Comparison of the Diagnostic Accuracy of Brain Natriuretic Peptide (BNP) and the N-Terminal Part of the Propeptide of BNP Immunoassays in Chronic and Acute Heart Failure: A Systematic Review

Aldo Clerico,1,2* Marianna Fontana,1 Luc Zyw,1 Claudio Passino,1,2 and Michele Emdin1

Background: We used evidence-based laboratory medicine principles to compare the diagnostic accuracy of brain natriuretic peptide (BNP) and the N-terminal part of the propeptide of BNP (NT-proBNP) assays for the diagnosis of heart failure.

Methods: In May 2006, we performed a computerized literature search of the online National Library of Medicine to select studies specifically designed to compare the diagnostic accuracy of BNP and NT-proBNP assays. The comparison took into account the area under the curve and diagnostic odds ratio (DOR) derived from ROC analysis of original studies.

Results: Both BNP and NT-proBNP assays were found to be clinically useful for the diagnosis of heart failure. Metaanalysis of these data was difficult because of the heterogeneity of data regarding patient population, diagnostic criteria, end-points, and immunoassay methods for both BNP and NT-proBNP. Separate metaanalyses were performed for acute and chronic heart failure. In chronic heart failure, the diagnostic DOR for BNP [8.44, 95% confidence interval (CI) 4.66–15.30] was not significantly different from that of NT-proBNP (23.36, 95% CI 9.38–58.19). In patients with acute heart failure, the mean DOR for BNP (16.46, 95% CI 10.65–25.43) was not significantly different from that of NT-proBNP (18.61, 95% CI 12.99–26.65).

Conclusion: Our results indicate that both BNP and NT-proBNP assays have a high degree of diagnostic accuracy and clinical relevance for both acute and chronic heart failure.
steps of the diagnostic algorithm, along with clinical examination, electrocardiogram, and chest x-ray examination.

The use of BNP or NT-proBNP measurement in clinical practice requires the commercial availability of robust and fully automated immunoassay methods (3, 7, 8). Although several studies have been published recently on the biochemical characteristics and pathophysiological relevance of BNP and NT-proBNP, no systematic reviews have been reported that specifically aimed to compare the diagnostic accuracy of commercial immunoassays for different B-type related natriuretic peptides.

Several recent studies suggested that the diagnostic accuracy of CNH immunoassay methods depends not only on the peptide measured but also on the platform used (3, 7, 8). Commercial immunoassays for BNP use different standard materials and antibodies that are specific for different epitopes located on the N-terminus, C-terminus, or cysteine-ring of the peptide chain (8). Experimental data support the hypothesis that BNP is cleaved by several plasma proteases; therefore, these enzymatic cleavages, particularly the one at the N-terminus, must be considered when choosing epitopes for antibody production and immunoassay design (3, 7, 8). Because of their different biochemical and physiological characteristics, BNP and NT-proBNP measurements may have different advantages (1, 3). BNP is a shorter peptide hormone than NT-proBNP (32 vs 76 amino acids), is inactive, and has a shorter plasma half-life (15–20 min vs >1 h) (1). For these reasons, BNP might show a better correlation with rapid changes in activation of neurohormone systems and hemodynamics than NT-proBNP. On the other hand, NT-proBNP degrades more slowly both in vivo and in vitro, has a higher circulating concentration, and is more stable, with less biological variability, than BNP (1).

We performed a computerized online literature search on the National Library of Medicine using as keywords “BNP assay” and “NT-proBNP assay”. We report results of our metaanalysis of selected studies specifically designed to compare the diagnostic accuracy of BNP and NT-proBNP assays for the differential diagnosis of heart failure.

