of Limburg.

Since 1993 he is working at the St Antonius Hospital in Nieuwegein, department of cardiology as clinical epidemiologist, research physician and physician for the heart failure outpatient clinic.

In 2002 he started his PhD work at the Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, supervised by professor A.W. Hoes and professor D.E. Grobbee and co-promotor M.J. Cramer, MD, PhD, consultant cardiologist at the department of cardiology of the University Medical Center Utrecht.

Hans is living with Chris Loth in Ruurlo and they have a daughter Emma Rosa (1998).