

Assessing the diagnostic test accuracy of natriuretic peptides and ECG in the diagnosis of left ventricular systolic dysfunction: a systematic review and meta-analysis

Clare Davenport, Elaine Yee Lan Cheng, Yip Tung Tony Kwok, Annie Hiu On Lai, Taka Wakabayashi, Chris Hyde and Martin Connock

ABSTRACT

Background

In 2003 the National Institute of Clinical Excellence published guidelines recommending the use of brain natriuretic peptide (BNP) and the electrocardiogram (ECG) as part of the diagnostic work up of individuals with heart failure. However, the guideline did not address whether one test was superior to the other or whether performing both tests was superior to performing single tests.

Aim

To investigate the relative test accuracy of the ECG, BNP, N terminal-pro brain natriuretic peptide (NT-proBNP) and combinations of two or more tests in the diagnosis of left ventricular systolic dysfunction (LVSD) in the primary care setting.

Design of study

Cohort studies making within-subject comparisons of intervention diagnostic test(s) with reference standard results.

Method

Standard systematic review methodology was followed.

Results

Thirty-two primary studies met the review inclusion criteria. Studies were of variable quality and highly clinically heterogeneous, therefore restricting the use of meta-analysis. Within these limitations BNP, NT-proBNP and the ECG all had similar test sensitivity (>80% in the majority of studies). Specificity of the three tests was not as good. Three studies directly comparing BNP and the ECG found no difference in sensitivity and limited support for improved specificity of BNP. Two studies found no difference in sensitivity and limited evidence for an improvement in specificity for the combination of the ECG and BNP compared to single tests.

Conclusion

On the basis of existing evidence, the ECG, BNP and NT-proBNP are useful in excluding a diagnosis of LVSD (good sensitivity). However, use of abnormal test results to select individuals for echocardiography may overwhelm services. There is currently no evidence to justify the use of one test over another or the use of tests in combination. The additional cost of BNP is not self-evidently justified by improved test accuracy. Further research is needed to directly compare the diagnostic performance of these tests in homogeneous, representative primary care populations.

Keywords

diagnostic tests; electrocardiography; natriuretic peptides; review, systematic.

INTRODUCTION

The crude prevalence of chronic heart failure is estimated to be between 0.4% and 3.2%.¹⁻¹⁰ The incidence and prevalence of heart failure in the population is rising mainly as a result of an ageing population and improved survival from the main aetiological cause, coronary heart disease.¹ The direct cost of health care for heart failure patients in the UK has recently been estimated at £716 million (1.83% of total NHS expenditure); the majority of costs are a result of hospitalisations.¹¹ Strategies to reduce hospitalisations (and particularly repeat hospitalisations) are best placed in primary care where the majority of the diagnosis and day-to-day management of heart failure occurs.

Left ventricular systolic dysfunction (LVSD) is one of the major underlying mechanisms causing chronic heart failure. However, recent evidence suggests that the drugs effective in reducing mortality and morbidity in this patient group continue to be under-prescribed and prescribed in sub-optimal doses.^{12,13} Diagnostic uncertainty in the primary care setting is argued to be a major cause

C Davenport, MB ChB, MSc, clinical research fellow; EYL Cheng, medical student; YTT Kwok, medical student; AHO Lai, medical student; T Wakabayashi, medical student; C Hyde, MB BS, senior clinical lecturer; M Connock, MSc, PhD, systematic reviewer, Department of Public Health and Epidemiology, University of Birmingham.

Address for correspondence

Dr Clare Davenport, Department of Public Health and Epidemiology, University of Birmingham, Edgbaston, Birmingham B15 2TT. E-mail: C.F.Davenport@bham.ac.uk

Submitted: 22 October 2003; Editor's response: 2 March 2004; final acceptance: 13 April 2005.

©British Journal of General Practice 2006; 56: 48-56.

of this inappropriate prescribing.¹⁴⁻¹⁷ Measurement of ventricular function is considered the reference standard for diagnosing LVSD¹⁸ due to the disparity that exists between ventricular function and associated signs and symptoms. However, primary care access to echocardiography is currently limited. The electrocardiogram (ECG), brain natriuretic peptide (BNP) and N terminal-pro brain natriuretic peptide (NT-proBNP) are being promoted as tests that may be used when echocardiography is not available, or in order to pre-select individuals with a suspected diagnosis of LVSD for further investigation with echocardiography.¹⁹⁻³⁰

Both the ECG and natriuretic peptides have a relatively high sensitivity (a 'normal' test result is good at ruling out a diagnosis of LVSD), but comparatively poor specificity with the potential to lead to considerable over-investigation of abnormal results. The National Institute of Clinical Excellence (NICE) published a guideline about the diagnosis and management of heart failure in 2003.²⁵ The NICE Guideline is ambiguous concerning whether natriuretic peptide testing alone, the ECG alone or a combination of natriuretic peptide testing and the ECG should be performed. Although this may reflect an inability to discriminate between the diagnostic accuracy of the three tests from the literature reviewed at the time of the Guideline's development, it has the potential to encourage indiscriminate and inefficient testing strategies.

METHOD

Comprehensive ascertainment was achieved through: (1) Electronic databases: (1980–March 2004) MEDLINE, EMBASE Cochrane Library 2003 Issue 4. Terms included a range of text words and MeSH terms concerning the condition of interest (suspected LVSD), the diagnostic tests being compared (BNP, NT-proBNP and the ECG) and the process (diagnosis). (2) Conference abstracts/hand-searching: The proceedings of the British Cardiac Society Annual Conferences 1980–March 2004 and the proceedings of the British Society for Heart Failure 1998 (inception)–March 2004 were hand-searched. The American Journal of Cardiology and the Journal of the American College of Cardiology (1980–2004) (the top two cardiology journals ranked according to the frequency with which diagnostic evaluation studies are published)³¹ were hand-searched for relevant articles. (3) Citation searches of identified included studies and reviews.

