B-Type Natriuretic Peptide and Clinical Judgment in Emergency Diagnosis of Heart Failure
Analysis From Breathing Not Properly (BNP) Multinational Study

Peter A. McCullough, MD, MPH; Richard M. Nowak, MD, MBA; James McCord, MD; Judd E. Hollander, MD; Howard C. Herrmann, MD; Philippe G. Steg, MD; Philippe Duc, MD; Arne Westheim, MD, PhD; Torbjørn Omland, MD, PhD, MPH; Cathrine Wold Knudsen, MD; Alan B. Storrow, MD; William T. Abraham, MD; Sumant Lamba, MD; Alan H.B. Wu, PhD; Alberto Perez, MD; Paul Clopton, MS; Padma Krishnaswamy, MD; Radmila Kazanegra, MD; Alan S. Maisel, MD; for the BNP Multinational Study Investigators

Background—We sought to determine the degree to which B-type natriuretic peptide (BNP) adds to clinical judgment in the diagnosis of congestive heart failure (CHF).

Methods and Results—The Breathing Not Properly Multinational Study was a prospective diagnostic test evaluation study conducted in 7 centers. Of 1586 participants who presented with acute dyspnea, 1538 (97%) had clinical certainty of CHF determined by the attending physician in the emergency department. Participants underwent routine care and had BNP measured in a blinded fashion. The reference standard for CHF was adjudicated by 2 independent cardiologists, also blinded to BNP results. The final diagnosis was CHF in 722 (47%) participants. At an 80% cutoff level of certainty of CHF, clinical judgment had a sensitivity of 49% and specificity of 96%. At 100 pg/mL, BNP had a sensitivity of 90% and specificity of 73%. In determining the correct diagnosis (CHF versus no CHF), adding BNP to clinical judgment would have enhanced diagnostic accuracy from 74% to 81%. In those participants with an intermediate (21% to 79%) probability of CHF, BNP at a cutoff of 100 pg/mL correctly classified 74% of the cases. The areas under the receiver operating characteristic curve were 0.86 (95% CI 0.84 to 0.88), 0.90 (95% CI 0.88 to 0.91), and 0.93 (95% CI 0.92 to 0.94) for clinical judgment, for BNP at a cutoff of 100 pg/mL, and for the 2 in combination, respectively (P<0.0001 for all pairwise comparisons).

Conclusions—The evaluation of acute dyspnea would be improved with the addition of BNP testing to clinical judgment in the emergency department. (Circulation. 2002;106:416-422.)

Key Words: heart failure ■ diagnosis ■ natriuretic peptides ■ lung ■ tests
over the conventional information obtained in the evaluation of patients with acute dyspnea.

Methods

Setting
The Breathing Not Properly (BNP) Multicenter Study was an international 7-center prospective study (5 US centers and 2 European centers). The study was conducted from April 1999 to December 2000. The institutional review boards of all study centers approved the study protocol, and all participants provided written informed consent.

Study Sample
A total of 1666 patients presenting to the emergency departments (EDs) of the study centers with a primary complaint of dyspnea were screened. Eighty patients were excluded from the study on the basis of the protocol exclusion criteria, which included the presence of advanced renal failure (calculated creatinine clearance <15 mL/min), acute myocardial infarction, and overt cause of dyspnea, including chest wall trauma or penetrating lung injury. A total of 1586 participants were enrolled in the present study. For those 1586 individuals, 48 had records that did not have the ED physician assessment of clinical probability of CHF; hence, those patients were excluded, leaving a final set of 1538 to be analyzed.

Data Collection
Baseline demographics, clinical history, and objective assessment of clinical signs were gathered by trained ED research personnel who were present continuously during the evaluation of the consenting individuals. All participants were seen and examined by an attending physician, and findings from the ECG, chest x-ray, and blood tests were categorized in a structured checklist. On disposition from the ED, research personnel recorded the attending physician’s estimate of clinical probability of CHF on a visual analog scale.

