Use of Natriuretic Peptide Measurement in the Management of Heart Failure

Executive Summary

Background

Heart failure (HF) is a major concern for health care systems because of its chronic nature and resource implications. HF affects approximately 5.7 million Americans, and 670,000 new cases are diagnosed annually.1 Based on current population estimates,2 HF is present in 1.8 percent of Americans. The estimated total cost for HF in 2010 was $39,2 billion, or 1 to 2 percent of all health care expenditures.1 Health care professionals, who face an aging population coupled with the need to be efficient with health care dollars, require sound evidence regarding the diagnosis and management of this disease.

The diagnosis of HF remains a difficult clinical challenge. The diagnosis is based on a constellation of symptoms and signs, supported by objective evidence of impairment of heart function.

B-type natriuretic peptide (BNP) and N-terminal proBNP (NT-proBNP) have emerged as promising markers for HF diagnosis, prognosis, and treatment. These peptides are secreted into the bloodstream by cardiac myocytes in response to increased ventricular wall stress, hypertrophy, and volume overload. Elevated levels of these peptides are evident in persons with HF, and it is well established that a low result can exclude HF.3

Effective Health Care Program

The Effective Health Care Program was initiated in 2005 to provide valid evidence about the comparative effectiveness of different medical interventions. The object is to help consumers, health care providers, and others in making informed choices among treatment alternatives. Through its Comparative Effectiveness Reviews, the program supports systematic appraisals of existing scientific evidence regarding treatments for high-priority health conditions. It also promotes and generates new scientific evidence by identifying gaps in existing scientific evidence and supporting new research. The program puts special emphasis on translating findings into a variety of useful formats for different stakeholders, including consumers.

The full report and this summary are available at www.effectivehealthcare.nih.gov/reports/final.cfm.

Reviews of the prognostic use of BNP and NT-proBNP have shown that these peptides are independent predictors of mortality and other cardiac outcomes in patients with HF.3-7 In addition, the reviews suggest that discharge or post-treatment BNP and NT-proBNP are the
optimal predictors of prognosis compared with BNP or NT-proBNP measured at other points in time. The reviews also found that BNP and NT-proBNP could add useful information to the standard cardiovascular disease (CVD) risk assessment in certain populations.

Optimization of therapy for patients with HF remains challenging due to the difficulty of diagnosing the condition in the absence of clinically evident signs and symptoms. Measurement of BNP or NT-proBNP has been advocated to guide treatment. This approach is taken because the peptides are independently associated with prognosis and their concentrations decrease with effective therapy. It is unclear whether biomarker-assisted therapy (to achieve a concentration below a target value) or intensified therapy (adjustment of therapy based on a change in biomarker concentration) reduces mortality, rehospitalization, or quality of life (QOL) compared with usual care.

Furthermore, knowledge of the variation of a test measure is important when treatment is based on a difference between serial measurements. We do not currently know how much of a difference in BNP or NT-proBNP concentrations is clinically important. Variation in a test measure is a function of the analytical variation of the assay method (bias and precision) and the inherent biological variation of the molecule tested. The biological variation may also be a function of disease severity, sex, medications, and comorbidity.

A comprehensive systematic review of BNP and NT-proBNP was completed in 2006 by the McMaster University Evidence-based Practice Center (EPC) for the Agency for Healthcare Research and Quality (AHRQ). Due to the vast amount of literature published since the last review, the obsolescence of certain assay types used in earlier studies of BNP and NT-proBNP, and new Key Questions (KQs) that account for the evolution of (and continuing uncertainty within) the field, an entirely new systematic review was required to provide an assessment of the “state of the science” in this field. To summarize the current body of scientific knowledge, this review examined the diagnostic, prognostic, and therapeutic use of BNP and NT-proBNP and whether the biological variation of BNP and NT-proBNP differs in HF and non-HF populations.

Key Questions

The Key Questions for our review are as follows:

Key Question 1: In patients presenting to the emergency department or urgent care facilities with signs or symptoms suggestive of heart failure:

a. What is the test performance of BNP and NT-proBNP for HF?
b. What are the optimal decision cutpoints for BNP and NT-proBNP to diagnose and exclude HF?
c. What determinants affect the test performance of BNP and NT-proBNP (e.g., age, gender, comorbidity)?

Key Question 2: In patients presenting to a primary care physician with risk factors, signs, or symptoms suggestive of HF:

a. What is the test performance of BNP and NT-proBNP for HF?
b. What are the optimal decision cutpoints for BNP and NT-proBNP to diagnose and exclude HF?
c. What determinants affect the test performance of BNP and NT-proBNP (e.g., age, gender, comorbidity)?

Key Question 3: In HF populations, is BNP or NT-proBNP measured at admission, discharge, or change between admission and discharge an independent predictor of morbidity and mortality outcomes?

Key Question 4: In HF populations, does BNP measured at admission, discharge, or change between admission and discharge add incremental predictive information to established risk factors for morbidity and mortality outcomes?

Key Question 5: Is BNP or NT-proBNP measured in the community setting an independent predictor of morbidity and mortality outcomes in general populations?

Key Question 6: In patients with HF, does BNP-assisted therapy or intensified therapy improve outcomes compared with usual care?

Key Question 7: What is the biological variation of BNP and NT-proBNP in patients with HF and without HF?

Analytic Framework

To guide this systematic review and facilitate the interpretation of the KQs, we developed an analytic framework (Figure A) that depicts the logical progression and interconnection of all seven KQs.

The analytic framework describes the interconnection among the study questions examining diagnosis, prognosis, therapy, and screening. For diagnosis of patients with signs and symptoms compatible with HF, the two settings are acute care (KQ1) and primary care (KQ2). A third setting is the general, undifferentiated population without overt signs or symptoms of HF (KQ5). KQ5 examines the ability of BNP/NT-proBNP to predict mortality and morbidity.
outcomes in this population. Prognosis of patients with established HF is addressed in KQ3 and KQ4. Prognosis in which the outcome is associated with the concentration of BNP/NT-proBNP is addressed in KQ3, whereas other prognostic measures are dealt with in KQ4. Once a diagnosis of HF has been made, patients are treated. KQ6 examines randomized controlled trials (RCTs) comparing usual care with therapy guided by BNP/NT-proBNP to assess outcome measures. The outcomes to be examined, if reported, include mortality, hospitalization, change in New York Heart Association (NYHA) class, and quality of life. In addition, information on the biological variation of BNP and NT-proBNP was gathered (KQ7).

**Figure A. Analytic framework**

![Analytic framework diagram](Image)

Note: BNP = B-type natriuretic peptide; ED = emergency department; KQ = Key Question; NT-proBNP = N-terminal proBNP; NYHA = New York Heart Association.

**Methods**

**Input From Stakeholders**

The EPC convened a group of experts in the fields of BNP, NT-proBNP, HF, and systematic review methods to form the Technical Expert Panel (TEP). Members of the TEP provided clinical and methodological expertise and input to help interpret the KQs guiding this review, identify important issues, and define parameters for the review of evidence. Discussions among the EPC, the AHRQ Task Order Officer, and the TEP occurred during a series of teleconferences and via email.

The KQs were nominated by a professional society. The KQs were revised for scope and clarity in conjunction with the TEP and the Task Order Officer.

**Search Strategy**

Six databases (Medline®, Embase™, AMED, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, and CINAHL) were searched and results captured for the period from January 1989 to June 2012. Search strategies were adjusted to conform to the parameters of each database. We also reviewed the reference lists of eligible studies during full-text screening and cross-checked all potentially relevant citations with
our citation database. Hand-searching was not done. Gray literature searches included the U.S. Food and Drug Administration (FDA), Health Canada, and European Medicines Agency Web sites; clinical trial registers (clinicaltrials.gov, clinicaltrialsregister.eu, metaRegister of Current Controlled Trials, Clinical Trial Registries, Clinical Study Results, and World Health Organization Clinical Trials); and Conference Papers Index and Scopus (for the previous 2 years only). We limited conference searches to the American Heart Association and the American College of Cardiology conferences.

