A Systematic Review of the Diagnostic Accuracy of Natriuretic Peptides for Heart Failure

Jenny A. Doust, BMBS, FRACGP; Paul P. Glasziou, FRACGP, PhD; Eva Pietrzak, PhD; Annette J. Dobson, PhD

**Background:** The diagnosis of heart failure is difficult, with both overdiagnosis and underdiagnosis occurring commonly in practice. Natriuretic peptides have been proposed as a possible test for assisting diagnosis. We assessed the diagnostic accuracy of brain natriuretic peptide (BNP), including a comparison with atrial natriuretic peptide (ANP).

**Methods:** Electronic searches were conducted of MEDLINE and EMBASE from January 1994 to December 2002 and handsearches of reference lists of included studies. We included studies that assessed the diagnostic accuracy of BNP against echocardiographic or clinical criteria or that compared the diagnostic accuracy of BNP with ANP. Two reviewers assessed studies for inclusion and quality and extracted the relevant data. A meta-analysis was performed by pooling the diagnostic odds ratios for studies that used a common reference standard.

**Results:** Twenty studies were included. For the 8 studies (n=4086) that measured BNP against the criterion of left ventricular ejection fraction of 40% or less (or equivalent), the pooled diagnostic odds ratio was 11.6 (95% confidence interval, 8.4-16.1). The pooled diagnostic odds ratio was greater, 30.9 (95% confidence interval, 27.0-35.4), in the 7 studies (n=2374) that measured BNP against clinical criteria (generally a consensus view using all other clinical information). The diagnostic odds ratio was similar in studies conducted in general practice and in hospital settings. Three studies compared BNP with N-terminal–ANP, a precursor form of ANP, and pooling of the results of these studies showed BNP to be a more accurate marker of heart failure than NT-ANP.

**Conclusions:** Brain natriuretic peptide is an accurate marker of heart failure. Use of a cutoff value of 15 pmol/L achieves high sensitivity, and BNP values below this exclude heart failure in patients in whom disease is suspected. As the diagnostic odds ratio for BNP is greater when assessed against clinical criteria than against left ejection fraction alone, BNP may also be detecting patients with “diastolic” heart failure.


---

**T**he rising prevalence\(^1\) and cost\(^2\) of heart failure and increasing treatment options\(^3\) have made the accurate diagnosis of heart failure increasingly important. Heart failure is difficult to diagnose correctly, with both overdiagnosis and underdiagnosis occurring commonly in practice. Studies that have investigated the prevalence of echocardiographic abnormalities in populations have found that at least half of patients with significant left ventricular dysfunction on echocardiogram are asymptomatic or have not previously been diagnosed as having heart failure.\(^4,5\) In the Rotterdam Study, a population-based cohort study of chronic disease in persons aged 55 years or older, 60% of people with a left ventricular ejection fraction of 25% or less were asymptomatic.\(^6\) Conversely, patients with signs and symptoms suggestive of heart failure are frequently found not to have the disease when measured against more objective criteria. One study investigated all suspected new cases of heart failure referred by general practitioners during a 15-month period against a consensus decision of 3 cardiologists, including the results of further investigations such as echocardiogram. Against this definition, only 30% of the referred patients had heart failure.\(^7\)

The diagnosis of heart failure is hindered if access to further investigations is limited. It is especially difficult for general practitioners, who are faced with many patients at high risk of the disease and who must make decisions regarding appropriate further investigation, treatment, and referral. General practitioners in the United Kingdom identified a lack of confidence in establishing an accurate diagnosis as a major barrier to treating patients with heart failure.
cutoff. This was done by taking the average of the natural logarithm of the odds ratio and the average of the variance of the natural logarithm of the odds ratio for cutoff points within a study (unweighted, because the study size was the same in each case), and back-transforming to calculate the average DOR and confidence interval (CI). The studies were grouped so that a DOR was calculated against each reference standard, using a DerSimonian and Laird random-effects model on a logarithmic scale with a corresponding test of heterogeneity. Where possible, the positive and negative likelihood ratios were calculated where studies had used similar cutoff levels and the same reference standard, again using a DerSimonian and Laird random-effects model.