Materials and Methods

METAANALYSIS

In May 2006, we performed a computerized literature search on the National Library of Medicine (i.e., PubMed access to MEDLINE citations, http://www.ncbi.nlm.nih.gov/PubMed/) using the keywords “BNP assay” (>1600 articles) and “NT-proBNP assay” (>450 articles). Then, we refined the metaanalysis to include only the studies specifically designed to compare the accuracy of BNP and NT-proBNP assays for the diagnosis of heart failure. To evaluate the diagnostic accuracy of these assays we followed the process usually recommended for systematic review of diagnostic accuracy of medical tests (9–13). This systematic review contained only data generated by parallel analysis of a BNP and an NT-proBNP assay on the same set of specimens. The criteria for excluding a metaanalysis from the study were (a) incomplete data on immunoassay methods used, (b) absence of ROC analysis, and (c) absence of diagnostic end-point. In addition, in cases in which several published studies (usually large clinical trials) included the same population (control or patient group), we selected only 1 article, the most recent or complete. One study compared the diagnostic accuracy of 5 commercial methods for BNP with that of the electrochemiluminescence immunoassay (ECLIa) method for NT-proBNP ([7]). Because the diagnostic accuracy did not significantly differ among the BNP methods, only data concerning the IRMA method for BNP (which performed best in the study) were included in the metaanalysis. This study also demonstrated that differences in diagnostic accuracy are more evident when patients with mild heart failure [New York Heart Association (NYHA) classes I and II] are considered rather than patients with severe disease (NYHA classes III and IV). For this reason, when studies (14–19) offered different options for the end-point (severity of left ventricular dysfunction, structural myocardial disease and symptoms) we chose the values associated with diagnosis of milder disease.

The statistical comparison was made by taking into account the area under the curve (AUC) of an ROC analysis and the values of the diagnostic odds ratio (DOR). This double approach was chosen because some articles did not report some statistical parameters, such as confidence limits for AUC or sensitivity/specificity at cutoff values. Thus we could not use only 1 type of accuracy test (e.g., consider only the AUC or DOR values) to analyze data from all studies.

We performed statistical analyses with a Macintosh PowerBook G4 using R statistical software (Version R 2.4.1, December 2006, by the R Foundation for Statistical Computing; section, Meta Package for metaanalysis), which is a free software environment for statistical computing and graphics (http://www.r-project.org). In particular, the random-effects model according to the DerSimonian–Laird method for pooled AUC and DOR values and the Q test for heterogeneity statistics were used. For all studies the counts used for the calculation of 2 × 2 tables were deduced from reported sensitivity, specificity, disease prevalence, and total number of patients studied. Some data concerning the study by Clerico et al. (7) were obtained directly from the authors. To test the heterogeneity of data, the true-positive rates were plotted against the false-positive rates on a Moses-type summary ROC curve (sROC) (13) (Fig. 1). In this plot, the area of the circles is proportional to the inverse of the SE, as reported in the original studies. To calculate the pooled AUC and DOR values, both fixed-effects and random-effects models were used: however, we chose to report only data obtained with the random-effects model with inverse
variance weighting (Figs. 2, 3, and 4), owing to the significant heterogeneity of data (11), as also previously reported in the literature (4).

Results
To reduce the heterogeneity among the clinical studies (as demonstrated by data reported in Fig. 1), we divided the metaanalysis into 2 parts according to the different clinical settings of patients with acute and chronic heart failure.

CHRONIC HEART FAILURE
We included 15 clinical studies specifically designed to compare the diagnostic accuracy of BNP and NT-proBNP assays in patients with left ventricular dysfunction and/or structural myocardial disease and symptoms. The characteristics and number of patients enrolled, the type of immunoassay, and the diagnostic end-points of these studies are reported in Table 1. Echocardiographic evaluation was the reference method for the assessment of myocardial dysfunction (both systolic and systodiastolic) in all studies; in only 1 study (20), diagnosis was performed by use of echocardiographic evaluation combined with clinical criteria (including clinical history, physical examination, electrocardiogram, and chest x-ray).

In the 15 studies reported in Table 1, different immunoassay methods were used to measure BNP and NT-proBNP. The most frequently used immunoassay for NT-proBNP (11 studies) was the ECLI A method (Elesys System, Roche Diagnostics). Other immunoassays used were an EIA (Biomedica; 3 studies) and an in-house RIA (1 study).