**Study Selection**

For KQs 1, 2, and 7, the only excluded study design was the case report. For KQs 3 to 5, cross-sectional and case-control studies were excluded. For KQ6, only RCTs were included. In addition, we excluded letters, editorials, commentaries, and conference proceedings. Systematic reviews and meta-analyses were excluded, although their reference lists were examined for potentially relevant citations. Table A shows study selection criteria.

**Data Extraction**

Trained data extractors compiled relevant information from individual studies using standardized forms and a reference guide. During the course of writing the report, investigators reviewed the extracted information for accuracy and made corrections as necessary.

### Table A. Participant selection criteria

<table>
<thead>
<tr>
<th>Category</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Populations</strong></td>
<td>KQs 1–2: Adults presenting to emergency department or urgent care (KQ1) or primary care settings (KQ2) with signs or symptoms consistent with HF. KQs 3–4: Adults with all types of HF. KQ5: Adults in community settings with no disease specified for the study. KQ6: Adults being treated for chronic HF. KQ7: Adults with and without HF.</td>
</tr>
<tr>
<td><strong>Interventions and Prognostic Factors</strong></td>
<td>KQs 1–2: FDA-approved assay for BNP or NT-proBNP at admission or discharge or change in BNP/NT-proBNP between admission and discharge using any cutpoint. KQs 3–4: BNP or NT-proBNP measured at admission or discharge or change between admission and discharge; analysis done by appropriate statistical metrics. KQ5: BNP or NT-proBNP assay using any cutpoint. KQ6: Medical therapy based on BNP or NT-proBNP concentration. KQ7: Multiple measurements of BNP or NT-proBNP per subject.</td>
</tr>
<tr>
<td><strong>Comparators</strong></td>
<td>KQs 1–2: Any method of diagnosing HF that does not use BNP or NT-proBNP. KQs 3–4: NYHA class of HF, ejection fraction, degree of hyponatremia, decreasing peak exercise oxygen uptake, decreasing hematocrit, widened QRS interval on 12-lead ECG, chronic hypotension, resting tachycardia, renal insufficiency, intolerance to conventional therapy, and refractory volume overload, or risk prediction scores. KQ5: Any predictive scoring system. KQ6: Medical therapy based on usual care for HF patients. KQ7: No comparators.</td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td>KQs 1–2: Test performance characteristics (i.e., sensitivity, specificity, positive and negative LR, DOR, and area under ROC curve). KQs 3–6: Mortality, including all cause and HF; morbidity, including hospitalization (HF, all cause, planned, and unplanned); change in NYHA class; and quality of life. Composite outcomes of mortality or morbidity that were not cardiac or HF specific were excluded. KQ7: Calculation of biological variation.</td>
</tr>
</tbody>
</table>
Table A. Participant selection criteria (continued)

<table>
<thead>
<tr>
<th>Category</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timing or Followup</td>
<td>Any length of followup.</td>
</tr>
<tr>
<td>Setting</td>
<td>KQ1: Emergency or urgent care departments only.</td>
</tr>
<tr>
<td></td>
<td>KQ2: Primary care settings only.</td>
</tr>
<tr>
<td></td>
<td>KQs 3–4: Limited to patients admitted to acute care hospitals or recruited from outpatient clinics/ambulatory care settings, hospital settings, or family practice settings.</td>
</tr>
<tr>
<td></td>
<td>KQ5: Primary care (i.e., community or family practice or equivalent).</td>
</tr>
<tr>
<td></td>
<td>KQs 6–7: No restriction on inclusion of articles based on setting.</td>
</tr>
</tbody>
</table>

Note: BNP = B-type natriuretic peptide; DOR = diagnostic odds ratio; ECG = electrocardiogram; FDA = U.S. Food and Drug Administration; HF = heart failure; KQ = Key Question; LR = likelihood ratio; NT-proBNP = N-terminal proBNP; NYHA = New York Heart Association; ROC=receiver operating characteristic.

Assessment of Risk of Bias

To assess the risk of bias for individual studies, we followed the methods recommended by AHRQ’s “Methods Guide for Effectiveness and Comparative Effectiveness Reviews” (Methods Guide)9 and “Methods Guide for Medical Test Reviews.”10 A single rater assessed each study using prescribed tools, clear decision rules, and standardized forms. Piloting of the standardized guide, followed by discussion among the raters, ensured clarity and consistency across raters.

A number of published systems were adapted for use, depending on the study design and the type of analysis. For observational studies, the Newcastle-Ottawa Scale was used;11 for RCTs, the Jadad scale;12 for prognosis studies, a modified version of the guidelines proposed by Hayden et al.;13 and for diagnosis, the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies-2).14 All modifications and instruments used can be found in the full report.

Data Synthesis

We present study results in four key sections based on diagnosis (KQs 1 and 2), prognosis (KQs 3 to 5), treatment (KQ6), and biological variation (KQ7). All included studies are summarized in narrative form and in summary tables in the full report.

Meta-analysis was carried out only for KQs 1 and 2. Two-by-two contingency tables were created for each study where true positive, false positive, false negative, and true negative could be estimated. Sensitivity and specificity, diagnostic odds ratio, and likelihood ratios with 95% confidence intervals were recalculated for each primary study from the contingency tables. Extracted data were pooled using exact binomial rendition15 of the bivariate mixed-effects regression model developed by van Houwelingen16,17 and modified for synthesis of diagnostic test data.18 The bivariate regression model fits a two-level model, with independent binomial distributions in each study and a bivariate normal model for the logit transforms between studies. Summary sensitivity, specificity, and the corresponding positive likelihood, negative likelihood, and diagnostic odds ratios are derived as functions of the estimated model parameters. This approach corresponds to the empirical Bayesian approach to fitting the hierarchical summary receiver operating characteristic (HSROC) model.19 Initial analyses considered the level of statistical heterogeneity across the individual studies that were included in the meta-analysis. The Cochran’s Q test was used as a measure of statistical heterogeneity in all the meta-analyses and the I² as a measure of inconsistency.20

Evaluating the Strength of the Evidence

Evaluating the strength of the body of evidence was conducted according to the Methods Guide9 and “Methods Guide for Medical Test Reviews.”10 We graded the strength of evidence (SOE) for KQs1 and 2 (outcomes of sensitivity and specificity) and KQ6 (death, all cause). We omitted KQs 3 to 5 because criteria to evaluate and score prognostic studies have not been fully developed.10 We also omitted KQ7 because it asks about biological variation rather than a clinical or diagnostic outcome.
The following strength ratings were used:

- High: High confidence that the evidence reflects the true effect. Further research is very unlikely to change our confidence in the estimate of effect.
- Moderate: Moderate confidence that the evidence reflects the true effect. Further research may change our confidence in the estimate of effect and may change the estimate.
- Low: Low confidence that the evidence reflects the true effect. Further research is likely to change the confidence in the estimate effect and is likely to change the estimate.
- Insufficient: Evidence either is unavailable or does not permit a conclusion.

Results

Results of Literature Search

Results of the review are organized by KQ. The full report includes evidence and summary tables showing findings from individual studies for each KQ.

The search yielded 25,864 records identified from six bibliographic databases. An additional 35 records were identified from three gray literature sources: regulatory agency Web sites, clinical trial databases, and conference sources. After duplicates were removed, a total of 16,893 records were screened at the title-and-abstract level; a total of 3,616 citations moved on to be screened at full text.

Following the application of full-text screening criteria, 310 papers were eligible for all research questions in this review.

A total of 104 papers were allocated for diagnostic accuracy. From these, 76 articles were evaluated for KQ1 and 28 for KQ2. For KQ3, KQ4, and KQ5, 190 unique articles were eligible to address the research questions related to prognosis; of these, 183 were eligible for KQ3, 22 for KQ4, and 7 for KQ5. A total of nine articles were evaluated for treatment guided by BNP or NT-proBNP for KQ6. Seven articles for KQ7 focused on biological variation.

