# Critical Review Form

## Diagnostic Test

Improving the Diagnosis of Acute Heart Failure Using a Validated Prediction Model, *Journal of the American College of Cardiology* 2009; 54: 1515-1521

**Objective:** **To analyze "NT-proBNP as a continuous variable by using data from a previously reported study of patients presenting to the ED with undifferentiated shortness of breath, deriving and externally validating a novel mathematical prediction model for diagnosing AHF and assessing this approach for appropriately redirecting the clinician's diagnostic impression". (p. 1516)**

**Methods:** **Model derived from previously collected, industry-sponsored trial of 534 patients presenting to 1 of 7 Canadian urban ED's with undifferentiated shortness of breath between December 2004 and December 2005 (IMPROVE-CHF). Exclusion criteria included AMI (elevated troponin or ST-T change ≥ 1 mV), renal failure (creatinine > 2.8 mg/dL), malignancy or clear etiology (wheezing in young healthy asthmatic).**

**After history and physical exam, chest x-ray and ECG, the emergency physician was asked to estimate the probability of AHF without knowledge of the NT- proBNP value.* After all subjects had been enrolled, AHF diagnosis criteria standard was adjudication by two Cardiologists using the Framingham Heart Score and NHANES I as guides. These adjudicating Cardiologists were blinded to the NT-proBNP level but had access to all other clinical data, including a 60-day follow-up telephone conversation.**

* the study population was divided into low (≤ 20%), intermediate (21% - 79%), or high (> 80%) pre-test probability.

**The diagnostic performance characteristics of NT- proBNP were analyzed as a categorical (< 300 pg/mL, ≥ 300 – 900 pg/mL, and ≥ 900 pg/mL) and as a continuous variable via the logarithmic ranges < 300, ≥ 300 and < 900, ≥ 900 and < 2700, ≥ 2700 and < 8100, and ≥ 8100. Investigators then fit a multiple logistic regression model using a pre-test probability combined with NT and tested the concordance index (*c* statistic, equivalent to the area under an ROC curve) and tested the concordance index (α-statistic equivalent to AUC for ROC) and tested to discriminatory power of the model via bootstrap method to ensure against over-fitting. Finally, the**

**investigators validated their model against a distinct cohort of 573 patients, from the US-based PRIDE study. Specifically, they analyzed the ability of the model by the net reclassification improvement (NRI) and the integrated discrimination improvement (IDI) which assesses appropriate re-classification of patients (for example if the test in question changes non-CHF patient from pre-test intermediate risk to post-test low risk then the test has *appropriately* re-classified the patient).**

**The algebraic model was published in an online appendix (p. 1521):**

$$\text{Probability AHF} = \frac{1}{1+\exp(8 + 0.01 \times \text{age} - 5.9 \times \text{pt prob} - 2.3 \times \text{lnt bnp} + 0.82 \times \text{pt prob} \times \text{lnt bnp})}$$

**Where**

**pt prob = patient's pre-test probability**
**lnt bnp = log (to base 10) of NT- proBNP**

| Guide | | Comments |
|---|---|---|
| **I.** | **Are the results valid?** | |
| **A.** | **Did clinicians face diagnostic uncertainty?** | Yes. "After the chest radiograph and electrocardiogram were reviewed, the emergency physician estimated the probability of AHF (from 1% to 100%) without knowledge of the drawn NT-proBNP value". (p. 1516) |
| **B.** | **Was there a blind comparison with an independent gold standard applied similarly to the treatment group and to the control group?** **(Confirmation Bias)** | Yes. For both the IMPROVE CHF and PRIDE studies adjudication for acute heart failure was determined independently by two cardiologists with access to all clinical and follow-up data except the NT- proBNP level". (p. 1516) |
| **C.** | **Did the results of the test being evaluated influence the decision to perform the gold standard?** **(Ascertainment Bias)** | No, all subjects analyzed had NT- proBNP Cardiology adjudication (criterion standard) for CHF. |
| **II.** | **What are the results?** | |