Study selection

Explicit, pre-determined inclusion and exclusion criteria were applied to abstracts or full articles of

How this fits in

The clinical diagnosis of left ventricular systolic dysfunction (LVSD) is difficult, and misdiagnosis results in inappropriate and sub-optimal treatment of patients. The National Institute of Clinical Excellence has recently published a diagnostic algorithm for the diagnosis of LVSD in primary care. However, the algorithm does not clarify whether brain natriuretic peptide (BNP), N terminal-pro brain natriuretic peptide (NT-proBNP), the electrocardiogram or a combination of these three tests should be used routinely in the diagnosis of LVSD. This review provides a comprehensive and up-to-date synthesis of the literature concerning the role of the natriuretic peptides and the electrocardiogram in the diagnosis of LVSD.

potential relevance to the review topic. Inclusion criteria were as follows:

Population. Adults suspected of having LVSD with or without comorbid conditions.

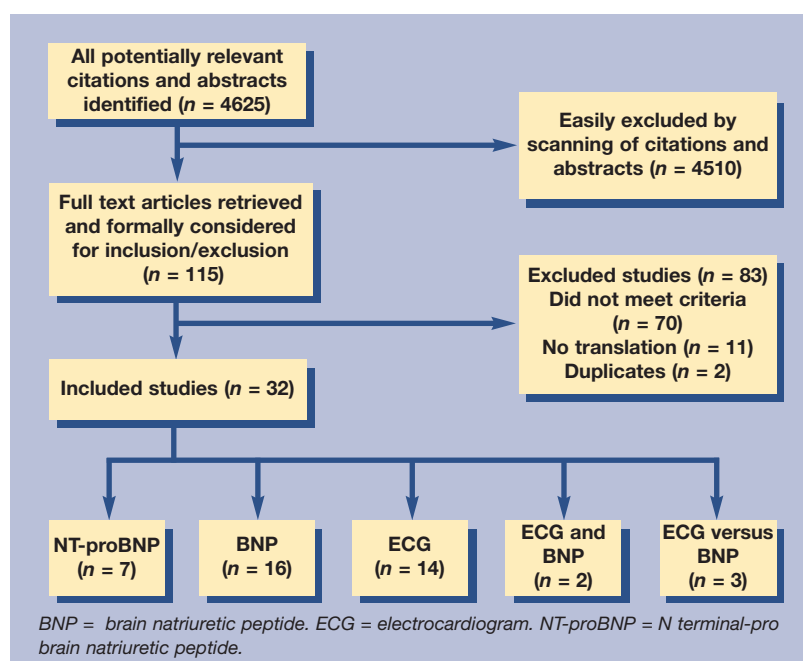
Intervention. One or more of the ECG, or the natriuretic peptides BNP and NT-proBNP.

Comparator. 'Gold' or reference standard for the diagnosis of LVSD (nuclear cardiology investigative techniques or 2-D echocardiography^{18,32-34} defining LVSD using a quantitative or qualitative measure of ejection fraction).

Study design. Cohort studies making within-subject comparisons of test results (interventions) with comparators in the same individuals.

Outcome measure. Derivation of a 2x2 diagnostic table in order to calculate test accuracy measures.

Figure 1. Summary of the study inclusion and exclusion process.



Inclusion and exclusion. See Figure 1. Abstracts and titles were initially scanned by one reviewer. Potentially relevant articles were further reviewed by at least two independent reviewers with disagreements resolved by a third reviewer according to the following exclusion criteria: studies with insufficient data to construct a 2x2 diagnostic table; studies concerned with the diagnosis of acute decompensated heart failure; studies concerned with the diagnosis of ventricular diastolic dysfunction alone; studies in which the majority of the target population had been on long-term treatment with ACE inhibitors and/or diuretics for presumed heart failure. Full details of the characteristics of included studies are presented in Supplementary Table 1.

Assessment of study quality and data extraction

Existing quality checklists^{35–37} were adapted to reflect the topic area and a pro-forma was used by two independent reviewers with disagreements resolved by a third reviewer. The criteria used for the quality assessment of included studies is outlined in Supplementary Table 2 and encompasses the domains of selection bias, verification bias, measurement bias and treatment paradox and disease progression bias. Data extraction was undertaken by two independent reviewers with uncertainty resolved by a third reviewer.

Data synthesis

Analysis was conducted using the Meta-DiSc software.³⁸ Data concerning disease spectrum (prevalence of LVSD), the reference standard used and test accuracy measures (true positives, false positives, false negatives and true negatives) was initially extracted into a spreadsheet.

Where multiple thresholds were provided within a study the worst and best estimates of the diagnostic odds ratio (DOR) for the individual study were used to investigate heterogeneity and to derive pooled estimates of test accuracy where appropriate. This approach was used in the absence of an agreed method for pooling studies comprising a mix of single and multiple thresholds. Analysis proceeded with an investigation for the presence of diagnostic threshold³⁹ and subsequently an investigation for other sources of heterogeneity. The *a priori* hypotheses were that heterogeneity in observed test accuracy is likely to be affected by differences in the prevalence of LVSD (representing the probability of having LVSD prior to testing with either the ECG or the natriuretic peptides) and according to the reference

standard test employed encompassing the test (echocardiography or nuclear cardiology), whether there was a requirement for symptoms of heart failure to be present and whether measurement of ejection fraction was qualitative or quantitative.⁴⁰ Heterogeneity was assessed using both χ^2 and I^2 statistics in order to take into account the low power of the χ^2 test.^{41–42} Where $P < 0.01$ and I^2 was $\leq 50\%$ studies were considered sufficiently homogenous to proceed with pooling to derive a summary estimate of test accuracy. In the absence of both diagnostic threshold and other sources of heterogeneity pooled sensitivities and specificities were calculated according to the Der Simonian-Laird random effects model.⁴¹ In the presence of a diagnostic threshold effect, but an apparent absence of other sources of heterogeneity, summary sensitivity and specificity⁴³ were derived from summary ROC curves according to the Moses-Shapiro-Littenberg method.⁴⁴ Where significant heterogeneity persisted despite subgroup analysis, the range of sensitivities and specificities across included studies are presented as an indication of test accuracy.

