Accuracy of B-type natriuretic peptide levels in the diagnosis of left ventricular dysfunction and heart failure: a systematic review

CRD summary
This review found that B-type natriuretic peptide levels can be used to rule out heart failure but their use for identifying systolic dysfunction is more limited. The data presented support these conclusions, but the review suffered from a number of limitations in relation to the literature search, inclusion criteria and reporting of methods of analysis. These findings should therefore be interpreted with caution.

Authors' objectives
To review the accuracy of B-type natriuretic peptide (BNP) levels in the diagnosis of left ventricular dysfunction (LVD) and heart failure (HF).

Searching
MEDLINE (1966 to October 2004), EMBASE (1993 to October 2003), the Cochrane Library (Issue 3, 2003), and MEDION (1998 to 2002) were searched. The search terms, which were reported, included diagnostic filters. Reference lists of retrieved studies and review articles were screened for additional studies. Authors were contacted as necessary for additional study details. Unpublished studies were not included. Only studies published in English, French, German, Italian, Norwegian, Portuguese or Spanish were eligible for inclusion.

Study selection
Study designs of evaluations included in the review
No inclusion criteria relating to the study design were specified. The included studies were both diagnostic cohort and diagnostic case-control studies.

Specific interventions included in the review
Studies that assessed BNP for the diagnosis of HF or LVD were eligible for inclusion. Studies that used NT-pro-BNP were excluded. The specific tests assessed included immunoradiometric assays (IRMA Shionogi and IRMA BNP), Triage BNP, radioimmunoassay (RIA), RIA Peninsula, RIA extractive and RIA Phoenix. Thresholds varied considerably between the studies.

Reference standard test against which the new test was compared
No inclusion criteria relating to the reference standard were specified. The reference standards among the included studies were described as systolic dysfunction, systolic/diastolic dysfunction and cardiac failure. It was unclear how these were assessed and defined.

Participants included in the review
Studies of patients with suspected HF or LVD were eligible for inclusion. Studies in very restrictive subgroups of patients, such as patients with Duchenne disease, Chagas disease or Brugada syndrome, were excluded. The included studies assessed patients with acute myocardial infarction; stable out-patients with coronary artery disease; volunteers with congestive heart failure versus healthy controls; referred patients (for radionuclide study, heart disease evaluation, suspected heart failure, catheterisation, evaluation of dyspnoea, echocardiography); patients in haemodialysis without signs of HF; survivors of acute myocardial infarction; mass screening; hospitalised patients with HF; out-patients treated with diuretics; cardiomyopathy versus healthy controls; intensive care unit patients; diabetic patients; patients with acute dyspnoea; and miscellaneous patients.

Outcomes assessed in the review
The studies had to report sufficient data to construct a 2x2 table of test performance.
How were decisions on the relevance of primary studies made?
The authors did not state how the papers were selected for the review, or how many reviewers performed the selection.

Assessment of study quality
The studies were assessed for methodological quality using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool and a modified checklist based on the Lijmer study. Two reviewers independently assessed the validity of the studies and any disagreements were resolved through consensus or by referral to a third reviewer.

Data extraction
Two reviewers independently extracted the data and any disagreements were resolved through consensus or by referral to a third reviewer. The data were extracted as 2x2 tables of test performance. For studies that contained zero cells in the 2x2 table, 0.5 was added to all cells for that table. If studies provided data for more than one BNP threshold, the threshold that minimised the sum of false positive and false negatives was chosen. The sensitivity, specificity, positive and negative likelihood ratios (LRs), and diagnostic odds ratios (DORs) were calculated for each study, together with their standard errors and 95% confidence intervals (CIs).

Methods of synthesis
How were the studies combined?
The sensitivity, specificity, positive and negative LRs, and DORs were pooled using random-effects models. The Moses-Littenberg model was used to construct summary receiver operating characteristic (SROC) curves. Publication bias was investigated statistically and graphically.