| | | |
|---|---|---|
| **A.** | **What likelihood ratios were associated with the range of possible test results?** | • 483 subjects from IMPROVE-CHF were analyzed with mean age 70-years.<br>• Adjudication resulted in 10 discordant cases.<br>• Pre-test probabilities were as follows:<br><br>Pre-test: Low 26 (16%) AHF, 137 (84%) No AHF, 163 (33.7%) Total; Intermediate 80 (43.5%) AHF, 104 (56.5%) No AHF, 184 (38.1%) Total; High 115 (84.6%) AHF, 21 (15.4%) No AHF, 136 (28.2%) Total<br><br>• Overall, median NT- proBNP values were 320 pg/mL for no CHF and 3820 pg/mL for CHF, but there was significant overlap between groups. (Fig 1, p. 1517)<br>• NT- proBNP LR's standard cut points:<br><br>NT- proBNP: < 300: 12 AHF, 129 No AHF, 0.11 (0.06-0.19)*; 300-899: 14 AHF, 49 No AHF, 0.34 (0.19-0.60); ≥ 900: 195 AHF, 84 No AHF, 2.75 (2.29-3.30)<br><br>• NT- proBNP LR's multiple cut points:<br><br>< 300: 12 AHF, 129 No AHF, 0.11 (0.06-0.19)*<br>300-899: 14 AHF, 49 No AHF, 0.34 (0.19-0.60)<br>900-2699: 57 AHF, 50 No AHF, 1.35 (0.97-1.89)<br>2700-8099: 84 AHF, 29 No AHF, 3.43 (2.34-5.03)<br>≥ 8100: 54 AHF, 5 No AHF, 12.80 (5.21-31.45)*<br>*only the < 300 or ≥ 8100 pg/mL is clinically useful to substantially alter post-test probability.<br><br>• The model displayed negligible over-fitting and excellent discriminatory power (C=097), but did under-estimate AHF probability.<br>• Validation of the model on the PRIDE cohort (with statistically significant differences in age, AHF probability and NT- proBNP levels) most of the reclassification occurred in the intermediate probability group, but NT-proBNP had 89% and 95% accuracy in the low and high prob groups, respectively. (Table 4, p.1519) |

| | | |
|---|---|---|
| | | Intermediate Probability Group (N=139)<br><br>Post-Test Stratification: Low prob (N=38): no AHF 37, AHF 1, % appropriately reclassified 97.3%; Int prob (N=77): no AHF 25, AHF 52, % appropriately reclassified -; High prob (N=24): no AHF 0, AHF 24, % appropriately reclassified 100%<br><br>• 44% of intermediate probability patients would be appropriately reclassified as low-probability or high-probability by this model, substantially improving upon clinical judgment alone with NRI 0.23 ($p < 0.001$) and IDI 0.11 ($p<0.001$). |
| **III.** | **How can I apply the results to patient care?** | |
| **A.** | **Will the reproducibility of the test result and its interpretation be satisfactory in my clinical setting?** | "Although currently not generalizable to all settings, the fact that the two study cohorts were from different countries and so different (Table 3) suggests the model may perform well in other patient populations". (p. 1520) |
| **B.** | **Are the results applicable to the patients in my practice?** | Yes. ED adult patients in urban ED's with undifferentiated dyspnea. |
| **C.** | **Will the results change my management strategy?** | Yes – will incorporate model into CHF probability assessment pending confirmatory trials (see Excel file). |
| **D.** | **Will patients be better off as a result of the test?** | Uncertain Green, et al suggest that ED patients with undifferentiated dyspnea and clinical uncertainty (intermediate probability) for AHF have longer hospital length of stay and increased mortality between those in whom clinical certainty is attained. Future RCT's will need to determine whether clinician awareness and interpretation of NT- proBNP/BNP results can positively impact these patient important outcomes. |

## Limitations

1) **Industry relationships (Dr. Januzzi) mandate healthy skepticism regarding data interpretation in lieu of non-industry sponsored trial.**
2) **Complicated algebraic equation limiting bedside application.**
3) **Uncertain applicability to BNP since NT-proBNP was used.**
4) **No assessment of pre-test probability inter-observer variability which could significantly impact model stability across departments or institutions.**
5) **Cannot apply model in setting of AMI, ARF, or malignancy.**

## Bottom Line

**A complicated algebraic model derived in the multi-center Canadian IMPROVE-CHF cohort and validated retrospectively in the U.S. PRIDE cohort improves the diagnostic accuracy for AHF in ED patients with undifferentiated dyspnea. Future studies should validate and/or refine this model while assessing mechanisms for clinical uptake, cost, and impact on patient important outcomes (length-of-stay, morbidity, mortality).**