Key Question 1: In patients presenting to the emergency department or urgent care facilities with signs or symptoms suggestive of heart failure:

- What is the test performance of BNP and NT-proBNP for HF?
- What are the optimal decision cutpoints for BNP and NT-proBNP to diagnose and exclude HF?
- What determinants affect the test performance of BNP and NT-proBNP (e.g., age, gender, comorbidity)?

**BNP**

Fifty-one publications met the criteria for KQ1 and examined cutpoints for BNP. Two of these papers were RCTs, 9 were cohort studies, and the remaining 40 were cross-sectional studies.

**Test Performance and Optimal Decision Cutpoints.**

Papers reporting information on the lowest cutpoint presented by the authors returned a pooled estimate for sensitivity of 95 percent (95% confidence interval [CI], 93 to 97%) and a pooled estimate for specificity of 67 percent (95% CI, 58 to 75%). Twenty-one papers reported on the manufacturers’ suggested cutpoint of 100 pg/mL, resulting in a pooled estimate for sensitivity of 95 percent (95% CI, 93 to 96%) and for specificity of 66 percent (95% CI, 56 to 74%).

Twenty-eight papers examined an optimal cutpoint, which was defined using various definitions, such as the cutpoint that would maximize accuracy. The pooled estimate for sensitivity was 91 percent (95% CI, 88 to 94%) and for specificity was 80 percent (95% CI, 74 to 85%). Using the optimal cutpoint resulted in a higher overall estimate of the positive likelihood ratio (LR+) of 4.61 (95% CI, 3.49 to 6.09) compared with either the lowest cutpoint (2.85; 95% CI, 2.23 to 3.65) or the manufacturers’ suggested cutpoint (2.76; 95% CI, 2.12 to 3.59). The negative likelihood ratio (LR-) was not statistically significantly different (p >0.05).

Choosing the lowest cutpoint, the manufacturers’ suggested cutpoint, or the optimal cutpoint had little effect on the diagnostic performance of the test. The test displayed high sensitivity and a high LR-, but low specificity and low LR+.

**Determinants Affecting Test Performance. Age:** Eight articles found increasing age to be associated with increased BNP concentrations, but the effect on the diagnostic performance of the test was not clear in the papers.

**Sex:** Maisel et al. reported that the difference in BNP concentrations between men and women was not significant. Conversely, Knudsen et al. noted differences in sensitivity and specificity between males and females using 100 pg/mL as the decision point (males: sensitivity 94.3%, specificity 54.9%; females: sensitivity 90.0%, specificity 55.2%).
Ethnicity: Maisel et al.\textsuperscript{22} reported that the prevalence of HF in their study population was significantly greater among whites than among African Americans. Similarly, the mean concentration of BNP was significantly greater in the white population with HF than in the African American population with HF (200 vs. 117 pg/mL; \(p < 0.001\)).

Obesity: Three papers\textsuperscript{41,59,60} showed that increasing body mass index (BMI) was inversely associated with BNP concentrations. This finding was consistent whether BMI and BNP were examined in the whole population\textsuperscript{59,60} or the population was examined in two groups, namely those with or without HF.\textsuperscript{41}

Renal function: Four\textsuperscript{42,48,51,67} articles examined estimated glomerular filtration rate (eGFR), and one\textsuperscript{59} examined serum creatinine concentration. The BNP concentration was inversely related to renal function. As eGFR decreased or creatinine concentration increased, the BNP concentration increased.

Diabetes: One study\textsuperscript{34} reported a nonsignificant difference in areas under the curve (AUCs) calculated for patients with or without diabetes. AUC was 0.878 (95% CI, 0.837 to 0.913) for patients with diabetes and 0.888 (95% CI, 0.860 to 0.912) for patients without diabetes.

**NT-proBNP**

Thirty-nine articles met the criteria for KQ1 and examined NT-proBNP\textsuperscript{25,38,42,45-48,51,55,61,63,64,66,67,69,72-95}. Eleven papers were prospective cohort studies,\textsuperscript{61,63,64,66,67,69,85,86,90,94,95} one was a case-control study,\textsuperscript{81} and the study design could not be determined in two papers.\textsuperscript{82,92} The remaining papers (n = 25) used a cross-sectional design.

**Test Performance and Optimal Decision Cutpoints.** The 39 papers evaluating NT-proBNP in the emergency department used several cutpoints, ranging from 100\textsuperscript{88} to 6,550\textsuperscript{42} pg/mL or ng/L. Reported sensitivities ranged from 53 percent\textsuperscript{47} to 100 percent\textsuperscript{38,47,51,76} (mean = 85.1%; median = 88%); specificities from 5 percent\textsuperscript{47} to 100 percent\textsuperscript{48} (mean = 70.9%; median = 73.2%); LR+ from 1.05\textsuperscript{47} to 115.03\textsuperscript{38} and LR- from 0.02\textsuperscript{38,51} to 0.35\textsuperscript{66} AUCs ranged from 0.6\textsuperscript{61} to 0.99\textsuperscript{72} (mean = 0.88; median = 0.89).

**Determinants Affecting Test Performance.** Age: The effect of age-optimized cutpoints was unclear. Some articles suggested improved test performance with age-optimized cutpoints and others did not.

Race and sex: Krauser et al.\textsuperscript{76} reported that the area under the receiver operating characteristic (ROC) curve was not different for men versus women or for African Americans versus others. There was no difference in the median NT-proBNP concentration between men and women or between African Americans and others.

**Obesity:** A single paper\textsuperscript{74} concluded that BMI-adjusted cutpoints performed well over a wide variety of BMIs. Despite lower sensitivity at the high range of BMI, the predictive values were unchanged.

**Renal function:** Two papers\textsuperscript{48,80} reported an inverse association between renal function and NT-proBNP concentration.

**Key Question 2: In patients presenting to a primary care physician with risk factors, signs, or symptoms suggestive of HF:**

a. What is the test performance of BNP and NT-proBNP for HF?

b. What are the optimal decision cutpoints for BNP and NT-proBNP to diagnose and exclude HF?

c. What determinants affect the test performance of BNP and NT-proBNP (e.g., age, gender, comorbidity)?

**BNP**

Twelve articles met the criteria for this KQ.\textsuperscript{96-107} One study used a prospective cohort design,\textsuperscript{103} and the remaining studies (n = 11) used a cross-sectional design.

**Test Performance and Optimal Decision Cutpoints.** Three cutpoints were selected: lowest presented, manufacturers’ suggested, and the optimal cutpoint as chosen by the authors. The pooled sensitivity using the optimal cutpoint was 82 percent (95% CI, 69 to 90%), and the pooled specificity was 64 percent (95% CI, 45 to 79%). Summary LR+ and LR- were 2.27 (95% CI, 1.43 to 3.62) and 0.28 (95% CI, 0.16 to 0.49), respectively.

Pooling using the lowest cutpoint produced slightly higher sensitivity and correspondingly lower specificity: 89 percent (95% CI, 77 to 93%) and 54 percent (95% CI, 41 to 66%), respectively. The LR+ and LR- gave similar results: 1.94 (95% CI, 1.47 to 2.57) and 0.20 (95% CI, 0.09 to 0.44), respectively.

The pooled sensitivity of 76 percent (95% CI, 59 to 87%) based on the manufacturers’ cutpoint of 100 pg/mL was lower than that for the optimal cutpoint. Corresponding specificity was increased to 71 percent (95% CI, 52 to 85%), compared with 64 percent for the optimal cutpoint. The LR+ and LR- gave results similar to those for the optimal cutpoint: 2.63 (95% CI, 1.59 to 4.36) and 0.34 (95% CI, 0.20 to 0.57), respectively.
**Determinants Affecting Test Performance.** 

**Age:** A single study examined the effect of age on BNP. A higher cutpoint was required in older patients (≥65 years) than in younger patients (<65 years) to detect left ventricular ejection fraction (LVEF) <45 (250 vs. 82 pg/mL) and advanced diastolic dysfunction (DD) (236 vs. 70 pg/mL).