Measurement of BNP
During initial evaluations, a blood sample (5 mL) was collected into tubes containing potassium EDTA (1 mg/mL blood). In a 15-minute period, BNP was measured by using the Triage BNP Test (Biosite Inc.). The Triage BNP Test is a fluorescence immunoassay for the quantitative determination of BNP in whole-blood and plasma specimens. Precision, analytical sensitivity, and stability characteristics of the system have been previously described. The coefficient of variation for BNP levels of 28.8, 584.0, and 1180.0 pg/mL, respectively. The mean of the 3 samples. Of note, the current approved clinical method is to measure BNP in a single run of the test. Test results

Statistical Analysis
Baseline characteristics were reported in counts and proportions or mean ± SD values as appropriate. Univariate comparisons were made with χ² or 2-sample t tests as appropriate. Because this was the largest and most broadly inclusive population with dyspnea to be tested for BNP to date, we decided a priori to derive the optimum cut point for BNP from the parent population of 1586 participants. We arrived at the optimum cut point of 100 pg/mL by selecting the point on the receiver operating characteristic (ROC) curve that maximized both sensitivity and 1 − specificity. The optimum cut point for ED clinical certainty of CHF was chosen at ≥ 80%, a cut point providing reasonable and actionable certainty of a cardiovascular syndrome. Decision statistics were computed from 2 × 2 tables and reported as sensitivity, specificity, and positive and negative predictive value. Diagnostic accuracy was computed as the sum of the concordant cells divided by the sum of all cells in the 2 × 2 table. Agreement between clinical judgment and BNP was quantified by using Cohen’s κ statistic. The positive likelihood ratio was taken as the slope of the ROC curve for the optimum cut point and was expressed as sensitivity/1 − specificity. Pairwise comparisons among the areas under ROC curves were made by using Delong’s method. Logistic regression was used to combine clinical judgment with BNP data in predicting final adjudicated diagnosis, generating a graphic displayed as a heart failure diagnosis nomogram. Judgments of 0% were set to 1%, and judgments of 100% were set to 99% so that the log of the odds ratios could be computed.

Results

Baseline Characteristics
The demographics for the study sample were as follows: age 64.0±16.7 (range 18 to 105) years; 883 (55.7%) men and 703 (44.3%) women; and 773 (48.7%) white, 715 (45.1%) African American, and 98 (6.2%) other race. Additional baseline characteristics are reported in Table 1 according to the ED setting. Approximately 30 days after the ED visit, the case report form (excluding the estimate of CHF probability), ECG, chest x-ray, echocardiogram, and all other clinical tests and consultations were reviewed by 2 independent cardiologists at the local study center who were not treating physicians. In addition, case report information was used to calculate the Framingham scores (requiring 2 major or 1 major and 2 minor criteria for CHF) and National Health and Nutrition Examination Survey (NHANES) scores (requiring ≥ 3 points for CHF) for CHF. After reviewing all information, if agreement was achieved, then the case was categorized as one of the following: (1) dyspnea due to CHF, (2) history of CHF but dyspnea due to noncardiac cause, or (3) dyspnea due to noncardiac cause. In the event of disagreement (n=164; 10.7%, range 0% to 24.3% across 7 sites), cases were adjudicated by the study end-points committee. For binary analyses of CHF versus no CHF, groups 2 and 3 were combined.