RESULTS

Number of studies

Of 4625 potentially relevant citations and abstracts, 115 needed detailed scrutiny of their whole text to make an inclusion decision. Seventy studies were excluded on the basis of full publications: 27 studies not concerned with the accuracy of the natriuretic peptides or the ECG; five not using an appropriate reference standard; 24 not concerned with a population suspected of having chronic LVSD; one comprising solely of patients on long-term treatment (diuretics and ACE inhibitors); six not including sufficient data to construct a 2x2 diagnostic table and seven erroneously using linear regression to measure agreement between a test and the reference standard.⁴⁵

Twenty-nine primary studies and three posters reported in conference proceedings were included in the analysis^{46–77} ($n = 32$), of which 14 investigated the diagnostic accuracy of the ECG; 16 BNP; seven NT-proBNP, two BNP and the ECG combined and three directly comparing the ECG and BNP in the same study population. In some instances a single study investigated the diagnostic accuracy of more than one test.

Characteristics of included studies

See Supplementary Table 1. All studies were of cross-sectional design. Fourteen studies were conducted in the UK, eight in Europe and nine elsewhere.

Spectrum of study participants. The mean age of study participants ranged from 53–79 years. In 16 out of 32 studies mean age was not reported.

Eleven studies were conducted on primary care populations, eight studies on secondary care populations, five studies on patients post myocardial infarction (MI) and eight on populations constructed by a mixture of primary and secondary care physicians.

Forty-three per cent of ECG studies employed exclusion criteria that would affect diagnostic accuracy or they lacked sufficient detail on which to base a judgement. The corresponding figures for BNP were 69%; 71% for NT-proBNP and 100% for BNP and the ECG combined. As an illustration, in four out of 16 BNP studies, and one out of seven NT-proBNP studies, patients were excluded if they exhibited characteristics or morbidity common in patients with LVSD but which independently increase BNP levels (for example renal dysfunction, diuretic treatment or increasing age). This degree of selection of study samples is likely to alter test accuracy to a degree that precludes transferability to the primary care setting in practice and is reflected in part by the observed range of prevalence rates across included studies.

LVSD prevalence ranged between 4.5% and 83% in primary care settings, between 23% and 55% in secondary care settings, between 4.5% and 52% in populations constructed by a mixture of primary and secondary care physicians and between 36% and 48% in post-MI patients. The prevalence of LVSD in studies investigating the diagnostic accuracy of the ECG ranged between 4.5% and 83% (median = 24.5%); BNP 7% and 52% (median = 31.5%); NT-proBNP 6.7% and 48% (median = 27%) and ECG and BNP combined 10–52%. In two studies, LVSD prevalence was not reported.

Definition of an 'abnormal' test result (diagnostic threshold). ECG: five of the 14 studies investigating the diagnostic accuracy of the ECG alone or in combination with BNP relied on ECG interpretation by secondary care physicians, in four out of 14 studies ECG reporting was automated, and in one study ECG reporting was performed independently by primary and secondary care physicians. In four out of 14 studies details of ECG interpretation were not reported. The definition of an 'abnormal' ECG varied across studies although the majority (71%) used six or more ECG abnormalities to define 'abnormal'.

Natriuretic peptides: the definition of 'abnormal' varied widely in both BNP and NT-proBNP studies and ranged between 5 pmol/L and 49 pmol/L (median 16 pmol/L) for BNP studies and 5 pmol/L

and 250 pmol/L (median 31 pmol/L) for NT-proBNP studies. This variation is likely to reflect a desire on behalf of researchers to optimise test performance in each study population.

Reference (gold) standard: a range of reference standards were used for each of the ECG, BNP and NT-proBNP. These included nuclear cardiology, nuclear cardiology or echocardiography, echocardiography and symptoms of chronic heart failure, and echocardiography alone. In addition, studies using echocardiography employed a range of methods of measurement (quantitative and qualitative), and where ejection fraction was quantified studies employed a range of definitions of abnormal varying between 30% and 50%.

In summary, included studies varied widely in terms of the spectrum of patients being considered, the application of the index test and the reference test employed.

Quality assessment

Quality assessment of included studies is presented in Supplementary Table 2. Where an aspect of quality was not clear or not reported a conservative approach was taken and that quality component was assumed to be absent.

Sample size. No studies reported sample size calculations. Sample size of studies ranged between 83 and 14507 (median = 287).

Thirty-six per cent of ECG studies; 38% of BNP studies; 57% of NT-proBNP studies and one of two (50%) studies investigating the diagnostic accuracy of BNP and the ECG combined were of relatively poor quality as indicated by a total score of three out of a total of six criteria.

Selection bias. In this review selection bias is most likely to operate for the ECG (a test routinely available in contrast to the natriuretic peptides) where preferential forward referral of positive and indeterminate ECGs as opposed to normal ECG results will result in an over-estimation of sensitivity. Eighty-six per cent of the ECG studies and both of the studies investigating the accuracy of the ECG and BNP combined were judged as likely to have been affected by selection bias.

Availability of reference test result (verification bias). Overall, in 29% of ECG studies, individuals were excluded from the analysis of test accuracy because reference test results were unavailable (indeterminate or lost). Corresponding figures for BNP were 38%, NT-proBNP 43%, and for studies investigating the diagnostic accuracy of BNP and the ECG combined, 50%.