**Sex:** Test performance did not show statistically significant sex differences in a study by Fuat et al. in which the AUC was 0.79 for men and 0.80 for women. In a study by Park et al., for patients with LVEF <45, the AUC was 0.89 for men and 0.93 for women; for patients with advanced DD, the AUC was 0.89 for men and 0.91 for women.

**BMI:** An inverse correlation of BNP with BMI was shown in one study: AUCs for diagnosis of decompensated HF were 0.78 (95% CI, 0.71 to 0.84) for normal-weight patients; 0.72 (95% CI, 0.66 to 0.79) for overweight patients; and 0.62 (95% CI, 0.54 to 0.70) for obese patients. For detecting LVEF <45 in another study, the AUC was 0.93 in patients ≥25 kg/m² (cutpoint, 151 pg/mL; sensitivity, 85%; specificity, 85%) and 0.90 in patients <25 kg/m² (cutpoint, 154 pg/mL; sensitivity and specificity, 81%). For detecting advanced DD, the AUC was 0.84 in patients ≥25 kg/m² (cutpoint, 82 pg/mL; sensitivity and specificity, 80%) and 0.92 in patients <25 kg/m² (cutpoint, 140 pg/mL; sensitivity and specificity, 83%).

**Renal function:** One study assessed the effect of renal function on test performance. Patients were grouped by clearance rates (≥60 mL/min and <60 mL/min). For detecting LVEF <45, AUC estimates were 0.92 (cutpoint, 89 pg/mL; sensitivity and specificity, 82%) for clearance rates ≥60 mL/min and 0.87 (cutpoint, 264 pg/mL; sensitivity and specificity, 78%) for clearance rates <60 mL/min. For detecting advanced DD, AUC estimates were 0.89 (cutpoint, 70 pg/mL; sensitivity, 83%; specificity, 82%) for clearance rates ≥60 mL/min and 0.88 (cutpoint, 247 pg/mL; sensitivity and specificity, 78%) for clearance rates <60 mL/min.

**NT-proBNP**

Twenty articles met the criteria for KQ2 examining NT-proBNP in primary care settings. Two studies used a prospective cohort design. Study design could not be determined in one of the articles. The remaining studies (n = 17) used a cross-sectional design. The 19 studies evaluating NT-proBNP in primary care settings used several cutpoints ranging from 25 pg/mL to 6,180 ng/L (mean = 635; median = 379).

**Test Performance and Optimal Decision Cutpoints.**

Three cutpoints were selected: lowest presented, the optimal cutpoint as chosen by the authors, and the manufacturers’ recommended cutpoint of 125 pg/mL for patients <75 years of age and 450 pg/mL for patients ≥75 years of age. When the optimal cutpoint chosen by the authors was used, the pooled sensitivity was 0.88 (95% CI, 0.81 to 0.93), and seven of the studies produced sensitivities greater than 0.90.

Choosing the lowest cutpoint selected by the authors produced increased pooled sensitivity when compared with the optimal cutpoint, with no decrease in pooled specificity. All but three studies produced sensitivities greater than 0.90.

It was determined that at least four studies were needed in each group to present summary estimates; however, only two studies satisfied our criteria for NT-proBNP according to manufacturers’ cutpoint, and thus they were not presented.

**Determinants Affecting Test Performance.**

**Age:** Two studies investigated the influence of age on the diagnostic ability of NT-proBNP. As was seen in the studies of BNP, the optimal cutpoint was higher in older patients. For detecting LVEF <45 in one study, the AUC was 0.88 in patients ≥65 years (cutpoint 1,446 pg/mL; sensitivity 82%; specificity 81%) and 0.91 in patients <65 years (cutpoint, 379 pg/mL; sensitivity and specificity, 83%). One study determined optimal cutpoints of 1,446 pg/mL for those >65 years and 379 pg/mL for those <65. A second study determined cutpoints of 652 pg/mL for those >75 years and 357 pg/mL for those <75 years.

**Sex:** Five studies investigated the relationship between sex and NT-proBNP’s ability to diagnose HF. Using optimal AUC analysis, a range of different cutpoints can be established for men and women. Typically the optimized cutpoint for men was lower than that for women.

**BMI:** Two studies examined the relationship between NT-proBNP and BMI. One study showed an inverse correlation of NT-proBNP with BMI.

**Renal function:** One study examined the effect of renal function on the ability of NT-proBNP to identify patients with LVEF <45 and advanced DD. The optimized cutpoints were higher with lower creatinine clearance.
Strength of Evidence for BNP and NT-proBNP for All Cutpoints in KQ1 and KQ2

Risk of Bias
Using the QUADAS-2 tool, we rated the risk of bias for both sensitivity and specificity. In the four domains (patient selection, index test, reference standard, and flow and timing), the risk of bias was rated as low.

Directness
KQ1 and KQ2 pertain to diagnostic accuracy and assessment of sensitivity and specificity. These concepts are well understood by clinicians and can be applied in a clinical setting, so we rate this domain as direct.

Precision
For both BNP and NT-proBNP, the CIs around the summary estimates for sensitivity and specificity are not precise. We rate this domain as imprecise.

Consistency
In terms of BNP sensitivity, the directions of the estimates are consistent, and with the exception of a single study,¹⁰⁵ are very similar. In terms of NT-proBNP sensitivity, the directions of the estimates are consistent and the CIs are small. Therefore, we rate this domain as consistent for both BNP and NT-proBNP. However, we rate the specificity as inconsistent because the range of estimates across studies for both BNP and NT-proBNP is large.

The overall SOE estimate for both BNP and NT-proBNP in emergency department and primary care settings is high for sensitivity and moderate for specificity.

Key Question 3: In HF populations, is BNP or NT-proBNP measured at admission, discharge, or change between admission and discharge an independent predictor of morbidity and mortality outcomes?

Patients With Decompensated Heart Failure
Seventy-nine publications (cohorts, case series, and RCTs) evaluated concentrations of BNP (n = 38), NT-proBNP (n = 35), or both (n = 6) as predictors of mortality and morbidity outcomes. Subjects were recruited from emergency or inpatient acute care centers. The majority of studies (n = 55) assessed BNP and NT-proBNP concentrations at admission, with fewer studies evaluating serial measurements while hospitalized (n = 4) or concentrations at hospital discharge (n = 21) as potential prognostic factors. Additionally, the majority of studies (n = 50) evaluated all-cause mortality and composite outcomes; cardiovascular mortality and morbidity outcomes were measured less frequently. In general, higher concentrations of admission BNP and NT-proBNP were predictive of outcomes of mortality and morbidity, but the range of thresholds for “high” varied markedly across studies. Similarly, for the studies evaluating BNP at discharge, a decrease in BNP concentrations was protective of subsequent mortality and morbidity. Four studies evaluated serial measurements during hospitalization and showed that higher BNP concentrations after admission could also predict mortality. Overall, we judge this body of evidence to be at moderate risk of bias because of the uncertainty with respect to the validity and reliability of the methods used to ascertain the outcome, confounding (inconsistent adjustment for age, sex, BMI, and renal function), and inappropriate statistical analyses (poorly reported).

Generally, studies predicting short-term mortality (up to 31 days) and longer term mortality (24 months or greater) were few in number. Most studies evaluated medium-range time intervals (6 to 12 months), and they consistently showed that BNP or NT-proBNP concentrations are independent predictors of all-cause and cardiovascular mortality, morbidity, and composite outcomes. This was shown across studies for both BNP and NT-proBNP despite the variations in the factors included within the statistical models, including different cutpoints (when used as dichotomous data), other potential prognostic factors included in the statistical models, and time intervals. Conversely, the challenge with these differing study factors was in interpreting the magnitude of the predictive values across studies. Far fewer studies evaluated longer term prognosis (>12 months), and these studies measured admission, discharge, or change from admission concentrations, further limiting the comparisons.

