Academic Emergency Medicine
Official Journal of the Society for Academic Emergency Medicine

EVIDENCE-BASED DIAGNOSTICS

Diagnosing Acute Heart Failure in the Emergency Department: A Systematic Review and Meta-analysis

Jennifer L. Martindale, MD, Abel Wakai, MD, Sean P. Collins, MD, MSc, Phillip D. Levy, MD, MPH, Deborah Diercks, MD, Brian C. Hiestand, MD, Gregory J. Fermann, MD, Ian deSouza, MD, and Richard Sinert, DO

Abstract

Background: Acute heart failure (AHF) is one of the most common diagnoses assigned to emergency department (ED) patients who are hospitalized. Despite its high prevalence in the emergency setting, the diagnosis of AHF in ED patients with undifferentiated dyspnea can be challenging.

Objectives: The primary objective of this study was to perform a systematic review and meta-analysis of the operating characteristics of diagnostic elements available to the emergency physician for diagnosing AHF. Secondary objectives were to develop a test–treatment threshold model and to calculate interval likelihood ratios (LRs) for natriuretic peptides (NPs) by pooling patient-level results.

Methods: PubMed, EMBASE, and selected bibliographies were searched from January 1965 to March 2015 using MeSH terms to address the ability of the following index tests to predict AHF as a cause of dyspnea in adult patients in the ED: history and physical examination, electrocardiogram, chest radiograph (CXR), B-type natriuretic peptide (BNP), N-terminal proB-type natriuretic peptide (NT-proBNP), lung ultrasound (US), bedside echocardiography, and bioimpedance. A diagnosis of AHF based on clinical data combined with objective test results served as the criterion standard diagnosis. Data were analyzed using Meta-DiSc software. Authors of all NP studies were contacted to obtain patient-level data. The Quality Assessment Tool for Diagnostic Accuracy Studies-2 (QUADAS-2) for systematic reviews was utilized to evaluate the quality and applicability of the studies included.

Results: Based on the included studies, the prevalence of AHF ranged from 29% to 79%. Index tests with pooled positive LRs ≥ 4 were the auscultation of S3 on physical examination (4.0, 95% confidence interval [CI] = 2.7 to 5.9), pulmonary edema on both CXR (4.8, 95% CI = 3.6 to 6.4) and lung US (7.4, 95% CI = 4.2 to 12.8), and reduced ejection fraction observed on bedside echocardiogram (4.1, 95% CI = 2.4 to 7.2). Tests with low negative LRs were BNP < 100 pg/mL (0.11, 95% CI = 0.07 to 0.16), NT-proBNP < 300 pg/mL (0.09, 95% CI = 0.03 to 0.34), and B-line pattern on lung US LR (0.16, 95% CI = 0.05 to 0.51). Interval LRs of BNP concentrations at the low end of "positive" results as defined by a cutoff of 100 pg/mL were substantially lower (100 to 200 pg/mL; 0.29, 95% CI = 0.23 to 0.38) than those associated with higher BNP concentrations (1000 to 1500 pg/mL; 7.12, 95% CI = 4.53 to 11.18). The interval LR of NT-proBNP concentrations even at very high values (30,000 to 200,000 pg/mL) was 3.30 (95% CI = 2.05 to 5.31).

Conclusions: Bedside lung US and echocardiography appear to the most useful tests for affirming the presence of AHF while NPs are valuable in excluding the diagnosis.

ACADEMIC EMERGENCY MEDICINE 2016;23:223–242

From the Department of Emergency Medicine, SUNY Downstate Medical Center (JLM, IdS, RS), New York, NY; the Emergency Care Research Unit, Royal College of Surgeons in Ireland (AW, RCSI), Dublin, Ireland; the Department of Emergency Medicine, Vanderbilt University (SPC), Nashville, TN; the Department of Emergency Medicine, Wayne State University School of Medicine (PL), Detroit, MI; the Department of Emergency Medicine, University of Texas Southwestern (DD), Dallas, TX; the Department of Emergency Medicine, Wake Forest University School of Medicine (BCH), Winston-Salem, NC; and the Department of Emergency Medicine, University of Cincinnati (GJF), Cincinnati, OH.

Received June 1, 2015; revision received August 31, 2015; accepted September 16, 2015.
The authors have no relevant financial information or potential conflicts to disclose.
Supervising Editor: Christopher Carpenter, MD.

Address for correspondence and reprints: Jennifer L Martindale, MD; e-mail: martindale.jen@gmail.com.

A related article appears on page 347.