Outcome measurement. Blind measurement of outcomes (intervention and reference test results) is important to avoid overestimation of test accuracy. In 64% of ECG studies, outcome measurement was blind. The corresponding figure for BNP studies was 69%; NT-proBNP studies 71%; and for studies investigating the diagnostic accuracy of BNP and the ECG combined, 50%.

Repeatability. Seventy-nine per cent of ECG studies had methods that were judged to be repeatable. The corresponding figure for BNP was 81%, NT-proBNP 29%, and for studies investigating the diagnostic accuracy of BNP and the ECG combined, 100%.

Reliability. Twenty-one per cent of ECG studies provided estimates of reliability for either the index test and/or the reference test. The corresponding figure for BNP was 25%, NT-proBNP 43%, and for studies investigating the diagnostic accuracy of BNP and the ECG, 0%.

Timing of diagnostic test under investigation. Left ventricular function may change over time as a result of the natural progression of disease and fluctuations in disease severity (disease progression bias) and as a result of changes in treatment (treatment paradox). For 21% of ECG

studies the time interval between application of the diagnostic test under evaluation and the reference standard was greater than 1 day or unknown. The corresponding figure for BNP was 63%, for NT-proBNP 71%, and for studies investigating the diagnostic accuracy of BNP and the ECG combined, 0%.

In summary, the quality of included studies was variable. Verification bias in the natriuretic peptide studies in particular is likely to result in an overestimation of test accuracy while selection bias in the ECG studies in particular is likely to result in an overestimation of test sensitivity.

Diagnostic test accuracy

See Table 1, Supplementary Table 3, and Supplementary Figures 1–6.

Investigation for diagnostic threshold. Unsurprisingly, given the wide variation in threshold employed across BNP and particularly NT-proBNP studies, there was evidence of a significant threshold effect for the group of BNP studies ($P<0.03$) and for the group of NT-proBNP studies ($P<0.001$). By contrast the group of ECG studies did not demonstrate a significant threshold effect ($P<0.639$), which was surprising given the range of methods of interpretation employed across included studies.

Table 1. Within study comparisons of the test accuracy of the ECG and BNP.

Reference	n	ECG		BNP		ECG+BNP		ECG versus BNP		ECG+BNP versus ECG alone or BNP alone	
		Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Hutcheon <i>et al</i> 2002 ⁵⁸ (BNP cut-off 35pmol/l)	304	0.96 (0.83–1.0)	0.50 (0.44–0.56)	0.94 (0.79–1.0)	0.63 (0.57–0.69)	0.94 (0.79–0.99)	0.63 (0.57–0.69)	NSD	BNP higher specificity v ECG ($P<0.05$)	NSD	ECG+BNP v BNP: NSD. ECG+BNP higher specificity v ECG alone ($P<0.05$)
Hutcheon <i>et al</i> 2002 ⁵⁸ (BNP cut-off 49pmol/l)	304	0.96 (0.83–1.0)	0.50 (0.44–0.56)	0.84 (0.66–0.95)	0.69 (0.63–0.74)	0.84 (0.66–0.86)	0.69 (0.63–0.74)	NSD	BNP higher specificity v ECG ($P<0.05$)	NSD	ECG+BNP v BNP: NSD. ECG+BNP higher specificity v ECG alone ($P<0.05$)
Kruger <i>et al</i> 2004 ⁶¹ (BNP cut-off 23pmol/l)	66	0.84 (0.73–0.92)	0.63 (0.5–0.75)	0.89 (0.79–0.96)	0.56 (0.42–0.68)	0.77 (0.64–0.86)	0.83 (0.72–0.92)	NSD	NSD	NSD	ECG+BNP higher specificity v ECG alone or BNP alone ($P<0.05$)
Landray <i>et al</i> 2000 ⁶² (BNP cut-off 5pmol/l)	126	0.42 (0.26–0.58)	0.87 (0.78–0.93)	0.88 (0.73–0.96)	0.34 (0.24–0.45)	–	–	BNP higher specificity v ECG ($P<0.05$)	BNP higher specificity v ECG ($P<0.05$)	–	–

BNP = brain natriuretic peptide. ECG = electrocardiogram. NSD = no significant difference.

Indirect comparison of the ECG, BNP and NT-proBNP. See Supplementary Table 3 and Supplementary Figures 1–6. The sensitivity of the ECG in this review ranged from 41.5% (26.3–57.9%)⁶² with a corresponding specificity of 87% (78.3–93.4%)⁶² to 98.4% (94.2%–99.8%)⁷² with a corresponding specificity of 66.1% (58.6–73.0%).⁷² Supplementary Figure 1 shows that the majority of ECG studies demonstrate a point estimate of sensitivity >80%, while Supplementary Figure 2 indicates that the estimates of specificity appear less good and more heterogeneous, with the majority of studies demonstrating a specificity of <80%. Those studies with sensitivity estimates <80% were not remarkable with respect to any study characteristics presented in Supplementary Table 1 or in terms of study quality.

The pooled DOR for all ECG studies was highly heterogeneous ($P<0.000$). This heterogeneity did not decrease when sub-grouping studies according to variations in disease spectrum or variation in reference standards employed (see Supplementary Table 3). It was, therefore, not possible to derive a pooled summary estimate of sensitivity and specificity for the ECG or sub-groups of ECG studies.

The sensitivity of BNP in this review ranged from 20% (13.3–45.5%)⁶⁵ with a corresponding specificity of 89% (80–93.6%)⁶⁵ to 100% (86.8–100%)⁷¹ with a corresponding specificity of 47% (34–61%).⁷¹ Supplementary figures 3 and 4 show a similar pattern of test accuracy for BNP as for the ECG with the majority of BNP studies demonstrating a point estimate of sensitivity >80% with more heterogeneous and poorer estimates of specificity. Those studies with sensitivity estimates of <80% were performed on patients post MI but were not remarkable in terms of study quality.