Sample Size and Power
The primary end point was diagnostic accuracy at the optimum cutoff of BNP and at ≥ 80% ED physician estimate of clinical probability of CHF. The following assumptions were made in the sample size calculation: diagnostic accuracy of the ED physician, 85%; prevalence of CHF as a final diagnosis in the ED dyspnea population, 30%; and effect size of ≥ 5% absolute difference between clinical judgment and BNP, β=0.20 and α=0.05 (2-sided). The calculated sample size of 1613 was set for study to have 80% power to observe a ≥ 5% absolute difference in diagnostic accuracy between the groups. With the 1538 participants evaluated in this analysis having a higher prevalence of CHF and larger effect size than expected, the observed power was 99%.
Symptoms and Physical Examination Findings
All participants in the present study required dyspnea on exertion or at rest for study inclusion. Table 2 indicates that the cardinal symptoms and signs (paroxysmal nocturnal dyspnea, elevated jugular venous pressure, pulmonary rales, cardiac enlargement, third heart sound, hepatic enlargement, and edema) of CHF were more common as ED clinical judgment was more certain of the diagnosis of CHF. Conversely, approximately one fourth of all participants had wheezing, regardless of pretest probability for CHF.

Diagnostic Testing Performed
All participants were subjected to ECG, and a majority, 1476 (96.0%), had chest x-rays performed in the ED. Table 3 lists the results of these tests stratified by the clinical probability of CHF as assessed by the ED physicians. Rates of all ECG abnormalities were more frequent in those with high clinical probabilities of CHF. Likewise, the rates of chest x-ray abnormalities indicating signs of CHF were more frequent in the high probability of CHF category. However, the presence of pneumonic infiltrate was not statistically significant across the categories (all <10%).

Reference Standard for Heart Failure
Two independent cardiologists at each study center evaluated all clinical data, including echocardiograms with reported
ejection fractions in 689 (44.8%) cases. There was initial agreement between the 2 cardiologists in 1374 (89.3%) of the cases. The remaining 164 cases required adjudication locally between the 2 cardiologists, including requesting additional data from the treating physicians and, finally, review by the end-points committee if disagreement remained. The diagnosis of CHF (n = 722) was supported by positive NHANES and Framingham scores in 599 (83.0%) and 621 (86.0%) individuals, respectively. The cardiologists reported that the diagnosis of CHF was supported in 587 (81.3%) by chest x-ray, 448 (62.0%) by echocardiography, 34 (4.7%) by nuclear ventriculography, and 55 (7.6%) by cardiac catheterization. In addition, ejection fractions in 689 (44.8%) cases. There was initial agreement between the 2 cardiologists in 1374 (89.3%) of the cases. The remaining 164 cases required adjudication locally between the 2 cardiologists, including requesting additional data from the treating physicians and, finally, review by the end-points committee if disagreement remained. The diagnosis of CHF (n = 722) was supported by positive NHANES and Framingham scores in 599 (83.0%) and 621 (86.0%) individuals, respectively. The cardiologists reported that the diagnosis of CHF was supported in 587 (81.3%) by chest x-ray, 448 (62.0%) by echocardiography, 34 (4.7%) by nuclear ventriculography, and 55 (7.6%) by cardiac catheterization.