Patients With Chronic Stable Heart Failure
One hundred four publications (cohorts, case series, and RCTs) at moderate risk of bias evaluated concentrations of BNP (n = 15), NT-proBNP (n = 88), or both (n = 1) as predictors of mortality and morbidity in patients with chronic stable HF. In patients with chronic stable HF, there is an association between BNP and the outcome of all-cause mortality. The other mortality outcomes (i.e., cardiac and sudden cardiac death) demonstrated less convincing associations. The importance of BNP as an independent predictor appears to correlate with severity of HF and possibly length of followup. The outcome of hospitalization and the composite outcome of all-cause mortality and cardiovascular morbidity demonstrated a significant independent association with BNP.

Eighty-eight publications evaluated NT-proBNP levels as predictors of mortality and morbidity in patients with chronic stable HF. Overall, the evidence consistently
supports the trend that NT-proBNP is an independent predictor of mortality and morbidity outcomes in people with chronic stable HF. The applicability of these results in chronic stable HF patients rests largely in middle-aged or elderly males. The included studies did not explore whether the prognostic effects of NT-proBNP differ by age, sex, or time period. Also, the studies did not suggest a single cutpoint to optimize the prognostic ability of the peptide. In general, the studies were not consistent with respect to measuring the outcome and including our predefined set of variables in the analysis. The largest number of studies and the strongest evidence concerned the outcome of all-cause mortality. Fifty-two publications included all-cause mortality as an outcome, and all of the point estimates measuring association indicated positive associations between NT-proBNP and all-cause mortality. This conclusion applies across all periods of followup, from 12 months to 44 months. For cardiovascular mortality, the evidence in 17 publications also suggests a positive association with NT-proBNP.

For morbidity outcomes (n=12), we found some evidence to suggest that higher concentrations of NT-proBNP predict hospitalization. Twenty-six publications evaluated composite outcomes and showed that NT-proBNP is an independent predictor; the results also suggest that higher levels of NT-proBNP predicted greater numbers of composite events.

**Patients With Decompensated Heart Failure Having Surgical Procedures**

To predict subsequent outcomes, six studies at low risk of bias evaluated BNP levels measured prior to or during cardiac resynchronization therapy (n=4), cardiac resynchronization defibrillation therapy (n=1), and noncardiac surgery (n=1) in stable HF patients, as well as in patients undergoing peritoneal dialysis (n=1) with decompensated HF. All except the peritoneal dialysis study showed that higher BNP levels were associated with subsequent mortality and morbidity.

Three publications evaluated NT-proBNP levels in stable HF patients undergoing cardiac resynchronization therapy (n=2) and intracoronary infusion of bone marrow–derived mononuclear progenitor cells (n=1). All studies (for both types of surgeries) showed that higher NT-proBNP levels were associated with subsequent mortality.

**Key Question 4: In HF populations, does BNP measured at admission, discharge, or change between admission and discharge add incremental predictive information to established risk factors for morbidity and mortality outcomes?**

Of 183 studies eligible for KQ3, 39 publications used methods that would allow assessment of the incremental value of adding BNP or NT-proBNP when predicting subsequent outcomes (KQ4). Of these 39 publications, 2 studies79,123 reported that they undertook statistical computations yet did not present any data for incremental value. Additionally, 15 studies included BNP in the base prognostic model,71,124-127 NT-proBNP in the base prognostic model,128-136 or both assays in the base model.137 Including these assays in the base model does not allow for the assessment of the predictive incremental value of BNP/NT-proBNP. The study findings from the remaining 22 publications that provided the appropriate computations to assess incremental value are presented below.

**Patients With Decompensated Heart Failure**

Seven publications (six studies) included patients with decompensated HF and evaluated the incremental value of admission BNP53,138-141 or admission NT-proBNP;142,143 one study53 evaluated both BNP and NT-proBNP but reported results only for BNP. Two publications138,139 pertaining to one study contained overlapping cohorts of consecutive patients recruited from the same center because the study was ongoing and more patients were added to the database; we report findings from both publications even though the cohorts overlap and the publications are considered to be from a single study.

**Added Value of BNP to Prognostic Risk Prediction.**

Data from five studies53,138-141 suggest that there may be differences in risk prediction by type of mortality outcome (all cause, cardiovascular) in decompensated HF patients. Risk prediction improved incrementally when admission BNP was added to the predictive models that did not contain other markers, despite differences in the models and lengths of followup (which varied from 31 days to 12 months). In some cases, risk prediction improved further when BNP was combined with other markers such as carbohydrate antigen 125 (CA125)138 or midregional proadrenomedullin (MR-proADM).53

**Added Value of NT-proBNP to Prognostic Risk Prediction.** One study142 of acutely ill patients with HF reported that the inclusion of NT-proBNP alone to a base model failed to show a statistically significant improvement in risk prediction. Conversely, statistically significant improvement was shown when NT-proBNP was combined with other markers in the form of a multimarker risk score based on optimal cutpoints (ROC analysis). Two other studies79,123 claimed to look at this issue yet did not report any results.
**Patients With Stable Heart Failure**

**Added Value of BNP to Prognostic Risk Prediction.** No studies evaluated the incremental predictive value of using BNP as a prognostic risk predictor in stable HF patients.

**Added Value of NT-proBNP to Prognostic Risk Prediction.** Fifteen publications evaluating patients with chronic stable HF considered the prognostic value of NT-proBNP. Overall, NT-proBNP demonstrated incremental predictive value in mortality outcomes, with some evidence suggesting that the incremental value might be more evident in cardiovascular versus all-cause mortality. In one cardiovascular mortality study, the addition of NT-proBNP to the base model resulted in better discrimination for risk prediction than the addition of C-terminal endothelin (CT-proET) (c-statistic = 0.78 vs. 0.77), although the highest value of discrimination was achieved when both NT-proBNP and CT-proET were added to the base model at the same time (c-statistic = 0.79). For all-cause mortality, the base model (clinical variables) with NT-proBNP had a higher discriminatory ability than the base model without NT-proBNP (c-statistic = 0.74 vs. 0.70). The study data also showed that for all-cause mortality, the discriminatory ability for risk prediction was improved by adding copeptin to the model with clinical variables and NT-proBNP (c-statistic = 0.76).

**Key Question 5: Is BNP or NT-proBNP measured in the community setting an independent predictor of morbidity and mortality outcomes in general populations?**

Seven studies were eligible for inclusion in this section of the systematic review. A total of 15,656 individuals were included in the seven studies. The smallest study included 274 individuals and the largest 5,447. The length of followup ranged from 3.5 to 13.8 years. All seven studies measured NT-proBNP. No studies used BNP, and this has been identified as a research gap.

**Mortality**

All-cause mortality was the outcome in three studies, and in all three there was an increasing adjusted hazard ratio (HR) with increasing NT-proBNP measured by tertiles, by increases of 1 standard deviation (SD) unit, and by log(NT-proBNP). The relationship between baseline NT-proBNP and all-cause mortality appeared to be log-linear in nature.

Sudden cardiac death had increasing HRs across the quintiles of NT-proBNP and an adjusted HR = 1.9 (95% CI, 1.7 to 2.1) for ln-NT-proBNP.

Cardiovascular death had a significant adjusted HR for log(NT-proBNP)/SD and log(NT-proBNP). A cutpoint of 100 pg/mL was applied to one population, and results showed an adjusted HR = 1.0 (95% CI, 1.0 to 1.001). However, in a model that was adjusted for known baseline CVD, the adjusted HR became nonsignificant (HR=1.61; 95% CI, 0.79 to 3.28).

**Morbidity**

Onset of atrial fibrillation (AF) was associated with ln-NT-proBNP in a model including conventional risk factors (adjusted HR = 1.45; 95% CI, 1.28 to 1.65) but not in a model that included midregional pro-atrial natriuretic peptide and c-reactive protein. Onset of incident HF was associated with ln-NT-proBNP in models that included other markers of cardiac risk.