The pooled DOR for all BNP studies was highly heterogeneous ($P<0.001$). Heterogeneity decreased when sub-grouping studies according to the reference standard employed (see Supplementary Table 3 for summary estimates of sensitivity and specificity). Sub-grouping studies according to disease spectrum (LVSD prevalence) did not consistently reduce heterogeneity and summary estimates are only available for an LVSD prevalence of <20%.

NT-proBNP. The sensitivity of NT-proBNP in this review ranged from 24.5%⁴⁶ (13.8–38.3%) with a corresponding specificity of 95% (92.2–97%)⁴⁶ to 98.1% (90.1–100%)⁴⁶ with a corresponding specificity of 23% (18.7–27.7%).⁴⁶ Supplementary Figures 5 and 6 show a similar pattern of test accuracy for NT-proBNP as for BNP and the ECG,

with the majority of BNP studies demonstrating a point estimate of sensitivity >80% with poorer and more heterogeneous estimates of specificity. Those studies with poorer sensitivity estimates were all studies with an NT-pro BNP cut-off ≥ 100 pmol/L but were not remarkable in terms of any other study characteristic or in terms of study quality.

The pooled DOR for all NT-pro BNP studies was highly heterogeneous ($P<0.006$). Heterogeneity was decreased when sub-grouping studies according to the reference standard used (Supplementary Table 3) but not according to LVSD prevalence.

Heterogeneity precluded indirect comparisons of test accuracy for the ECG, BNP and NT-proBNP. Although heterogeneity was reduced in BNP and NT-proBNP studies when grouped according to the reference standard test used, the small size of the sub-groups did not allow calculation of confidence intervals and thus it was not possible to determine if there was a significant difference in test accuracy between the two natriuretic peptide tests.

Direct comparison of the ECG with BNP (same patient population). Three studies allow the direct comparison of the ECG with one of the natriuretic peptides, BNP^{58,61,62} (Table 1). This eliminates sources of heterogeneity as comparisons between tests are within-study. However, each study has to be considered in isolation. Table 1 illustrates that in all three studies sensitivity did not differ between the ECG and BNP. Two of three studies directly comparing BNP and the ECG demonstrated an improved specificity with BNP and one of the studies showed no difference in specificity.

ECG combined with BNP compared to BNP alone and the ECG alone. Two studies allow comparison of the performance of the ECG and BNP combined compared to individual tests.^{58,61} The pooled DOR for the two studies was heterogeneous ($P>0.01$) and so it was not possible to drive a summary estimate of the test accuracy of the ECG and BNP combined. Both studies demonstrate an improvement in specificity for the combination of the ECG and BNP compared to the use of the ECG alone, but no improvement in sensitivity.

DISCUSSION

Clinical significance of findings

The NICE guidelines concerning the management of chronic heart failure²⁵ advise that the ECG or one of the natriuretic peptides (BNP or NT-proBNP) or a combination of natriuretic peptide testing and ECG testing should be employed as part of the diagnostic work up for individuals with suspected chronic heart failure.

Use of single tests. This review demonstrates that the ECG, BNP and NT-proBNP have good sensitivity but that it is not possible to distinguish between them using either direct or indirect comparisons on this measure of test performance. Estimates of specificity for the three tests are more heterogeneous but there is limited evidence that BNP may have superior specificity to the ECG. However, the clinical significance of any improvement in terms of reducing referrals for echocardiography has not been explored.

Use of a combination of tests. This review demonstrates that a combination of the ECG and BNP does not improve sensitivity. There is very limited evidence (two studies) for an improvement in specificity with a combination of the two tests, but again the clinical significance of this improvement has not been explored.

On the basis of existing evidence, it is, therefore, recommended that either a natriuretic peptide test (BNP or NT-proBNP) or the ECG should be used as part of the diagnostic work up of individuals with suspected chronic heart failure and that there is no evidence to justify the use of both tests. The choice of employing ECG or a natriuretic peptide will be affected by issues such as relative cost and availability. The estimated cost of an ECG is approximately £10 (personal communication, Finance Department, Northern General Hospital, Birmingham, August 2003) compared to approximately £20 for a natriuretic peptide test (personal communication, Roche Diagnostics and Bayer Diagnostics, August 2003).

Strengths and limitations of this review

This review represents an up-to-date and comprehensive review of primary research investigating the diagnostic accuracy of the natriuretic peptides (BNP and NT-proBNP) and the ECG in patients with suspected chronic heart failure. In addition, this review clearly sets out the clinical context in which existing studies of diagnostic test accuracy for these tests have been conducted and the associated problems when trying to synthesise primary research in this area as a result of clinical and methodological heterogeneity. Our chosen methods for synthesising this group of studies is based on the premise that heterogeneity in meta analysis is inevitable and should be quantified.⁷⁸ Our conclusions are that the degree of heterogeneity present in all but a few small sub-groups of our included studies would mean that pooling and the production of summary test accuracy estimates would be inappropriate and misleading.

Implications for further research

A substantial and recent body of work exists around the use of the natriuretic peptides and, to a lesser extent, the ECG in the diagnostic work up of individuals with chronic heart failure. However, the research is characterised by substantial heterogeneity and an absence of direct test comparisons that reduces its use in clinical practice. In addition, there is an absence of work investigating the test performance of current normal diagnostic practice for diagnosing LVSD in the primary care setting. It is, therefore, not possible to conclude whether the addition of new testing strategies, including use of the natriuretic peptides, would lead to an improvement in referral practices. Any further work with existing primary research would require the use of individual patient data. Variations in test accuracy with changes in clinical characteristics of patients or changes in the application of the diagnostic tests, such as the operator, the reference standard employed or the threshold used to define abnormality, could then be investigated.