### Table 2: Symptoms and Physical Exam Findings by Pretest Probability of CHF

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Clinical Judgment of Probability of CHF as Cause of Dyspnea</th>
<th>P (Low vs High)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low (0%–20%)</td>
<td>Intermediate (21%–79%)</td>
</tr>
<tr>
<td>N</td>
<td>721 (46.9%)</td>
<td>427 (27.8%)</td>
</tr>
<tr>
<td>Symptoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orthopnea</td>
<td>307 (42.6%)</td>
<td>277 (64.9%)</td>
</tr>
<tr>
<td>Paroxysmal nocturnal dyspnea</td>
<td>290 (40.2%)</td>
<td>267 (62.5%)</td>
</tr>
<tr>
<td>Night cough</td>
<td>306 (44.3%)</td>
<td>189 (44.3%)</td>
</tr>
<tr>
<td>Fatigue or weakness</td>
<td>484 (67.1%)</td>
<td>341 (79.9%)</td>
</tr>
<tr>
<td>Physical exam</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart rate, bpm</td>
<td>94±22</td>
<td>92±24</td>
</tr>
<tr>
<td>Systolic blood pressure, mm Hg</td>
<td>137±29</td>
<td>142±30</td>
</tr>
<tr>
<td>Diastolic blood pressure, mm Hg</td>
<td>78±16</td>
<td>78±20</td>
</tr>
<tr>
<td>Elevated jugular venous pressure</td>
<td>50 (6.9%)</td>
<td>116 (27.2%)</td>
</tr>
<tr>
<td>Wheezing</td>
<td>214 (29.7%)</td>
<td>109 (25.5%)</td>
</tr>
<tr>
<td>Rales (any lung field)</td>
<td>168 (23.3%)</td>
<td>223 (52.2%)</td>
</tr>
<tr>
<td>Cardiac enlargement</td>
<td>37 (5.1%)</td>
<td>52 (12.2%)</td>
</tr>
<tr>
<td>Third heart sound</td>
<td>15 (2.1%)</td>
<td>38 (8.9%)</td>
</tr>
<tr>
<td>Hepatic congestion</td>
<td>33 (4.6%)</td>
<td>58 (13.6%)</td>
</tr>
<tr>
<td>Edema</td>
<td>151 (20.9%)</td>
<td>227 (53.2%)</td>
</tr>
</tbody>
</table>

Values are n (%) or mean±SD.

### Table 3: ECG and Chest X-Ray Findings by Pretest Probability of CHF

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Clinical Judgment of Probability of CHF as Cause of Dyspnea</th>
<th>P (Low vs High)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low (0%–20%)</td>
<td>Intermediate (21%–79%)</td>
</tr>
<tr>
<td>ECG, n</td>
<td>721 (46.9%)</td>
<td>427 (27.8%)</td>
</tr>
<tr>
<td>Q waves</td>
<td>86 (11.9%)</td>
<td>83 (19.4%)</td>
</tr>
<tr>
<td>ST depression</td>
<td>14 (1.9%)</td>
<td>19 (4.4%)</td>
</tr>
<tr>
<td>T-wave inversion</td>
<td>29 (4.0%)</td>
<td>43 (10.1%)</td>
</tr>
<tr>
<td>Atrial fibrillation or flutter</td>
<td>39 (5.4%)</td>
<td>53 (12.4%)</td>
</tr>
<tr>
<td>QRS duration, ms</td>
<td>94±24</td>
<td>102±28</td>
</tr>
<tr>
<td>Right bundle branch block</td>
<td>34 (4.7%)</td>
<td>28 (6.6%)</td>
</tr>
<tr>
<td>Left bundle-branch block</td>
<td>19 (2.6%)</td>
<td>23 (5.4%)</td>
</tr>
<tr>
<td>Chest x-ray, n</td>
<td>691 (47.0%)</td>
<td>412 (28.2%)</td>
</tr>
<tr>
<td>Normal</td>
<td>325 (47.2%)</td>
<td>116 (28.2%)</td>
</tr>
<tr>
<td>Enlarged cardiac silhouette</td>
<td>118 (24.6%)</td>
<td>182 (61.1%)</td>
</tr>
<tr>
<td>Cephalization of pulmonary vessels</td>
<td>47 (6.8%)</td>
<td>86 (20.9%)</td>
</tr>
<tr>
<td>Pleural effusion</td>
<td>54 (7.8%)</td>
<td>79 (19.2%)</td>
</tr>
<tr>
<td>Interstitial edema</td>
<td>26 (3.8%)</td>
<td>70 (17.0%)</td>
</tr>
<tr>
<td>Alveolar edema</td>
<td>13 (1.9%)</td>
<td>19 (4.6%)</td>
</tr>
<tr>
<td>Pneumonic infiltrate</td>
<td>51 (7.4%)</td>
<td>35 (8.5%)</td>
</tr>
<tr>
<td>Hyperinflated lungs</td>
<td>60 (8.7%)</td>
<td>14 (3.4%)</td>
</tr>
</tbody>
</table>

Values are n (%) or mean±SD.
tion, the cardiologists reported that 490 (67.9%) of those with CHF had an expected response to CHF therapy. Conversely, 684 (91.4%) of those 748 found not to have CHF had cumulative evidence from chest x-ray, echocardiography, or ventriculography suggesting that CHF was not the cause of dyspnea.