Commercial BNP methods used (Table 1) were an IRMA (Shionoria; 7 studies), a point-of-care-testing method (TRIAGE System, Biosite; 5 studies), the ADVIA method (Centaur System, Bayer HealthCare; 3 studies), the MEIA method (AxSYM System, Abbott Laboratories; 2 studies), and an in-house RIA (1 study). The values for AUC and sensitivity/specificity reported for BNP and NT-proBNP assays in the original articles are also included in Table 1.

Higher AUC values were observed with the ECLI A NT-proBNP method than with the 5 commercial BNP assays, whereas other NT-proBNP methods did not show AUC values different from BNP methods (Table 1). Metaanalysis revealed no significant difference between diagnostic accuracy of NT-proBNP and BNP assays for 12 studies reporting the AUC and its respective 95% confidence interval (CI) values (mean AUC for NT-proBNP 0.8308, 95% CI 0.7738–0.8878, test of heterogeneity $P <0.0001$; mean AUC for BNP 0.8182, 95% CI 0.7561–0.8803, test of heterogeneity $P <0.0001$ (7, 14–24).

For Moses-type sROC curve plots of true-positive vs false-positive rates for individual studies including patients with chronic heart failure (Fig. 1A), the Q test of heterogeneity demonstrated a wide dyshomogeneity among all studies ($P <0.0001$), including both BNP and NT-proBNP results. To analyze results of the 12 studies reporting the sensitivity and specificity at cutoff values, we used the sROC approach, which converts each pair of sensitivity and specificity data into a single measure of accuracy (DOR) (9–13), calculated with a random effects model, for BNP (Fig. 2A) and NT-proBNP (Fig. 2B). The pooled DOR for the BNP assay (DOR 8.44, 95% CI
4.66–15.30, test of heterogeneity \( P < 0.0001 \); Fig. 2A) was not significantly different from that for the NT-proBNP assay (23.36, 95% CI 9.38–58.19, test of heterogeneity \( P < 0.0001 \); Fig. 2B).

**ACUTE HEART FAILURE**

Our 2nd metaanalysis included 9 clinical studies (25–33) specifically designed to compare the diagnostic accuracy of BNP and NT-proBNP assays in patients with symptoms of acute heart failure treated in an emergency department. The characteristics and numbers of patients enrolled, the type of immunoassay, and the diagnostic end-points of these studies are reported in Table 2, and the metaanalysis results are reported in Figs. 3 and 4. All studies except the study by Chien et al. (31), used clinical and echocardiographic criteria for the diagnosis of acute heart failure. Chenevier-Gobeaux et al. (27) studied patients with acute heart failure associated with moderate to severe renal dysfunction, stratified in some groups according to the glomerular filtration rate. Therefore we
separately analyzed patient data divided into 3 groups on the basis of glomerular filtration rates: <30, 30–59, and 60–90 [mL·min⁻¹·(1.73 m²)⁻¹].

All studies but 1 (32) used ECLIA methods for NT-proBNP measurement and 7 studies used the TRIAGE method or the MEIA method for BNP measurement (Table 2). The results of metaanalyses including the 6 studies reporting AUC values are shown in Fig. 3A for BNP and Fig. 3B for NT-proBNP. We observed no significant difference in diagnostic accuracy between NT-proBNP and BNP. The mean AUC values calculated with a random-effects model were 0.8391 for BNP (95% CI 0.7816–0.8966, test of heterogeneity P < 0.0001) and 0.8689 for NT-proBNP (95% CI 0.8287–0.9091, test of heterogeneity P < 0.0009). With metaanalyses based on calculation of the DOR values for BNP and NT-proBNP (Fig. 4), the pooled DOR estimate calculated by random effects model for BNP was 16.46 (95% CI 10.65–25.43, test of heterogeneity P = 0.0462) and for NT-proBNP was 18.61 (95% CI 12.99–26.65, test of heterogeneity P = 0.2654), with no significant difference between BNP and NT-proBNP (Fig. 4).
Discussion