Alternatively, further primary research is needed to directly compare the diagnostic performance of these tests in representative primary care populations and whether their use results in improved test accuracy compared to current diagnostic practice. It could be argued that there is a need for more diagnostic tests studies to be performed in the clinical setting in which they are to be used (or are being used) and to incorporate health outcomes, impact on referral practices and cost-effectiveness. Assessing diagnostic tests in isolation and out of clinical context has limited use past the assessment of a test's potential usefulness in phase I and II diagnostic study designs.⁷⁹

Further reading

Since this paper was written a further piece of related work has been completed by the QIS in Scotland. See: Craig J, Bradbury I, Cummins E, et al. *Health Technology Assessment report 6. The use of B-type natriuretic peptides (BNP and NT-proBNP) in the investigation of patients with suspected heart failure.* Edinburgh: NHS QIS, May 2005. It can be accessed at: http://www.nhshealthquality.org/nhsqis/qis_display_finding_s.jsp?pContentID=2456 (accessed 7 Dec 2005.)

Supplementary information

Additional information accompanies this article at <http://www.rcgp.org.uk/journal/supp/index.asp>

Competing interests

The authors have stated that there are none

Acknowledgements

Thank you to Sue Bayliss (information specialist) for help with devising the search strategy.

REFERENCES

1. Mair FS, Crowley TS, Bundred P. Prevalence, aetiology and management of heart failure in general practice. *Br J Gen Pract*

- 1996; **46**: 77–79.
2. Parameshwar J, Shackell MM, Richardson A, *et al*. Prevalence of heart failure in three general practices in north west London. *Br J Gen Pract* 1992; **42**: 287–289.
 3. Wheeldon NM, MacDonald TM, Flucker CJ, *et al*. Echocardiography in chronic heart failure in the community. *Q J Med* 1993; **86**(1): 17–23.
 4. Lamberts H, Brouwer HJ, Mohrs J. *Reason for encounter — episode- and process-oriented standard output from the transition project*. Amsterdam: Department of General Practice, 1990.
 5. Ambrosio GB, Riva L, Casiglia E. Prevalence of congestive heart failure (CHF) in elderly. A survey from a population of Veneto region. *Acta Cardiologica* 1994; **49**: 324–325.
 6. Clarke KW, Gray D, Hampton JR. The prevalence of heart failure estimated from prescription data. *Br Heart J* 1994; **71**(Suppl): 33.
 7. Ho KKL, Pinsky JL, Kannel WB, Levy D. The epidemiology of heart failure: the Framingham Study. *J Am Coll Cardiol* 1993; **22**(4 Suppl A): 6A–13A.
 8. Kannel WB, Kalon Ho, Thorn T. Changing epidemiological features of cardiac failure. *Br Heart J* 1994; **72**(2 Suppl): S3–9.
 9. McDonagh TA, Morrison CE, Lawrence A, *et al*. Symptomatic and asymptomatic left ventricular systolic dysfunction in an urban population. *Lancet* 1997; **350**(9081): 829–833.
 10. Davies M, Hobbs F, Davis R, *et al*. Prevalence of left-ventricular systolic dysfunction and heart failure in the Echocardiographic Heart of England Screening Study: a population based study. *Lancet* 2001; **358**: 439–444.
 11. Stewart S, Jenkins A, Buchan S, *et al*. The current cost of heart failure to the National Health Service in the UK. *Eur J Heart Fail* 2002; **4**(3): 361–371.
 12. Cleland JGF, Cohen-Solal J, Cosin Aguilar J, *et al*. Management of heart failure in primary care (the IMPROVEMENT of Heart Failure Programme): an international survey. *Lancet* 2002; **360**: 1631–1639.
 13. Luthi JC, Lund MJ, Sampietro-Colom L, *et al*. Readmissions and the quality of care in patients hospitalized with heart failure. *Int J Qual Health Care* 2003; **15**(5): 413–421.
 14. Fuat A, Hungin AP, Murphy JJ. Barriers to accurate diagnosis and effective management of heart failure in primary care: qualitative study. *BMJ* 2003; **326**: 196–202.
 15. Mair F. Do GPs underuse ACE inhibitors in heart failure? *Prescriber* 1996; **7**(12): 99–101.
 16. Mayor S. Integrate care for heart failure. *Medical Interface* 1996; **May**.
 17. Khunti K, Hearnshaw H, Baler R, Grimshaw G. Heart failure in primary care: qualitative study of current management and perceived obstacles to evidence-based diagnosis and management by general practitioners. *Eur J Heart Fail* 2002; **4**: 771–777.
 18. Naik MM, Diamond G, Pai T, *et al*. Correspondence of left ventricular ejection fraction determinations from two-dimensional echocardiography, radionuclide angiography and contrast cineangiography. *J Am Coll Cardiol* 1995; **25**(4): 937–942.