**Decision Statistics**

The diagnostic accuracy for high (80% to 100%) ED probability of CHF on clinical grounds was 74.0%. The other decision statistics for this category were as follows: sensitivity 49% (95% CI 47% to 52%), specificity 96% (95% CI 95% to 97%), positive predictive value 91% (95% CI 90% to 92%), negative predictive value 68% (95% CI 66% to 71%), and positive likelihood ratio 11.5. Diagnostic accuracy for BNP ≥100 pg/mL was 81.2%. The other decision statistics for BNP were as follows: sensitivity 90% (95% CI 89% to 92%), specificity 73% (95% CI 71% to 73%), positive predictive value 75% (95% CI 72% to 77%), negative predictive value 90% (95% CI 88% to 91%), and positive likelihood ratio 3.4. For a composite decision based on clinical probability of 80% to 100% or BNP ≥100 pg/mL, or both, the diagnostic accuracy was 81.5%, sensitivity was 94% (95% CI 93% to 95%), specificity was 70% (95% CI 68% to 73%), positive predictive value was 74% (95% CI 71% to 76%), negative predictive value was 93% (95% CI 92% to 94%), and the positive likelihood ratio was 3.2. As an overall measure of diagnostic value, BNP levels ≥100 pg/mL would have added to clinical judgment, thus boosting accuracy from 74.0% to 81.5% (P<0.0001) (Figure 2). Overall, BNP at a cut point of 100 pg/mL and clinical judgment ≥80% certainty were relatively independent indicators, as reflected by a κ value of 0.30 (P<0.0001). In participants without a self-reported history of CHF (n=1027), the diagnostic accuracy of BNP was 80.4%.

In other important subgroups, including men, women, whites, African Americans, the elderly (aged >70 years), and those with ischemic heart disease, the diagnostic accuracy of BNP was 83.6%, 78.0%, 80.7%, 81.0%, 78.1%, and 81.2%, respectively. Compared through a range of values with the use of ROC curves (Figure 3), the areas under the ROC curve were 0.86, 0.90, and 0.93 for clinical judgment, for BNP, and for the 2 in combination, respectively (P<0.001 for all pairwise comparisons).

**Heart Failure Diagnosis Nomogram**

Figure 4 displays a CHF diagnosis nomogram with the estimate of pretest probability being the certainty in the ED that dyspnea is due to CHF. The rates of actual CHF by final adjudicated diagnosis were 17.1%, 33.6%, and 49.3% for the low-, intermediate-, and high-probability groups, respectively (P<0.0001 for trend). The middle line represents BNP level in picograms per milliliter at the time of presentation. When a straight line is drawn through the pretest probability and BNP level in picograms per milliliter, the posttest probability is found on the right line. For example, a clinical judgment of 20% probability of CHF with a BNP of 1000 pg/mL yields an ≈85% probability of CHF based on these 2 predictors. As indicated, BNP has the greatest value as a diagnostic test in the intermediate zone of probability. In this category, BNP...
BNP Measurement and Diagnosis of CHF

 erection benefited from a comprehensive longitudinal electronic medical record, and, likely, there was less variation in the ascertainment of the original diagnosis and greater precision in the gold standard assessment of CHF.