ISSN 1069-6563
PII ISSN 1069-6563583

doi: 10.1111/acem.12878

Acute heart failure (AHF) is defined as a gradual or rapid deterioration in heart failure signs and symptoms in need of urgent treatment.[1] Dyspnea is the most common presenting complaint prompting AHF patients to seek acute care.[2,3] A primary diagnosis of AHF accounts for approximately one million emergency department (ED) visits in the United States.[4] Despite its high prevalence in the ED setting, diagnosing AHF in ED patients with undifferentiated dyspnea can be challenging, especially in patients with advanced age[5] and comorbid disease.[6] There is no single historical, physical examination, electrocardiographic (ECG) or radiographic finding that can on its own reliably diagnose or rule out AHF as the cause of dyspnea. An ED diagnosis of AHF based on history, physical examination, chest radiograph (CXR), and ECG is qualified as "uncertain" in 44% of cases[7] and is discordant with the final discharge diagnosis in nearly one out of every four cases.[8–11] Natriuretic peptide (NP) testing improves diagnostic uncertainty for acutely dyspneic patients[7,12] and is a now a routine component of the workup of patients with possible AHF.[13] However, even when NP testing is incorporated into the clinical workup of acute dyspnea, the misclassification rate remains 14% to 29%.[14–17] Other diagnostic modalities such as bedside echocardiography,[18–21] lung ultrasound (US),[22,23] and bioimpedance[24,25] have been shown to help discriminate between AHF and other primary causes of dyspnea, but their added clinical utility has yet to be fully characterized.

Reviews and meta-analyses aimed at helping clinicians sort through the array of available resources for evaluation of dyspneic patients have been previously published, but the most comprehensive of these is now more than a decade old.[26] As new diagnostic modalities have been integrated into clinical practice and new data have been published in the ensuing period, the primary objective of this systematic review was to cohere the current best evidence concerning the diagnostic accuracy of index tests that might help discriminate AHF from other clinical conditions in patients presenting to the ED with dyspnea. The index tests evaluated in this review include: clinical history, symptoms, physical examination findings, ECG, CXR, B-type natriuretic peptide (BNP), N-terminal proB-type natriuretic peptide (NT-proBNP), lung US, ED-based bedside echocardiography, and bioimpedance.

METHODS

Study Design

We conducted a systematic review of studies that examined the operating test characteristics of the modalities used by emergency physicians (EPs) for the diagnosis of AHF among patients presenting to the ED with dyspnea. The systematic review was conducted using the Preferred Reporting Items for Systematic Review and Meta-analyses (PRISMA) guidelines.[27]

Search Strategy

The design and manuscript structure of this systematic review conform to the recommendations from the Meta-analysis of Observational Studies in Epidemiology (MOOSE) statement.[28] The medical literature was searched using PubMed and EMBASE from their inception through March 2015. With the assistance of an experienced health sciences librarian, the selected Medical Subject Headings (MeSH) terms *heart failure and dyspnea* were individually combined with the MeSH terms *sensitivity and specificity, predictive value of tests, history taking, physical examination, electrocardiography, natriuretic peptide, ultrasonography, ultrasonics, echocardiography,* and *bioelectrical impedance*. References from review articles identified by these searches were searched for relevant studies. Two authors independently screened the search results of each diagnostic modality of interest.

Studies eligible for inclusion in the systematic review were those that focused on the diagnosis of AHF in the ED population. Studies that focused on diagnostic tests that were not available shortly after ED presentation were excluded, as were studies that focused on patients with compensated, chronic heart failure and those that focused on prognosis or therapeutics. Case studies and reports or studies published solely in abstract form were excluded. The search was also restricted to human studies published in English.

Shortly after this systematic review project began, a comprehensive systematic review by Hill et al.[29] on NPs for the diagnosis of AHF in the ED was published. All references identified by Hill et al.[29] were screened for possible inclusion in this review. To avoid duplicating the PubMed and EMBASE NP searches performed by Hill et al,[29] we limited our search for NP articles using these databases to a time frame beyond the stop point of their review (June 2012).

A review author (JLM) was tasked with ensuring that reported diagnostic data outside of the scope of each individual search was included in the analysis of the other relevant diagnostic elements in this review (e.g., a study identified only by the bioimpedance search strategy that reported BNP data was referred to reviewers also assigned to BNP/NT-proBNP).

Criteria for Considering Studies for this Review

Types of Participants. We included studies that recruited adult patients presenting to the ED with dyspnea as a primary complaint. The authors chose to exclude studies that recruited patients who presented to an urgent care setting, as we felt these patients might undergo an abbreviated diagnostic workup and likely represent a different spectrum of AHF than those who present to the ED with true, emergent dyspnea. Patients were not excluded based on comorbidities, etiology of AHF, or the presence of arrhythmia.