 19. Badgett RG, Lucey CR, Mulrow CD. Can the clinical examination diagnose left-sided heart failure in adults? *JAMA* 1997; **277**(21): 1712–1719.
 20. Badgett RG, Mulrow CD, Otto P, Ramirez G. How well can the chest radiograph diagnose left ventricular dysfunction? *J Gen Intern Med* 1996; **11**(10): 625–634.
 21. Khunti K, Baker R, Grimshaw G. Diagnosis of patients with chronic heart failure in primary care: usefulness of history, examination, and investigations. *Br J Gen Pract* 2000; **50**: 50–54.
 22. Cleland JG, Habib F. Assessment and diagnosis of heart failure. *J Intern Med* 1996; **239**(4): 317–325.
 23. Dargie H, McMurray JJV. Diagnosis and management of heart failure. *BMJ* 1994; **308**: 321–328.
 24. Davies M, Gibbs CR, Lipp GYH. ABC of heart failure: investigation. *BMJ* 2000; **320**: 297–300.
 25. National Collaborating Centre for Chronic Conditions on behalf of NICE. *Chronic heart failure. Management of chronic heart failure in adults in primary and secondary care*. Salisbury: Royal College of Physicians, 2003. <http://www.rcplondon.ac.uk/pubs/books/chf/heartfailure.pdf> (accessed 5 Dec 2005.)
 26. Centre for Health Services Research. *Evidence-based clinical practice guideline. ACE inhibitors in the primary care management of adults with heart failure*. Newcastle upon Tyne: University of Newcastle upon Tyne North of England Evidence Based Guideline Project, 1997. <http://www.nelh.nhs.uk/guidelinesdb/html/AFACEHF.HTM> (accessed 5 Dec 2005.)
 27. Scottish Intercollegiate Guidelines Network. *A National Clinical Guideline. Diagnosis and treatment of heart failure due to left ventricular systolic dysfunction*. Edinburgh: SIGN, 1999.
 28. Hunt SA, Baker DW, Chin MH, *et al*. ACC/AHA Guidelines for the Evaluation and Management of Chronic Heart Failure in the Adult: Executive Summary A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the 1995 Guidelines for the Evaluation and Management of Heart Failure): Developed in Collaboration With the International Society for Heart and Lung Transplantation; Endorsed by the Heart Failure Society of America. *Circulation* 2001; **104**(24): 2996–3007.
 29. National Heart Foundation of New Zealand. A guideline for the management of heart failure: health professionals guide. Auckland: National Heart Foundation of New Zealand, 2001. http://www.guideline.gov/summary/summary.aspx?ss=15&doc_id=3309&nbr=2535 (accessed 5 Dec 2005.)
 30. Remme WJ, Swedberg K, Task Force for the Diagnosis and Treatment of Chronic Heart Failure, European Society of Cardiology. Guidelines for the diagnosis and treatment of chronic heart failure. *Eur Heart J* 2001; **22**: 1527–1560.
 31. Ballaglia M, Bucher M, Egger M. The Bayes Library of diagnostic studies and reviews. 2nd edition, 2002. http://www.ispm.unibe.ch/files/file/261.Bayes_library_handbook.pdf (accessed 5 Dec 2005).
 32. British Society of Echocardiography Guidelines. Training in echocardiography. *Br Heart J* 1994; **71**(4 Suppl): 2–5.
 32. Sahn DJ, De Maria A, Kisslo J, Weyman A. Recommendations regarding quantitation in M-mode echocardiography: results of a survey of echocardiographic measurements. *Circulation* 1978; **58**: 1072–1083.
 34. Roelandt J, Gibson DG. Recommendations for standardization of measurements from M-mode echocardiograms. *Eur Heart J* 1980; **1**: 375–378.
 35. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1994; **271**(5): 389–391.
 36. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994; **271**(9): 703–707.
 37. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine*. 2nd edn. Boston: Little Brown, 1991.
 38. Zamora J, Muriel A, Abaira V. *Meta-DiSc for Windows: a software package for the meta-analysis of diagnostic tests*. Barcelona: XI Cochrane Colloquium, 2003. http://www.hrc.es/investigacion/metadisc_en.htm (accessed 5 Dec 2005.)
 39. Deville WL, Buntinx F, Bouter LM, *et al*. Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol* 2002; **2**: 9.
 40. McGowan J, Cleland JGF. Reliability of reporting left ventricular systolic function by echocardiography: a systematic review of 3 methods. *Am Heart J* 2003; **146**: 388–397.
 41. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. In: Chalmers I, Altman DG, Egger M (eds). *Systematic reviews in health care. Meta-analysis in context*. London: BMJ Books, 2001: 248–282.
 42. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21**(11): 1539–1558.
 43. Mitchell MD. Validation of the summary ROC for diagnostic test meta-analysis: a Monte Carlo simulation. *Acad Radiol* 2003; **10**(1): 25–31.
 44. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med* 1993; **12**(14): 1293–1316.