Several different methods can be used to measure diagnostic accuracy, but most diagnostic accuracy studies present estimates of sensitivity and specificity, either alone or in combination with other measures (9–13). Pooling pairs of sensitivity and specificity to perform a metaanalysis is not a straightforward process, however, because these measures are often negatively correlated within studies. Some authors have suggested that the sROC approach should be the method of choice for metaanalysis of studies reporting paired sensitivity and specificity data (9–13). The sROC approach converts each sensitivity and specificity data pair into a single measure of accuracy, the DOR (9–13). This approach has been used for systematic review of the diagnostic accuracy of measurement ANP- and BNP-related peptides for heart failure (4). In the present study, we used the sROC approach (12) to compare the diagnostic accuracy of BNP and NT-proBNP assays in patients with acute or chronic heart failure. In particular, because the AUC with respective

![Forest plots of DOR values for BNP (A) and NT-proBNP (B) assays reported for studies of patients with acute heart failure (Table 2). The black diamond with the dotted line indicates the pooled DOR with 95% CI.](image)
Table 1. Chronic heart failure: comparison of the diagnostic accuracy between BNP and NT-proBNP assays.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Number of subjects (patients)</th>
<th>Diagnostic end-point</th>
<th>NT-proBNP assay</th>
<th>AUC</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seino et al. (14)</td>
<td>172 (105)</td>
<td>EF &lt;50%</td>
<td>ECLIA, Roche</td>
<td>0.82</td>
<td>85</td>
<td>73</td>
</tr>
<tr>
<td>Hobbs et al. (21)</td>
<td>133 (10)</td>
<td>EF &lt;40%</td>
<td>ECLIA, Roche</td>
<td>0.73</td>
<td>100</td>
<td>46</td>
</tr>
<tr>
<td>Mikkelsen et al. (34)</td>
<td>150 (22)</td>
<td>EF ≤45%</td>
<td>ECLIA, Roche EIA, Biomedica</td>
<td>0.98</td>
<td>100</td>
<td>85</td>
</tr>
<tr>
<td>Pfister et al. (15)</td>
<td>150 (32)</td>
<td>EF &lt;60%</td>
<td>ECLIA, Roche</td>
<td>0.72</td>
<td>94</td>
<td>37</td>
</tr>
<tr>
<td>Clerico et al. (7)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>451 (279)</td>
<td>EF &lt;50%</td>
<td>ECLIA, Roche</td>
<td>0.95</td>
<td>90</td>
<td>85</td>
</tr>
<tr>
<td>Mueller et al. (16)</td>
<td>148 (37)</td>
<td>EF = 35%–60% vs 60%</td>
<td>ECLIA, Roche</td>
<td>0.93</td>
<td>87</td>
<td>87</td>
</tr>
<tr>
<td>Bhalla et al. (22)</td>
<td>193 (148)</td>
<td>Systolic dysfunction (EF &lt;50%) or diastolic dysfunction</td>
<td>ECLIA, Roche</td>
<td>0.67</td>
<td>53</td>
<td>78</td>
</tr>
<tr>
<td>Zaphiriou et al. (20)</td>
<td>306 (104)</td>
<td>Clinical + echo criteria</td>
<td>ECLIA, Roche</td>
<td>0.85</td>
<td>98</td>
<td>35</td>
</tr>
<tr>
<td>Vanderheyden et al. (38)</td>
<td>72</td>
<td>EF &lt;50%</td>
<td>ECLIA, Roche</td>
<td>0.85</td>
<td>98</td>
<td>35</td>
</tr>
<tr>
<td>Costello-Boerrigter et al. (17)</td>
<td>1869 (115)</td>
<td>EF &lt;50%</td>
<td>ECLIA, Roche</td>
<td>0.78</td>
<td>74</td>
<td>74</td>
</tr>
<tr>
<td>Fonseca et al. (24)</td>
<td>107 (86)</td>
<td>Systolic dysfunction (EF &lt;40%) or diastolic dysfunction</td>
<td>ECLIA, Roche</td>
<td>0.992</td>
<td>96</td>
<td>99</td>
</tr>
<tr>
<td>Mueller et al. (23)</td>
<td>149 (118)</td>
<td>Class B HF by echo</td>
<td>ECLIA, Roche</td>
<td>0.76</td>
<td>90</td>
<td>32</td>
</tr>
<tr>
<td>Hammerer-Lercher et al. (18)</td>
<td>57</td>
<td>EF &lt;48%</td>
<td>BA, Biomedica</td>
<td>0.67</td>
<td>70</td>
<td>73</td>
</tr>
<tr>
<td>Hammerer-Lercher et al. (35)</td>
<td>86 (19)</td>
<td>EF = 40–50%</td>
<td>BA, Biomedica</td>
<td>0.75</td>
<td>98</td>
<td>98</td>
</tr>
<tr>
<td>Richards et al. (19)</td>
<td>1049 (588)</td>
<td>EF &lt;50%</td>
<td>BA, Biomedica</td>
<td>0.73</td>
<td>63</td>
<td>73</td>
</tr>
</tbody>
</table>