**Key Question 6: In patients with HF, does BNP-assisted therapy or intensified therapy improve outcomes compared with usual care?**

Nine RCTs examined whether patients whose treatment for HF was guided by BNP or NT-proBNP displayed improved outcomes compared with patients treated for HF with usual care only. The term “usual care” encompassed standard of care, clinically guided care, symptom-guided care, or control group. One study used a congestion score strategy compared with BNP-guided therapy. Another study was a three-arm trial with an additional multidisciplinary group, but only the usual-care and NT-proBNP arms are included in this systematic review. There were 7 multicenter studies, including 3 to 45 sites with a minimum of 41 patients up to a maximum of 499 patients. The total number of patients included for all nine studies was 2,104. Four studies measured BNP, and five studies measured NT-proBNP. The risk of bias for the nine studies was low. Meta-analyses were not performed because of the substantial heterogeneity among the studies, and therefore no quantitative summary estimates could be made.

**Primary Endpoint**

A composite of endpoints was used in six studies, two studies used only one endpoint, and one study did not define a primary endpoint. Patients in the BNP/NT-proBNP group had fewer events compared with the usual-care group in three studies. The other studies showed no difference in the primary endpoint between treatment groups.
Clinic Visits
Clinic visits were reported in only two studies,168,169 of which one, but not the other, reported more visits for the BNP/NT-proBNP group than the usual-care group.168

Hospitalizations
Admissions were considered all cause unless otherwise specified. All studies except one174 reported on some parameter related to admissions. Most studies reported on cardiovascular admissions, and three studies168,170,173 reported fewer admissions in the BNP/NT-proBNP group than the usual-care group. The other studies had no difference in admissions between groups.

Deaths
Of the seven studies that reported on deaths, six reported all-cause mortality,167-169,171,173,175 four reported death due to a cardiovascular cause,170,171,173,175 and only two studies reported on death related to HF.173,175 The SOE was assessed using the single outcome of mortality. Relative risks, confidence intervals, and SOE are presented in Table B. Overall the SOE was rated as low, as two domains (consistency and precision) were not met. Future research is likely to change the magnitude and direction of the effects for the outcome of all-cause mortality.

### Table B. Strength of evidence for studies evaluating the benefit of therapy guided by BNP and NT-proBNP compared with usual care on all-cause mortality in patients with HF

<table>
<thead>
<tr>
<th>Design</th>
<th>Risk of Biasa</th>
<th>Consistency</th>
<th>Directness</th>
<th>Precision</th>
<th>Effect Size, RR (95% CI)</th>
<th>Strength of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCT</td>
<td>Low</td>
<td>Inconsistent (5 studies with no effect and 2 studies with a lower RR)</td>
<td>Direct</td>
<td>Imprecise (Unable to assess if the studies were adequately powered and the overall event rates were variable because of length of followup)</td>
<td>Beck–daSilva,167 2005: 0.48 (0.05 to 4.85) &lt;br&gt; Berger,168 2010: 0.56 (0.35 to 0.89) &lt;br&gt; PRIMA,169 2001: 0.79 (0.57 to 1.10) &lt;br&gt; STARS-BNP,173 2007: 0.64 (0.26 to 1.58) &lt;br&gt; UPSTEP,175 2011: 0.96 (0.61 to 1.50) &lt;br&gt; SIGNAL-HF,171 2010: 0.98 (0.36 to 2.72) &lt;br&gt; TIME-CHF,174 2009: 0.65 (0.52 to 0.81)</td>
<td>The strength of evidence was rated as low. Therapy guided by BNP/NT-proBNP, when compared with usual care, reduced all-cause mortality. Future research is likely to change the magnitude and direction of the effects for the outcome of all-cause mortality.</td>
</tr>
</tbody>
</table>

*aModified Jadad scale.

Note: BNP = B-type natriuretic peptide; ED = emergency department; CI = confidence interval; NT-proBNP = N-terminal proBNP; PRIMA = PRo-brain-natriuretic peptide guided therapy of chronic heart failure IMprove heart failure morbidity and mortality; RCT = randomized controlled trial; RR = relative risk; SIGNAL-HF = Swedish Intervention study – Guidelines and NT-proBNP AnaLysis in Heart Failure; STARS-BNP = Suivi du Traitement dans l-insuffisAnce caRdiaque Systolique-BNP; TIME-CHF = Trial of Intensified vs standard Medical therapy in Elderly patients with Congestive Heart Failure; UPSTEP = Use of PeptideS in Tailoring hEart failure Project.
Days Alive
Data on days alive, as opposed to death data, were captured in five studies.169,171-174 Two studies173,174 showed that patients in the BNP/NT-proBNP group had more days of survival outside the hospital than the usual-care group. The other studies showed no difference between groups.

Quality of Life
Three studies included a QOL questionnaire.167,171,174 One study167 used the Kansas City Cardiomyopathy Questionnaire (KCCQ) and showed improvement in score in the BNP/NT-proBNP group compared with the usual-care group. The other two studies used different QOL questionnaires and did not show a difference between groups.

Other Parameters
Studies also reported on acute coronary syndrome,170 cerebral ischemia,170 significant ventricular arrhythmia,170 a combined endpoint of time to cardiovascular death or cardiovascular hospitalization,171 congestion score,171 and worsening of HF.170,176 Only one parameter, worsening HF (new worsening symptoms and signs of HF requiring unplanned intensification of decongestive therapy), was different in the BNP/NT-proBNP group compared with the usual-care group. The study showed fewer events in the BNP/NT-proBNP group.170

Medications
Medication use was reported in all nine studies. Of the studies that showed differences in use between the BNP/NT-proBNP group and the usual-care group, most showed increased use in the BNP/NT-proBNP group. These included aldosterone antagonists (AA) in one170 of three studies,169,170,175 angiotensin-converting enzyme (ACE-I) in one172 of four studies,170-172,175 ACE-I or angiotensin receptor blockers (ARB) in four168,169,172,174 of five studies,168,169,171,172,174 ACE-I or ARB and beta-blocker in two172,177 of three studies,168,172,177 beta-blocker in two168,174 of eight studies,168-175 and spironolactone in one174 of three studies.168,172,174

Medication decreases were found for diuretics (two168,170 of six studies168-172,175) and ARB (one170 of five studies168-171,175) in the BNP/NT-proBNP group compared with the usual care group. No differences between BNP/NT-proBNP and usual-care groups were found for ACE-I and AA,171 ACE-I plus ARB and AA,171 digoxin,168,171 or nitrates.168,170

Key Question 7: What is the biological variation of BNP and NT-proBNP in patients with HF and without HF?
Seven studies included data on biological variation for BNP and NT-proBNP.178-182 All study designs were prospective cohort studies except for one that was a retrospective chart review.182 Studies varied in length from as short as 1 day to as long as 2 years. Overall, the number of patients or participants sampled was small (mean = 32; range = 5 to 78), as were the samples obtained to calculate biological variation (median = 4; range = 2 to 15). Blood collection parameters and analytical protocols varied among studies and were inconsistently reported.

The analytical coefficient of variation (CVa) values, or assay imprecision, for BNP were lowest for the Bayer Centaur method (1.8% to 4%) and highest for the Biosite Triage (8.6% to 13.7%), reflecting the higher imprecision for point-of-care devices. Similar CVa values were obtained for the Roche NT-proBNP method (1.4% to 3.0%). Review of the within-individual variation values (CVi) for BNP and NT-proBNP in patients with HF or healthy controls showed lower values (by about one-half) for within-hour180 and within-day178 values than for values from longer time intervals (1 to 12 weeks). Within-individual variation was similar for BNP (median = 25%) and NT-proBNP (median = 20%).

The relative change value (RCV) is a parameter that constitutes a clinically meaningful change in serial results. The largest RCV values were found for healthy individuals for BNP (123% and 139% for two different methods) and NT-proBNP (92%).183 The only other study with an RCV value for healthy individuals measured NT-proBNP and reported a much lower value (26%), but this value was log-transformed.184 For patients with HF, the RCV values were overall higher for BNP (32% to 113%) than for NT-proBNP (16% to 55%). In studies178,180,181 that analyzed both BNP and NT-proBNP, the RCV was lower for NT-proBNP, mostly as a function of the lower CVa for the method compared with the BNP methods.