How were differences between studies investigated?
The statistical heterogeneity of the DOR was investigated statistically and graphically. The SROC model was used to assess whether the DOR was independent of the threshold. This model was also used to explore sources of heterogeneity (no further details reported).

Results of the review
Fifty-two studies (16,730 patients) were included.

Four quality items were found to be associated with the DOR: adequate description of study population, adequate description of the test, prospective study and no case-control design. The 12 studies that did not fulfill these four criteria were considered to be of a low quality.

Initial analysis: investigation of heterogeneity (55 studies).
Low-quality studies overestimated the DOR by a factor of 3.7. The accuracy of the test was much greater for the identification of HF than LVD (relative DOR 6.4). Clinical setting, prevalence of cardiac dysfunction and type of BNP assay were not associated with the DOR. Low-quality studies were excluded from further analysis, which was stratified by target condition.

Heart failure (11 studies after the exclusion of 5 poor-quality studies).
There was no indication of a threshold effect (p=0.55) or publication bias (p=0.18). There was evidence of significant heterogeneity in the DOR (p=0.001); this disappeared on the exclusion of 2 outlying studies with very high DORs. After exclusion of these outliers, BNP levels were very accurate for the diagnosis of HF (DOR 28.94; area under the curve 0.93). The negative LR showed very little heterogeneity (p=0.09). The pooled negative LR was 0.11 (95% CI: 0.08, 0.16). There was greater heterogeneity in positive LRs, so these were not pooled.

Systolic and/or diastolic function (7 studies after the exclusion of 3 poor-quality studies).
There was significant heterogeneity between the studies (p<0.0001). The sensitivity ranged from 28 to 92% and the
specificity from 44 to 97%. The funnel plot suggested the possibility of publication bias (p=0.06).

Systolic dysfunction (25 studies after the exclusion of 4 poor-quality studies).

There was significant heterogeneity (p<0.0001) and the funnel plot suggested the presence of publication bias (p=0.0005). Despite the heterogeneity, it appeared that the accuracy was poorer than that of studies of HF.

**Authors' conclusions**
BNP can be used to rule out heart failure. However, its use for identifying patients with systolic dysfunction is more limited.

**CRD commentary**
This review suffered from a number of limitations. The objective was clearly stated and supported by defined inclusion criteria. However, it appears that studies were included that did not fulfill these criteria: the authors state that only studies that included an appropriate spectrum were included (defined as patients in whom it was sensible to suspect the target disorder) whereas the included studies did not all fulfill this criterion. The literature search was limited to two electronic databases and the search was restricted using diagnostic search terms. In addition the review was restricted to published studies in a number of defined languages. It is therefore likely that relevant studies have been missed and the review may be subject to language and publication bias. Where reported, appropriate steps were taken to minimise bias in the review process. A detailed quality assessment appears to have been carried out but the exact items assessed and the results of the quality assessment were not reported.

It appears that appropriate methods were used to pool the results but only limited details of the analysis were reported, in particular in relation to the methods used to investigate heterogeneity. The inclusion of SROC plots greatly helped the interpretation of the results, but further investigation of heterogeneity within the target condition subgroups would have been helpful. The authors' conclusions are supported by the results presented but should be interpreted with caution given the limitations outlined above.

**Implications of the review for practice and research**
Practice: The authors stated that low BNP levels provide reasonably convincing evidence for ruling out HF. However, the reliability of BNP in confirming the existence of HF is debatable. The authors also stated that the ability of BNP to diagnose ventricular dysfunction is only moderate.

Research: The authors did not state any implications for further research.

**Bibliographic details**

PubMedID
16305826

DOI
10.1016/j.ejheart.2005.10.004

Indexing Status
Subject indexing assigned by NLM

MeSH
Cardiac Output, Low /blood /diagnosis /physiopathology; Humans; Natriuretic Peptide, Brain /blood; Ventricular
Dysfunction, Left /blood /diagnosis /physiopathology

AccessionNumber
12006002526

Date bibliographic record published
29/02/2008

Date abstract record published
29/02/2008

Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.