<sup>a</sup>In the second column, both the total number of subjects considered in the study and the number of patients affected (in brackets) are reported for the calculation of prevalence of disease.

<sup>b</sup>Data concerning sensitivity and specificity were directly allowed by the authors of the study.
Table 2. Acute heart failure: comparison of diagnostic accuracy between BNP and NT-proBNP assays.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Number of subjects studied (patients)</th>
<th>Diagnostic end-point</th>
<th>NT-proBNP assay</th>
<th>AUC</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
<th>BNP assay</th>
<th>AUC</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lainchbury et al. (25)</td>
<td>205 (70)</td>
<td>Clinical + echocardiographic diagnosis of cardiac (HF-related) vs noncardiac dyspnea</td>
<td>ECLIA, Roche</td>
<td>0.89</td>
<td>80</td>
<td>87</td>
<td>TRIAGE, Biosite</td>
<td>0.89</td>
<td>94</td>
<td>70</td>
</tr>
<tr>
<td>Alibay et al. (26)</td>
<td>160 (60)</td>
<td>Clinical + echocardiographic diagnosis of cardiac (HF-related) vs noncardiac dyspnea</td>
<td>ECLIA, Roche</td>
<td>0.84</td>
<td>97</td>
<td>63</td>
<td>TRIAGE, Biosite</td>
<td>0.82</td>
<td>94</td>
<td>61</td>
</tr>
<tr>
<td>Ray et al. (29)</td>
<td>202 (88)</td>
<td>Clinical + echocardiographic diagnosis of cardiac (HF-related) vs noncardiac dyspnea</td>
<td>ECLIA, Roche</td>
<td>0.85</td>
<td>75</td>
<td>76</td>
<td>TRIAGE, Biosite</td>
<td>0.80</td>
<td>73</td>
<td>91</td>
</tr>
<tr>
<td>Chevenier-Gobeaux et al. (27)</td>
<td>141 (28) 187 (64) 41 (20)</td>
<td>Clinical diagnosis of cardiac vs noncardiac dyspnea</td>
<td>ECLIA, Roche</td>
<td>0.85</td>
<td>77</td>
<td>86</td>
<td>TRIAGE, Biosite</td>
<td>0.84</td>
<td>76</td>
<td>88</td>
</tr>
<tr>
<td>Jefic et al. (28)</td>
<td>41 (34)</td>
<td>Clinical + catheterization diagnosis of cardiac (HF-related) vs noncardiac dyspnea</td>
<td>ECLIA, Roche</td>
<td>0.89</td>
<td>86</td>
<td>83</td>
<td>TRIAGE, Biosite</td>
<td>0.64</td>
<td>■</td>
<td>■</td>
</tr>
<tr>
<td>Collin-Chavagnac et al. (32)</td>
<td>112 (42)</td>
<td>Clinical diagnosis of cardiac vs noncardiac dyspnea</td>
<td>Dimension Dade Behring</td>
<td>0.93</td>
<td>■</td>
<td>■</td>
<td>TRIAGE, Biosite</td>
<td>0.93</td>
<td>■</td>
<td>■</td>
</tr>
<tr>
<td>Bai et al. (30)</td>
<td>41 (25)</td>
<td>Clinical + echocardiographic diagnosis of cardiac (HF-related) vs noncardiac dyspnea</td>
<td>ECLIA, Roche</td>
<td>0.78</td>
<td>84</td>
<td>75</td>
<td>TRIAGE, Biosite</td>
<td>0.81</td>
<td>68</td>
<td>88</td>
</tr>
<tr>
<td>Gegenhuber et al. (33)</td>
<td>251 (137)</td>
<td>Clinical + echocardiographic diagnosis of cardiac (HF-related) vs noncardiac dyspnea</td>
<td>ECLIA, Roche</td>
<td>0.90</td>
<td>87</td>
<td>81</td>
<td>MEIA, Abbott</td>
<td>0.92</td>
<td>80</td>
<td>86</td>
</tr>
<tr>
<td>Chien et al. (31)</td>
<td>55 (19)</td>
<td>Diagnosis of HF by echocardiography</td>
<td>ECLIA, Roche</td>
<td>0.84</td>
<td>90</td>
<td>72.2</td>
<td>MEIA, Abbott</td>
<td>0.70</td>
<td>77</td>
<td>59</td>
</tr>
</tbody>
</table>