The index of individuality (IOI) is a useful parameter for assessing the degree of individuality for a biomarker and was assessed in four studies.179,181,183,184 The IOI for NT-proBNP in healthy individuals (0.64 and 0.90) was higher than for patients with HF (0.03 and 0.11). Similarly, the IOI for BNP was higher in healthy individuals (1.1 and 1.8; same patients but different methods) than for patients with HF (0.14). This means there is more individuality for BNP and NT-proBNP in patients with HF than in healthy individuals.
Discussion

Diagnostic Studies (Key Questions 1 and 2)

Key Findings for Emergency Settings
For patients who present to emergency departments or urgent care settings with signs and symptoms suggestive of HF, BNP and NT-proBNP have good diagnostic performance to rule out, but lesser performance to rule in, the diagnosis of HF compared with the reference standard of global assessment of the patient’s medical record. Covariates, especially age and renal function, have important effects on the performance of these tests. However, the findings about the effects of age were equivocal, with some studies reporting effects and others not.

Key Findings for Primary Care Settings
This review indicates that BNP and NT-proBNP are useful diagnostic tools to identify patients with HF in primary care settings, with pooled sensitivities ranging from 0.77 to 0.84 for BNP and 0.86 to 0.90 for NT-proBNP, depending on the cutpoint. Both BNP and NT-proBNP have good diagnostic performance in primary care settings for identifying patients who are either at risk of developing HF or have limited symptoms suggestive of HF. Using the manufacturers’ suggested cutpoint, BNP can effectively be used to rule out the presence of HF in primary care settings. In the case of NT-proBNP, limited evidence is available to determine if the manufacturers’ suggested cutpoint is as effective. Only one study evaluated the cutpoints recommended by the European Society of Cardiology. A single study looked at the age effect and showed that a higher cutpoint is required for both BNP and NT-proBNP in patients aged 65 years and older to maintain test sensitivity equivalent to that for patients less than 65 years. No sex differences were seen for BNP, and no clear conclusions could be drawn regarding optimal cutpoints for NT-proBNP in males and females. A negative correlation of BMI with BNP or NT-proBNP was reported, with decreasing sensitivities for diagnosing HF. However, no BMI-specific cutpoints were suggested in the included articles. Decreased renal function, measured by creatinine clearance (<60 mL/min), was shown to increase the levels of both BNP and NT-proBNP; however, the effect was more significant with NT-proBNP.

Applicability
The diagnosis of HF in patients presenting to emergency departments is difficult. The differential diagnosis for patients presenting with the chief report of dyspnea is large, including cardiac causes, pulmonary causes, combined cardiac and pulmonary causes, and neither cardiac nor pulmonary causes. This review focused on patients with acute or chronic HF who are admitted to emergency departments or followed in primary care settings, regardless of comorbidity, which helped maximize generalizability.

For BNP, we present data on the common cutpoint of 100 pg/mL proposed by all manufacturers of FDA-approved BNP assays. This should provide users of the test with robust information on the applicability of the test to patients. For NT-proBNP, cutpoints based on age varied among studies. This lack of uniformity for NT-proBNP suggests that clinicians should cautiously apply the findings of this report to their practices in emergency departments and urgent care centers.

In primary care settings, the majority of patients do not present to general practitioners with obvious serious symptoms of HF. Identifying at-risk patients or those with subclinical HF is critical, as undiagnosed HF leads to progression and worse QOL in patients and increased costs to the health care system. BNP, using both the optimal or manufacturers’ suggested cutpoint, is effective in identifying patients at risk of HF or identifying patients with little subclinical HF. NT-proBNP is effective at identifying patients at risk of HF using the optimal cutpoint; however, limited evidence exists for using the manufacturers’ suggested cutpoint.

Research Gaps
• More studies are needed to determine the effect of age on the diagnostic cutpoints, especially for NT-proBNP. Common cutpoints that can be used in all clinical situations, especially those suggested in recent guidelines, would increase the applicability of this test.
• More studies are needed to determine the effect of declining renal function on the diagnostic performance of both BNP and NT-proBNP, and to establish cutpoints in situations of reduced renal function.
• More studies are needed to determine the effect of sex, ethnicity, and BMI on natriuretic peptide concentrations and ultimately on the cutpoints for diagnosis.
• Studies are needed to examine the role of BNP and NT-proBNP in multimarker panels for the diagnosis of HF.
• A more detailed study of the effects of heterogeneity among the studies would allow a clearer understanding of the effects of various confounders, including comorbidities.
Prognosis Studies: Patients With Acute and Chronic Heart Failure (Key Question 3)

Key Findings
The findings demonstrate that BNP and NT-proBNP are independent predictors for outcomes of mortality and morbidity. All-cause mortality and composite outcomes across different time intervals (from 14 days to 7 years in decompensated HF patients and from 12 to 44 months in chronic stable patients) were most often evaluated; cardiovascular mortality and morbidity were less frequently evaluated and showed some inconsistency in demonstrating an association with these peptides. In general, higher levels of BNP/NT-proBNP were associated with greater risk, but the thresholds used to categorize groups varied widely. In studies of decompensated HF patients, a decrease in BNP/NT-proBNP levels relative to admission levels was also predictive of decreased rates of mortality and morbidity.

The studies were rated as having moderate risk of bias overall. However, it was observed that the majority of studies had high risk of bias in two main domains: control of confounding and adequate measurement of the outcome. Many of the studies failed to assess prediction of outcomes using multivariable models that included adjustments for age, sex, BMI, and renal function, the minimum set that we established based on expert consultation and our previous review. Despite this concern, the overall conclusion that BNP and NT-proBNP are independent predictors of mortality and morbidity outcomes in persons with decompensated and stable HF remains, given the consistent association across different time periods and HF populations. It should be noted that the majority of studies employed lower hierarchical statistical approaches, reflecting early-phase prognostic study development; few studies undertook validation or impact investigations.

Applicability
With respect to applicability, most papers pertained to populations aged 60 years or older. However, we could not find specific evidence to suggest that the predictive value of BNP or NT-proBNP varies by the age, sex, or race of the study population. Although many studies controlled for sex in multivariable regression models, few investigated sex as a potential effect modifier. Thus, we cannot comment on whether the results differ in males and females. Comparing across studies that considered various cutpoints, higher cutpoints appear to be associated with greater risk. However, the studies considered a wide variety of cutpoints. Also, proportions of change (relative to baseline) varied widely in the studies, thus rendering any clear thresholds for practical clinical guidance problematic.

From a clinical perspective it is challenging to apply the test result, as there are neither established cutpoints nor tools for interpreting logBNP or logNT-proBNP to help physicians apply the information to their patients. However, the association of higher levels of BNP or NT-proBNP with poor outcomes over a variety of time periods is consistent. Current clinical guidelines do not provide information on how to use BNP and NT-proBNP in prognosis but suggest that they add prognostic information.

Research Gaps
• Future studies should consider including more women and various races. Sex and age should be investigated as effect modifiers.
• Consensus should be obtained on some key predetermined cutpoints or change relative to baseline and on clinically meaningful intervals for followup that are relevant to decompensated patients and chronic stable patients.
• Researchers should agree on and use a standard group of covariates to account for potential confounding in nonrandomized studies. In particular, future studies should include either BMI or another measure of body fat (such as waist circumference or waist-to-hip ratio) and a measure of renal function in multivariable regression models.
• Outcome assessment should also be standardized, both in terms of the types of outcomes investigated and the ways in which these outcomes are defined and measured.
• We recommend consideration of a phased approach to establishing the predictive value of BNP or NT-proBNP. Attempts to validate predictive models (internal or external) are an important priority for future research.
• There is a need for more impact studies assessing the clinical utility of using the predictive models.
• For populations with acute HF, more studies are needed to evaluate the potential differences in predictive ability between admission and discharge levels of BNP and NT-proBNP.
Prognosis Studies: Adding Predictive Information to Other Prognostic Methods in Patients With Heart Failure (Key Question 4)

Key Findings
For patients with decompensated HF, only mortality outcomes were evaluated with respect to incremental prognostic value; in chronic stable HF patients, mortality, morbidity, and composite outcomes were assessed. Overall, despite the differences in base predictive models, cutpoints, and lengths of followup, BNP and NT-proBNP were both shown to add incremental predictive value in acutely ill HF patients for all-cause mortality; however, the highest incremental predictive value was achieved when BNP or NT-proBNP was combined with other markers such as CA125 or MR-proADM. Fewer studies evaluated cardiovascular mortality, but they also demonstrated the independent predictive value of BNP.