An unweighted least-squares regression model using the method of Moses et al was used to assess whether the odds ratio was independent of the cutoff point. The model is $D = a + bS$, where the difference $D$ is defined as the logit true-positive rate minus the logit false-positive rate, which is equivalent to the log of the DOR, and the sum $S$ is defined as the logit true-positive rate plus the logit false-positive rate, which is a measure in the variation of the threshold. For each study, the values $D$ and $S$ were calculated, then the regression line was fitted. An unpaired, 2-tailed t test was used to see whether the slope of the line was significantly different from zero, which would imply that the overall diagnostic accuracy of the test varied with the cutoff point used. In studies that compared the diagnostic accuracy of BNP with ANP, the estimated area under the curve (AUC) for each study was pooled by an inverse variance method. The diagnostic accuracy of the 2 natriuretic peptides was assessed by comparing the difference between the 2 pooled AUCs divided by the variance of the AUC,

$$\frac{AUC_{\text{BNP}} - AUC_{\text{ANP}}}{\sqrt{\text{var}_{\text{BNP}} + \text{var}_{\text{ANP}}}}$$

with a standard normal distribution. This method will underestimate the true measure of difference, as the data are derived from paired study designs, and would require individual patient data to estimate the true difference.

RESULTS

We identified 20 studies with published data that met the inclusion criteria of the review (Table 1). The quality of the studies was generally high, with most satisfying the following criteria (with the number reporting each criterion in parentheses): (1) patients were a consecutive series or random sample (15/20); (2) the index and reference tests were assessed independently and blinded to the other test result (15/20); (3) all patients received both tests (16/20); (4) the methods for performing both tests were described (20/20); (5) the characteristics of the study population were described (20/20); and (6) there was no time delay between the measurement of the 2 tests (14/20). The details of the first 2 quality criteria are also shown in Table 1.

The DORs of BNP against each reference standard are shown in Figure 1. In the 8 studies that measured BNP against a left ventricular ejection fraction of 40% or less, the pooled DOR was 11.6 (95% CI, 8.4-16.1). The results of the studies were consistent, with no evidence of heterogeneity. With less restrictive echocardiographic criteria, the DOR was smaller and the degree of heterogeneity greater, as would be expected with studies with a more imperfect reference standard and varying levels for the reference threshold. Using a reference standard of left ventricular ejection fraction of 40% or less and pooling studies that used a cutoff between 14 and 19 pmol/L gave an estimated positive likelihood ratio of 4.1 (95% CI, 2.6-6.6) and a negative likelihood ratio of 0.35 (95% CI, 0.17-0.72).

In the 7 studies that measured BNP against a consensus clinical opinion, generally using all other diagnostic information available, the DOR was greater (30.9; 95% CI, 27.0-35.4). The degree of heterogeneity between the studies was also low. This result suggests that BNP and clinical diagnosis are in greater agreement than BNP and left ventricular function, assuming no other differences between the 2 groups of studies. The results were heavily weighted by the results of one study, however, the Breathing Not Properly study reported by Maisel et al in 2002. This was a multicenter study with 1586 patients that assessed the diagnostic accuracy of BNP in patients presenting to 7 emergency departments against the diagnosis by 2 cardiologists who had access to all clinical data, includ-
ing results of radiology and echocardiography. In this study, using a cut-off level of 14.4 pmol/L, the positive likelihood ratio was 2.6 (95% CI, 2.3-2.8) and the negative likelihood ratio was 0.05 (95% CI, 0.03-0.07).

In the 2 studies that measured BNP against echocardiographic criteria for both systolic and diastolic heart failure, the DOR (37.7; 95% CI, 5.9-237.2) was again greater than in studies that measured only systolic function, but the results of the 2 studies were different.

Of the studies that investigated BNP vs systolic or systolic plus diastolic function, 7 studies were conducted in general practice or community settings and 11 were conducted in hospital settings. The DOR from the studies in the 2 settings showed similar results (Figure 1). As these groups of studies pool results using different reference standards, they both showed highly statistically significant levels of heterogeneity.
Table 1. Accuracy of Brain Natriuretic Peptides for Diagnosis of Heart Failure by Diagnostic Standard (cont)