95% CI or sensitivity/specificity values were not reported in all studies included in Tables 1 and 2, we made calculations with both pooled AUC and DOR values by means of a random effects model with inverse variance weighting.

Our results confirm that the measurement of both BNP and NT-proBNP shows a high degree of diagnostic accuracy and clinical relevance for the diagnosis of heart failure (3, 4). Furthermore, the present study demonstrates a wide dyshomogeneity among the results of published studies that used both BNP and NT-proBNP immunnoassays. The great dyshomogeneity and the low number of studies suggest the need for caution when interpreting clinical results reported in the literature.

The results reported in Fig. 1 suggest that the dyshomogeneity is greater for the studies concerning patients with chronic heart failure (Fig. 1A) than for those with acute heart failure (Fig. 1B). This finding is in part expected, because patients referred to the emergency department are relatively more clinically homogeneous than those enrolled in studies on chronic stable heart failure. The management of patients suspected to have acute heart failure is mainly based on the differential diagnosis of dyspnea in the emergency department and intensive coronary care unit. According to the definition of acute heart failure (5), patients with severe symptomatic disease (NYHA class III-IV) may present either with de novo dyspnea without previously ascertained cardiac disease or with an acute decompensation of chronic heart failure. On the other hand, patients with chronic heart failure defined as chronic tend to have milder, more stable disease (6). Moreover, studies of the diagnostic accuracy of BNP and NT-proBNP assay for diagnosis of heart failure usually enroll some patients who have mild symptoms (NYHA class II) or even are asymptomatic and have only structural disease (NYHA class I). Patients with chronic heart failure are usually enrolled and studied in cardiological outpatient clinics or during community screening of large general population studies performed with the collaboration of general practitioners. Dyshomogeneity also occurs as the result of differing inclusion (or exclusion) criteria for patient enrollment; in particular, for some studies exclusion criteria included the presence of some degree of renal failure (14, 18, 23, 24, 34, 35). Thus, the prevalence and severity of heart failure varied greatly among the studies, a circumstance that clearly affected the prevalence and severity of heart failure varied greatly among the studies, a circumstance that clearly affected the diagnostic accuracy of the BNP and NT-proBNP assays (3, 4). Finally, although all studies reported that the diagnosis of acute or chronic heart failure was made according to the most recent international guidelines (5, 6, 36), the authors actually chose different diagnostic end-points (Tables 1 and 2). Another cause of dyshomogeneity was the use of different methods for BNP and NT-proBNP measurement (Tables 1 and 2). This effect is even more impressive for the NT-proBNP assays. The analytical performance and clinical results of ECLIA methods differ greatly compared with other commercial or in-house methods for NT-proBNP measurement (Tables 1 and 2). For these reasons, our data confirm that the cutoff values are method dependent (3, 7, 8, 37).

In conclusion, our results, taken as a whole, indicate that both BNP and NT-proBNP assays, without significant differences between the 2 analytes, have high diagnostic accuracy and clinical relevance for assessment of both acute and chronic heart failure patients.

Grant/funding support: None declared.

Financial disclosures: None declared.

References