The present study confirms the value of the careful history and physical examination in patients with dyspnea. The sharpest gradients in the cardinal features of CHF were seen across our diagnostic probability categories. Indeed, the symptom of paroxysmal nocturnal dyspnea and the physical examination findings of rales, cardiac enlargement on palpation, third heart sound, and peripheral edema were all 2 to 3 times more likely in those patients with a high probability of CHF. Conversely, wheezing did not appear to be a factor in the clinicians' ability to discriminate among cases. The ECG and chest x-ray appeared to be valuable in the development of a clinical probability discrimination for CHF. Of note, only a pneumatic infiltrate appeared to be of little help in making an assignment of CHF probability. Despite the value of a careful clinical examination seen in the present study, it has been shown in several prior studies that the clinical examination for CHF is limited. The source of plasma BNP is cardiac ventricles, which suggests that BNP may be a more sensitive and specific indicator of ventricular disorders than other natriuretic peptides. This release appears to be responsive to wall tension, which, in turn, is affected by a variety of determinants that are deranged in CHF. The results of the BNP Multinational Study reported in the present study suggest that the biological properties of this peptide make it an attractive test for the acute ED diagnosis of CHF.

The present study has multiple limitations related to any study that attempts to create a gold standard for a clinical syndrome. Blind cardiologists used all possible information in making the final adjudicated CHF diagnosis. We attempted to aid in this process by creating standardized CHF scores from 2 prior validated methods for the cardiologists to view with all of the clinical data. We acknowledge that misclassification bias is possible and difficult to quantify. It is also possible that the measurement of BNP could have been confounded by other factors, including acute ischemia or renal insufficiency, in patients who were not excluded on these grounds. It is unlikely that missing data, either in the pretest, test, or posttest probability categories, have influenced the results, given the fact that the study sample was restricted to 1538 individuals, ensuring complete data in all cases.

We believe that the importance of the present study will be to advance the current state of certainty regarding the usefulness of BNP in the diagnosis of CHF as indicated by the most recently published set of CHF guidelines. Our findings are supportive of the recently published European guidelines for the diagnosis and treatment of CHF, which incorporate BNP as a diagnostic test for routine clinical practice. In addition to being a useful outpatient screening tool for left ventricular dysfunction, results of the BNP Multinational Study support the use of BNP in the ED. Routine use of BNP in the evaluation of suspected heart failure would be largely confirmatory, yet still valuable in cases in which the clinician has a high degree of certainty of the diagnosis. Importantly, BNP would clarify the final diagnosis in a large proportion of cases encountered in the ED. We anticipate that the published nomogram in this article will be useful to ED and other physicians in establishing the diagnosis of CHF in patients with dyspnea of uncertain etiology. This nomogram leverages an objective yet conservative approach, providing a safeguard for the highly subjective clinical assessment of patients who
have CHF presenting as dyspnea of uncertain etiology. To put this in context, 90% of the patients who had CHF but were thought by the ED physician to be of low probability (≤20%) would have been correctly diagnosed with a point-of-care blood test, allowing for rapid triage and appropriate care of these patients. Importantly, a final degree of clinical utility is achieved by integration of a careful history, physical examination, ECG, chest x-ray, and BNP level, as demonstrated in the ROC curves.

In conclusion, in a multinational sample of men and women seen in the ED with acute dyspnea, BNP measurement would have added to clinical judgment in establishing a final diagnosis of CHF. In those patients with an intermediate probability of CHF, BNP would have clarified the diagnosis in the majority of cases.

Acknowledgment
Some financial support and Triage BNP devices and meters were provided by Biosite, Inc. We are indebted to Roberta A. Sullivan, BSN, MPH, for preparation of this manuscript.

References
B-Type Natriuretic Peptide and Clinical Judgment in Emergency Diagnosis of Heart Failure: Analysis From Breathing Not Properly (BNP) Multinational Study


for the BNP Multinational Study Investigators

_Circulation_. 2002;106:416-422; originally published online July 1, 2002;
doi: 10.1161/01.CIR.0000025242.79963.4C

_Circulation_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2002 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://circ.ahajournals.org/content/106/4/416

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation_ is online at:
http://circ.ahajournals.org//subscriptions/