<table>
<thead>
<tr>
<th>Source</th>
<th>Setting, Location</th>
<th>N</th>
<th>Prevalence, %</th>
<th>Quality: Series,*</th>
<th>Reference Test Criteria</th>
<th>Index Test</th>
<th>Cutoff Closest to 15 pmol/L</th>
<th>TP, No.</th>
<th>FN, No.</th>
<th>TN, No.</th>
<th>FP, No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Davis et al,27 1994</td>
<td>Patients admitted for acute dyspnea, Christchurch</td>
<td>52</td>
<td>62</td>
<td>NR, NR</td>
<td>Committee of 3 physicians and radiologist</td>
<td>BNP, Christchurch assay</td>
<td>22</td>
<td>30</td>
<td>2</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>Hobbs et al,28 2002</td>
<td>General population aged &gt;45 y, England</td>
<td>307</td>
<td>2</td>
<td>+, +</td>
<td>ESC criteria (3 cardiovascular clinicians)</td>
<td>NT-proBNP, Roche</td>
<td>36</td>
<td>7</td>
<td>0</td>
<td>210</td>
<td>90</td>
</tr>
<tr>
<td>Hobbs et al,28 2002</td>
<td>Patients at high risk of heart failure</td>
<td>133</td>
<td>7</td>
<td>+, +</td>
<td>ESC criteria (3 cardiovascular clinicians)</td>
<td>NT-proBNP, Roche</td>
<td>36</td>
<td>9</td>
<td>0</td>
<td>55</td>
<td>69</td>
</tr>
<tr>
<td>Hobbs et al,28 2002</td>
<td>Patients taking diuretics</td>
<td>87</td>
<td>16</td>
<td>+, +</td>
<td>ESC criteria (3 cardiovascular clinicians)</td>
<td>NT-proBNP, Roche</td>
<td>36</td>
<td>13</td>
<td>1</td>
<td>29</td>
<td>44</td>
</tr>
<tr>
<td>Hobbs et al,28 2002</td>
<td>Patients with existing diagnosis of heart failure</td>
<td>103</td>
<td>34</td>
<td>+, +</td>
<td>ESC criteria (3 cardiovascular clinicians)</td>
<td>NT-proBNP, Roche</td>
<td>36</td>
<td>35</td>
<td>0</td>
<td>12</td>
<td>56</td>
</tr>
<tr>
<td>Maisel et al,29 2002</td>
<td>BNP study, patients attending emergency department with dyspnea (United States, France, Norway)</td>
<td>1586</td>
<td>47</td>
<td>NR, +</td>
<td>2 cardiovascular clinicians</td>
<td>BNP, Biosite</td>
<td>14.4</td>
<td>722</td>
<td>22</td>
<td>522</td>
<td>320</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Setting, Location</th>
<th>N</th>
<th>Prevalence, %</th>
<th>Quality: Series,*</th>
<th>Reference Test Criteria</th>
<th>Index Test</th>
<th>Cutoff Closest to 15 pmol/L</th>
<th>TP, No.</th>
<th>FN, No.</th>
<th>TN, No.</th>
<th>FP, No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Castro et al,30 2001</td>
<td>Patients referred with suspected heart failure and LVEF &gt;40%</td>
<td>65</td>
<td>71</td>
<td>+, NR</td>
<td>Symptoms and signs of HF and abnormal pattern of mitral inflow or atrial fibrillation and no other cause of symptoms</td>
<td>BNP, Shionogi</td>
<td>10.1</td>
<td>45</td>
<td>1</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>Krishnaswamy et al,25 2001</td>
<td>Patients referred for echocardiogram, San Diego</td>
<td>400</td>
<td>24</td>
<td>NR, +</td>
<td>Diastolic dysfunction = impaired relaxation or restrictive or pseudonormal pattern</td>
<td>BNP, Biosite</td>
<td>21.6</td>
<td>141</td>
<td>27</td>
<td>125</td>
<td>107</td>
</tr>
<tr>
<td>Lubien et al,31 2002</td>
<td>Patients referred for echocardiogram with normal LV function, San Diego</td>
<td>294</td>
<td>40</td>
<td>−, +</td>
<td>Diastolic dysfunction = impaired relaxation or restrictive or pseudonormal pattern</td>
<td>BNP, Biosite</td>
<td>17.9</td>
<td>101</td>
<td>18</td>
<td>145</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Setting, Location</th>
<th>N</th>
<th>Prevalence, %</th>
<th>Quality: Series,*</th>
<th>Reference Test Criteria</th>
<th>Index Test</th>
<th>Cutoff Closest to 15 pmol/L</th>
<th>TP, No.</th>
<th>FN, No.</th>
<th>TN, No.</th>
<th>FP, No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Krishnaswamy et al,25 2001</td>
<td>Patients referred for echocardiogram, San Diego</td>
<td>400</td>
<td>63</td>
<td>NR, +</td>
<td>Systolic or diastolic dysfunction</td>
<td>BNP, Biosite</td>
<td>14.1</td>
<td>230</td>
<td>23</td>
<td>121</td>
<td>26</td>
</tr>
<tr>
<td>Bettencourt et al,14 2000</td>
<td>Patients day 4 and 5 after MI, Portugal</td>
<td>101</td>
<td>66</td>
<td>+, +</td>
<td>LVEF ≤40% or LVEDD index &lt;3.2 cm²/m²</td>
<td>abnormal relaxation or restrictive pattern</td>
<td>BNP, Shionogi</td>
<td>27.1</td>
<td>55</td>
<td>12</td>
<td>26</td>
</tr>
</tbody>
</table>