When considering composite outcomes, NT-proBNP was shown to be an independent predictor; there are too few studies evaluating morbidity to assess incremental prognostic value. Only one study attempted internal validation and none employed external validation. Five publications undertook reclassification statistics, and results show inconsistency regarding the incremental prognostic value of NT-proBNP.

Applicability
Studies addressing KQ4 consisted predominately of middle-aged and elderly male subjects with HF. Time intervals were heterogeneous for studies of both decompensated HF (from 31 days to 6.8 years) and chronic stable HF (from 12 to 37 months), making comparisons across studies problematic. There were also differences in statistical base models, cutpoints, and lengths of followup, thereby suggesting that the studies are applicable to these specific factors.

Research Gaps
- There is a need to move to higher level hierarchical approaches (internal and external validation) when selecting statistical evaluations (i.e., reclassification methods), as well as designing impact studies.
- There is a need to evaluate outcomes of morbidity and composite outcomes in decompensated HF subjects with respect to the incremental value of BNP and NT-proBNP.
- There is a need to evaluate BNP in stable chronic populations with respect to incremental predictive value.

- Future research recommendations for KQ3 (see above) are also applicable for KQ4.

Prognosis Studies: General Populations (Key Question 5)

Key Findings
The adjusted HR demonstrates the log-linear relationship between baseline NT-proBNP and cardiovascular death as well as all-cause mortality, taking into consideration age, sex, BMI, and renal function. Our findings demonstrate clearly that there is an association between NT-proBNP and the outcomes of morbidity (HF and AF), as well as mortality (all cause, cardiovascular, and sudden cardiac).

For outcomes that are associated with cardiac disease (incident HF and AF), there appears to be a log-linear relationship between NT-proBNP and the outcome, taking into consideration age, sex, BMI, and renal function. In addition, NT-proBNP seems to perform well, even when adjusted for other conventional risk markers and biomarkers.

Applicability
While the association is clear, the directness or applicability of these findings to patient care is not demonstrated well in the included papers. Two papers considered the application of NT-proBNP to other traditional risk factors and used the c-statistic to assess the additional discrimination for risk prediction. To translate this into clinical practice will require the development of specific risk calculators that take into consideration confounders and any other established risk markers.

Research Gaps
Future research should develop specific risk calculators that take into consideration confounders and any other established risk markers. Such models will require testing in population cohorts before the use of NT-proBNP or BNP can be validated for use as a prognostic marker in community settings.

BNP-Assisted Therapy (Key Question 6)

Key Findings
Few RCTs have been undertaken to assess whether BNP-guided therapy has benefits over usual care. Studies varied in patient selection; baseline characteristics of patients; therapy (type, schedule, goals); BNP/NT-proBNP target; outcome types; and how the findings were reported. The conclusions from these studies are varied, in part because of the differences in study design and outcomes.
Meta-analyses were not performed because of the substantial heterogeneity among the studies, and therefore no quantitative summary estimates could be made. Differences among studies provide greater understanding of how BNP/NT-proBNP therapy can be used, despite whether trials succeeded or failed.

Four of five studies reported at least one outcome that was better in the group with therapy guided by BNP/NT-proBNP than in the usual-care group.168,170,173,174 Five studies reported negative results, three167,171,172 of which had short followups (3–9 months) that would have limited the number of long-term outcomes.

One limitation to this systematic review was the exclusion of two trials, the 2000 trial assessing therapy guided by NT-proBNP186 and a more recent study in 2010 done by the same research group.187 They were not included because the NT-proBNP assay used is not commercially available. These data would have strengthened the results of this systematic review but not altered the conclusions.

**Applicability**

Understanding the usefulness of BNP or NT-proBNP measurement in the assessment of HF status will allow better management of HF patients, essentially serving as a barometer. Currently, the data from the studies that have evaluated BNP or NT-proBNP for this purpose are inconclusive.

**Research Gaps**

Future trials should consider the following design features:

- Therapy optimized at baseline according to clinical guidelines.
- BNP or NT-proBNP target near the median value for patients with stable HF.
- Consideration of use of the relative change value when gauging the value of a change in therapy.
- Followup of 2 years or more.
- Inclusion of all relevant endpoints: cardiovascular mortality, total mortality, days alive and not hospitalized for HF, number of HF hospitalizations, number of HF events not requiring hospitalization, surrogate measures of renal function (e.g., creatinine) and ischemia (e.g., troponin), number of patients who have achieved target BNP/NT-proBNP concentration, and number of patients who have achieved recommended medication doses. Also, inclusion as part of medication information of the number of patients who are taking additional medications or doses above the recommended amounts. Inclusion of QOL questionnaires for additional value.
- Sample size calculations to demonstrate adequate study power for the outcomes selected.

**Biological Variation (Key Question 7)**

**Key Findings**

This systematic review of biological variation was specific to patients with stable HF or healthy controls. In the two studies in which healthy individuals were evaluated, the RCVs were higher than those in studies of patients with stable HF. Within-individual variation was similar for BNP (median = 25%) and NT-proBNP (median = 20%), but lower in short measurement intervals (hours, days) than longer measurement intervals (weeks, year). Although the circulating half-life of BNP is much shorter (21 minutes) than that for NT-proBNP (60–120 minutes), this did not seem to affect the within-individual variation (CVi) values much.188 No meta-analysis could be done to compute summary estimates for CVi or RCV, as confidence limits were not provided for variance data in any study.

Most studies included in this systematic review considered at least some known preanalytical factors and tried to minimize or address them. However, the determinants of within-person biological variation have not been well explored; more is known about between-person variation, such as sex, age, exercise, and comorbidity.189 The biological variations are likely due to subclinical changes in hemodynamics, hormonal regulation, and clearance, and perhaps even differences in the type of circulating forms of BNP.188

The IOI for BNP and NT-proBNP was between 0.03 and 0.14, which is lower than any of the common biochemistry analytes.190 A low IOI (<0.48) is considered to reflect strong individuality, which in turn indicates that an individual patient should be assessed with respect to his or her individual hormonal level.

**Applicability**

The applicability of the RCV values calculated from stable HF patients is to assess instability in HF patients. Although the inclusion criteria of patients with stable HF varied among studies, this did not seem to influence the RCV values by a large degree. The timeframe of collection for the biological variation data seemed to influence the RCV. The within-hour and within-day values were much lower, yet there was no discernible difference beyond this time period (up to 2 years). Interestingly, the RCV values for BNP were about double those for NT-proBNP, suggesting that NT-proBNP would be more sensitive than BNP for detecting a significant change. The implication is that NT-proBNP may be better than BNP for serial monitoring.
Research Gaps

Additional studies are needed to provide supporting evidence of the biological variation parameters. These studies should be designed to capture sources of biological variation determinants by multivariable regression analysis and would therefore require larger sample sizes than have been used thus far. Preanalytical and analytical variation should be minimized by collection of samples in the early morning, increasing the frequency of collection, and duplicating determinations to increase the accuracy of the measure. Calculations should include CIs to show reliability and allow meta-analyses to be done.

References


26. Maisel AS, McCoard J, Nowak RM, et al. Bedside B-Type natriuretic peptide in the emergency diagnosis of heart failure with reduced or preserved ejection fraction. Results from the


57. Rogers RK, Stoddard GJ, Greene T, et al. Usefulness of adjusting for clinical covariates to improve the ability of B-type natriuretic peptide to distinguish cardiac from non-cardiac dyspnea. Am J Cardiol. 2009;104(7):966-71. PMID: 19766765.


Full Report