45. Altman DG. Chapter 14. Practical statistics for medical research. *Some common problems in medical research*. London: Chapman and Hall, 1991.
46. Alehagen U, Lindstedt G, Eriksson H, Kahlstrom U. Utility of the amino-terminal fragment of pro-brain natriuretic peptide in plasma for the evaluation of cardiac dysfunction in elderly patients in primary health care. *Clin Chem* 2003; **49**(8): 1337–1346.
47. Baker DW, Bahler RC, Finkelhor RS, Lauer MS. Screening for left ventricular systolic dysfunction among patients with risk factors for heart failure. *Am Heart J* 2003; **146**(4): 736–740.
48. Bettencourt P, Ferreira A, Dias P, et al. Evaluation of brain natriuretic peptide in the diagnosis of heart failure. *Cardiology* 2000; **93**: 19–25.
49. Byrne J, Morton JJ, Metcalfe MJ, Dargie HJ. Baseline Assessment of Plasma Brain Natriuretic Peptide and Simple Clinical Variables Predicting Long-Term Left Ventricular Function Following Acute Myocardial Infarction. *Heart* 1996; **75**(5 Suppl 1): S31.
50. Campbell DJ, Mitchelhill KI, Schlicht SM, Booth RJ. Plasma amino-terminal pro-brain natriuretic peptide: a novel approach to the diagnosis of cardiac dysfunction. *J J Card Fail* 2000; **6**(2): 130–139.
51. Choy AMJ, Darbar D, Lang CC, et al. Detection of Ventricular Dysfunction after Acute Myocardial Infarction: comparison of clinical, echocardiographic and neurohormonal methods. *Br Heart J* 1994; **72**(1): 16–22.
52. Cosin Aguilar J, Hernandez Martinez A, et al. Value of NTproBNP concentration in an out-of-hospital adult population. *Rev Esp Cardiol* 2003; **56**(3): 236–244.
53. Cowie MR, Struthers AD, Wood DA, et al. Value of natriuretic peptides in assessment of patients with possible new heart failure in primary care. *Lancet* 1997; **350**: 1347–1351.
54. Davie AP, Francis CM, Love MP, et al. Value of the electrocardiogram in identifying heart failure due to left ventricular systolic dysfunction. *BMJ* 1996; **312**: 222.
55. Fox KF, Cowie MR, Wood DA, et al. 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**(4): 423–429.
56. Hobbs FDR, Davis RC, Roalfe AK, et al. 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): 1498–1500.
57. Houghton AR, Sparrow NJ, Toms E, Cowley AJ. Should general practitioners use the electrocardiogram to select patients with suspected heart failure for echocardiography? *Int J Cardiol* 1997; **62**: 31–36.
58. Hutcheon SD, Gillespie ND, Struthers AD, McMurdo ME. B-type natriuretic peptide in the diagnosis of cardiac disease in elderly day hospital patients. *Age Ageing* 2002; **31**(4): 295–301.
59. Khandekar S, Bossingham CM, Murphey JJ. Open access echocardiography for suspected heart failure: would a 12 lead electrocardiogram be just as good. *Heart* 1996; **75**(5 Suppl 1): 119.
60. Krishnaswamy P, Lubien E, Clopton P, et al. Utility of B-natriuretic peptide levels in identifying patients with left ventricular systolic or diastolic dysfunction. *Am J Med* 2001; **111**: 274–279.
61. Kruger S, Filzmaier K, Graf J, et al. QRS prolongation on surface ECG and brain natriuretic peptide as indicators of left ventricular systolic dysfunction. *J Intern Med* 2004; **255**: 206–212.
62. 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.
63. Lindsay MM, Goodfield NER, Hogg KJ, Dunn FG. Optimising direct access ECHO referral in suspected heart failure. *Scott Med J* 2000; **45**(2): 43–44.
64. Mattleman SJ, Hakki A-Hamid, Iskandran AS, et al. Reliability of bedside evaluation in determining left ventricular function: correlation with left ventricular ejection fraction determined by radionuclide ventriculography. *J Am Coll Cardiol* 1983; **1**(2): 417–420.
65. McClure SJ, Caruana L, Davie AP, et al. Cohort study of plasma natriuretic peptides for identifying left ventricular systolic dysfunction in primary care. *BMJ* 1998; **317**: 516–519.
66. Mosterd A, de Bruijine MC, Hoes AW, et al. Usefulness of echocardiography in detecting left ventricular dysfunction in population-based studies (The Rotterdam Study). *Am J Cardiol* 1997; **79**: 103–104.
67. Neilson OW, Hansen JF, Hilden J, 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.
68. Neilson LS, Svanegaard J, Klitgaard NA, Egeblad H. N-terminal pro-brain natriuretic peptide for discriminating between cardiac and non-cardiac dyspnoea. *Eur J Heart Fail* 2004; **6**: 63–70.
69. Richards M, Nicholls MG, Yandle TG, et al. Plasma n-terminal pro-brain natriuretic peptide and adrenomedullin: new neurohormonal predictors of left ventricular function and prognosis after myocardial infarction. *Circulation* 1998; **97**: 1921–1929.
70. Rihal CS, Davis KB, Kennedy JW, Gersh BJ. The utility of clinical, electrocardiographic and roentgenographic variables in the prediction of left ventricular function. *Am J Cardiol* 1995; **75**: 220–223.
71. Sim V, Hampton D, Phillips C, et al. The use of brain natriuretic peptide as a screening test for left ventricular systolic dysfunction — cost-effectiveness in relation to open access echocardiography. *Fam Pract* 2003; **20**(5): 570–574.
72. Talreja D, Gruver C, Sklenar J, et al. Efficient utilisation of echocardiography for the assessment of left ventricular systolic function. *Am Heart J* 2000; **139**(3): 394–398.
73. Valli N, Georges A, Corcuff JB, et al. Assessment of brain natriuretic peptide in patients with suspected heart failure: comparison with radionuclide ventriculography data. *Clin Chim Acta* 2001; **306**(1–2): 19–26.
74. Wright SP, Doughty RN, Pearl A, et al. Plasma amino-terminal pro-brain natriuretic peptide and accuracy of heart-failure diagnosis in primary care. *J Am Coll Cardiol* 2003; **42**(10): 1793–1800.
75. Yamamoto K, Burnett JC, Jougasaki M, et al. Superiority of brain natriuretic peptide as a hormonal marker of ventricular systolic and diastolic dysfunction and ventricular hypertrophy. *Hypertension* 1996; **28**: 988–994.
76. Yamamoto K, Burnett J, Bermudez EA, et al. Clinical criteria and biochemical markers of the detection of systolic dysfunction. *J Cardiac Failure* 2000; **6**(3): 194–200.
77. Zaphiriou A, Robb S, Murray-Thomas T, et al. Ruling out heart failure with BNP or NTproBNP: does it work in practice. Results of the UK Natriuretic Peptide Study. *Heart* 2003; **89**(Suppl 1): A49.
78. Alderson P, Green S, Higgins JPT (eds). *Cochrane Reviewers' Handbook 4.2.2*. [updated March 2004]. <http://www.cochrane.org/resources/handbook/hbook.htm> (accessed 6 Dec 2005.)
79. Knottnerus JA, van Weel C. General introduction: evaluation of diagnostic procedures. In: Knottnerus JA (ed.). *The evidence base of clinical diagnosis*. London: BMA books, 2002: 1–16.