Abbreviations: BNP, brain natriuretic peptide; ESC, European Society of Cardiology; FN, false negative; FP, false positive; FS, fractional shortening; HF, heart failure; LVEDD, left ventricular end-diastolic diameter; LVEF, left ventricular ejection fraction; LVSD, left ventricular systolic dysfunction; MI, myocardial infarction; MONICA, Multinational Monitoring of Trends and Determinants in Cardiovascular Disease; NT, N-terminal; TN, true negative; TP, true positive.

*Consecutive series or random sample of a consecutive series of patients.
†The results of the index and the reference test were read independently and blinded to the result of the other test; plus sign indicates fulfilled; minus sign, not fulfilled; and NR, not reported.
‡Manufacturer information is as follows: Peninsula Laboratories Europe Ltd (Peninsula), St Helens, England; Shionogi, Osaka, Japan; CIS Bio International (CIS Bio), Gif-sur-Yvette, France; Biosite Incorporated (Biosite), San Diego, Calif; and Roche Diagnostics (Roche), Mannheim, Germany.
The data for the individual studies and the curve derived from the pooled odds ratio for studies comparing BNP with a reference standard of left ventricular ejection fraction of 40% or less are shown in Figure 2. This figure illustrates the variation seen between studies in terms of the sensitivity and specificity for various cutoff points. The Moses et al study regression model was estimated for studies that used a reference standard of left ventricular ejection fraction of 40% or less and clinical diagnosis. These both showed a significant negative relationship, implying that the diagnostic accuracy decreases with increases in the cutoff.

Six studies compared the diagnostic accuracy of BNP and ANP. The results of these studies are

---

**Figure 2.** Estimates of diagnostic accuracy from individual studies that measured brain natriuretic peptide vs left ventricular ejection fraction of 40% or less plotted in receiver operating characteristic curve space, ie, sensitivity vs 1-specificity. Results from different cutoff points within the same study are joined, with the numbers next to the circles being the cutoff point (in picomoles per liter). The pooled diagnostic odds ratio (DOR) is used to estimate the plot of summary receiver operating characteristic curve and the area under the curve (AUC). The area of the circles is proportional to the size of the studies.

---

**Table 2. Studies That Compared BNP and ANP**

<table>
<thead>
<tr>
<th>Source</th>
<th>Setting</th>
<th>Reference Standard Used</th>
<th>Form of ANP</th>
<th>Result for ANP*</th>
<th>Result for BNP*</th>
</tr>
</thead>
<tbody>
<tr>
<td>McDonagh et al, 1998</td>
<td>MONICA study participants, Glasgow, Scotland</td>
<td>LVEF ≤30%</td>
<td>NT-ANP</td>
<td>AUC = 0.75 (0.65-0.85)</td>
<td>AUC = 0.88 (0.82-0.94)</td>
</tr>
<tr>
<td>Choy et al, 1994</td>
<td>Patients day = 3 after MI, Dundee, Scotland</td>
<td>LVEF ≤40%</td>
<td>ANP</td>
<td>OR = 1.45 (0.58-3.68)</td>
<td>OR = 8.57 (2.89-25.39)</td>
</tr>
<tr>
<td>Vasan et al, 2002</td>
<td>Female Framingham Study patients</td>
<td>LVEF ≤40%</td>
<td>NT-ANP</td>
<td>AUC = 0.88 (0.70-1.00)</td>
<td>AUC = 0.85 (0.67-1.00)</td>
</tr>
<tr>
<td>Vasan et al, 2002</td>
<td>Male Framingham Study patients</td>
<td>LVEF ≤40%</td>
<td>NT-ANP</td>
<td>AUC = 0.78 (0.70-0.85)</td>
<td>AUC = 0.79 (0.71-0.86)</td>
</tr>
<tr>
<td>Vasan et al, 2002</td>
<td>Female Framingham Study patients</td>
<td>LVEF ≤50%</td>
<td>NT-ANP</td>
<td>AUC = 0.54 (0.50-0.63)</td>
<td>AUC = 0.56 (0.50-0.65)</td>
</tr>
<tr>
<td>Vasan et al, 2002</td>
<td>Female Framingham Study patients (reference = LVEF &lt;40%)</td>
<td>LVEF ≤50%</td>
<td>NT-ANP</td>
<td>AUC = 0.70 (0.65-0.75)</td>
<td>AUC = 0.72 (0.67-0.77)</td>
</tr>
</tbody>
</table>

Abbreviations: ANP, atrial natriuretic peptide; AUC, area under the curve; BNP, brain natriuretic peptide; ESC, European Society of Cardiology; LVEF, left ventricular ejection fraction; MI, myocardial infarction; MONICA, Multinational Monitoring of Trends and Determinants in Cardiovascular Disease; NT, N-terminal; OR, odds ratio.

*The 95% confidence intervals are given in parentheses.
The results of these 20 studies show that BNP is accurate in the diagnosis of heart failure. However, there is considerable variation in the estimates of diagnostic accuracy between studies, and this variation does not seem to be accounted for by differences in the clinical setting or the type of test used. It would be helpful to understand the sources of this variation before recommending the routine clinical use of the test.

Measurement of BNP is cheaper and is potentially more accessible than echocardiography. Results of BNP testing can be obtained within 20 minutes of blood collection. As echocardiography provides additional information that may be important in the clinical treatment of patients with heart failure, the most likely use of BNP will be in the ambulatory care setting to determine which patients require further testing with echocardiogram. The cutoff level will therefore need to be sufficiently low that patients who have heart failure are not excluded from further testing. In the Breathing Not Properly study, combining BNP with clinical judgment improved diagnostic accuracy across the entire range of diagnostic certainty. Even when the clinician had a high degree of certainty in her or his diagnosis, combining this judgment with the BNP result improved diagnostic accuracy.

Measurement of BNP may play a role in the diagnosis of patients with diastolic heart failure. The diagnostic accuracy of BNP was greater when the definition of disease used as the reference standard included patients who were diagnosed as having heart failure but who had “preserved left ventricular systolic dysfunction.” This raises the question of whether BNP could be a better marker of disease, prognosis, and response to treatment than left ventricular function and whether comparison with echocardiographic criteria of left ventricular function may, in fact, be a comparison with a “silver” standard. To date, almost all of the trials of treatment for heart failure have been conducted in patients with impaired left ventricular function. Although there is a group of patients who have heart failure with preserved left ventricular systolic function (diastolic heart failure), it is not clear how these patients could be identified or how patients chosen by alternative criteria would respond to treatment.

Using a cutoff value of 15 pmol/L achieves high sensitivity, and BNP values below this are able to exclude heart failure in patients in whom disease is suspected. Early studies have also shown a potential role for BNP in monitoring the response to treatment. Future research will need to further define BNP’s role in clinical management, and to determine whether it is better at predicting prognosis and response to treatment than echocardiography.

Accepted for publication March 23, 2004.

We thank Jon Deeks, Senior Medical Statistician, Centre for Statistics in Medicine, Oxford, England, for comments on an earlier draft of this article.

Correspondence: Jenny A. Doust, BMBS, FRACGP, Centre for General Practice, School of Population Health, University of Queensland, Herston Road, Herston Qld 4006, Australia (j.doust@uq.edu.au).

REFERENCES