Types of Index Tests. We included studies that used history, symptoms, and physical examination findings as index tests for the diagnosis of AHF. Our search for NP testing was limited to BNP and NT-proBNP since these are the two most commonly used peptides for the ED diagnosis of AHF. We did not limit our search to point-of-care testing, as results from standard NP tests conducted in a hospital laboratory are typically made available to EPs during a patient's ED course. We also did not limit our search and inclusion to any specific

type of assay. Our search for lung US studies was not limited to a specific protocol. We narrowed our inclusion of studies investigating lung US and echocardiography to those that had EPs both performing and interpreting these tests. Inclusion of bioimpedance studies was not limited to any particular type of protocol, bioimpedance metric, or specific device.

Types of Reference Standard. Acute heart failure is a clinical diagnosis, and there are currently no universally accepted diagnostic criteria to serve as the criterion standard for AHF in the acute care setting. Previously published diagnostic criteria (Framingham,[30] Boston,[31] National Health and Nutrition Examination Survey [NHANES][32]) may be used as clinical guides but on their own lack sensitivity for the diagnosis of AHF.[33,34] The most prevalent accepted criterion standard for diagnostic research in AHF is an adjudicated diagnosis reached by physicians after retrospective review of inpatient medical records.[7,26,35,36] We therefore chose to include studies that used as a reference standard a final diagnosis of AHF based on adjudication of clinical data by independent reviewers who were blinded to the study's primary index test. We imposed no restrictions on the timing of the final diagnosis, the type of physician making the final diagnosis, or the type of clinical data upon which the final diagnosis was based.

Data Abstraction. Two or more authors for each index test independently selected articles from the combined PubMed/EMBASE search for full text review (history and physical, JM and AW; CXR, BH and GF; ECG, JM and AW; NT-proBNP and BNP, SC, DD, and PL; echocardiography, RS and IdS; lung US, RS and IdS; bioimpedance, BH and GF). Each reviewer independently selected potentially eligible studies before both authors agreed on the list of studies for full text review. Differences in study selection were resolved by consensus. Having read the methods sections of the full-text version of the studies potentially eligible for inclusion, each author then applied the stated inclusion and exclusion criteria to determine which studies to include in our systematic review. Differences were resolved by consensus after discussion and adjudication.

A standardized data collection form (see Data Supplement S1, available as supporting information in the online version of this paper) was used to abstract data pertaining to study funding, study location and setting, patient selection, patient demographics, prevalence of heart failure, manufacturer of the index test, definition of a positive diagnostic test, specialist performing and interpreting the index test, blinding with respect to the index test and criterion standard, and data incorporated into the criterion standard diagnosis of AHF. Study authors were contacted when study methodology or results required clarification.

Data Analysis

Sensitivities, specificities, and likelihood ratios (LRs) were calculated based on constructed two-by-two tables for each included study. To compute meta-analysis summary estimates when more than one study assessed the same index test, we combined test characteristic data using a random-effects model with MetaDiSc[37] software. Interstudy heterogeneity was assessed for pooled estimates of sensitivity and specificity using the DerSimonian-Laird random-effects model.[38] Publication bias, or overrepresentation of studies yielding positive results in published literature, was not assessed because a consensus approach toward funnel plot analysis is lacking for meta-analyses of diagnostic studies.[39]

BNP and NT-proBNP Analysis

Summary analyses for dichotomous NP results were performed separately for each type of peptide, assay manufacturer, and common cutoff value. Cutoff values within 5 pg/mL for BNP and 50 pg/mL for NT-proBNP were considered common. Because they are ubiquitous in clinical practice, and have various decision-making cut-points, we also used patient-level data to compute interval LRs for BNP/NT-proBNP. The interval LR represents the probability of a test result within a user-defined interval in patients with a criterion standard diagnosis of AHF divided by the probability of a result in the same interval in patients with an alternative (non-AHF) diagnosis.[40] To derive patient-level data, we contacted the authors of included studies that examined the test characteristics of BNP or NT-proBNP and requested actual patient-level NP results and their associated final diagnoses (AHF/not AHF).

Quality Assessment

Two authors independently used the revised Quality Assessment Tool for Diagnostic Accuracy Studies-2 (QUADAS-2) to evaluate the overall quality of evidence for included studies relating to each index test.[41] The QUADAS-2 tool assists review authors in making explicit judgments for the risk of bias in four domains of study methodology: 1) patient selection, 2) index test, 3) reference standard, and 4) flow and timing. The tool also allows authors to rate how well a study's 1) patient selection, 2) index test, and 3) reference standard apply to the specific research question posed by the systematic review. There are currently no criteria for assessing risk of bias in studies that compare multiple index tests.[41] Studies that evaluated more than one index test were excluded from our QUADAS-2 analysis. The following signaling questions and statements were used to tailor the QUADAS-2 tool to this systematic review and guide quality judgments:

Patient Selection. The risk of spectrum bias was considered high if 1) the study was a case-control design or 2) the study made inappropriate exclusions that eliminated diagnoses with overlapping features of AHF or reduced the burden of disease in the non-AHF population. Examples of unacceptable exclusions included renal insufficiency, cirrhosis, morbid obesity, and other causes of dyspnea such as pneumonia, pulmonary embolism, and chronic obstructive pulmonary disease (COPD). Excluding these conditions would inflate the specificity of diagnostic tests.[42] Acceptable exclusions included trauma patients, patients with an acute coronary syndrome unless dyspnea was the predominant complaint, and obvious noncardiac diagnoses such as pneumothorax. Loss of applicability was rated as "high" if the study

failed to consider all patients presenting to the ED with a chief complaint of dyspnea for which AHF was in the differential. Studies that excluded patients based on age demographic (other than > 18 years) or dyspnea severity or had significant gender imbalance were also at high risk for losing applicability.

Index Test. If the result of the index test was interpreted with the foreknowledge of the initial diagnosis or other clinical data, the study was at high risk of overestimating the accuracy of the diagnostic test. This applied to all index tests except history and examination, bioimpedance, and NP. If the study failed to use a predetermined cut-off value or definition of a positive test, the study was also considered to be at high risk for bias in this domain. Loss of applicability was rated as "high" if the diagnostic test was performed or interpreted by specialists other than EPs. Sonographic studies performed by fellowship-trained EPs were considered to be at high risk for losing applicability to the majority of EPs without formal US training.

Reference Standard. If the results from the index test were incorporated into the final diagnosis of AHF, this domain was considered at high risk for incorporation bias and inflated estimates of both sensitivity and specificity.[42,43] Exception for history and examination elements was made, as these were considered essential for the reference standard. If the approach to the criterion standard diagnosis of heart failure was not explicit or if it failed to incorporate previously established clinical criteria (for example, Framingham,[30] Boston,[31] European Society of Cardiology,[44] NHANES[32]), the risk of bias was considered high. If the study's criterion standard did not match our review's clinical definition (new or worsening dyspnea in addition to objective evidence supporting the diagnosis of AHF), this domain was considered less applicable to our review. For example, a study that diagnosed AHF based on echocardiographic evidence alone would be less applicable.

Flow and Timing. If there was a significant delay between ED presentation and performance of the index test, this domain was considered to be at high risk for bias. We used 6 hours as the cutoff after which a patient would likely have received therapeutic interventions and change clinical course enough to affect diagnostic testing. Sonographic and bioimpedance studies that excluded patients because of lack of feasibility (for example, body habitus, poor acoustic windows) were also considered to be at high risk for bias.

Each pair of authors independently piloted the QUADAS-2 tool with a sample of one to three included studies. Signaling questions that led to incongruent piloted responses were refined as deemed necessary. Kappa analysis using SPSS Statistics version 17.0 was used to calculate statistical agreement regarding blinded QUADAS-2 answers when more than three studies per index test could be evaluated by these criteria. When consensus agreement could not be reached between the two authors, a third reviewer (JLM) adjudicated.

Risk of bias and applicability for the 35 NP studies derived from the systematic review by Hill et al.[29] was already examined using the QUADAS-2 tool. Evaluation of bias and applicability was therefore limited to the more recent studies identified by our search and included in this review.

Test–Treatment Threshold

The Pauker and Kassirer decision threshold model,[45] which incorporates diagnostic test characteristics as well as estimated risks and benefits of treatment of AHF, was used to determine testing and treatment thresholds. Briefly, the Pauker and Kassirer decision threshold model is based on five variables: sensitivity, specificity, risk of a diagnostic test, risk of treatment, and anticipated benefit of treatment.

RESULTS

From the 9,405 citations identified by the PubMed searches, 9,317 citations identified by the EMBASE searches, and two studies identified from the bibliographies of review articles, a total of 57 studies including 52 unique patient cohorts were chosen for inclusion in this systematic review. A summary of the selection process for the systematic review is presented in Figure 1. Study selection diagrams illustrating the separate searches applied to each index test are shown in Data Supplement S2 (available as supporting information in the online version of this paper). Most studies were prospective and cross-sectional in design. Only two[15,46] of the included studies published after 2003 were conducted using the Standards for Reporting of Diagnostic Accuracy (STARD) criteria.[47]

Prevalence

The combined population from the 52 unique cohorts included in this review was 17,893 patients. The summary prevalence of AHF in these studies was 45.6% (95% CI = 44.9 to 46.4) with a range from 29% to 79%.

History and Physical Examination

A total of 31 studies[7,8,12,14–16,18–21,23,25,46,48–65] reporting test characteristics for history and physical examination were selected for inclusion and are fully described in Data Supplement S3 (available as supporting information in the online version of this paper). Studies that duplicated the cohort of an included study,[36,66] reported insufficient information to calculate test characteristics[67,68] or took place in the prehospital setting rather than the ED[69] were excluded after full text review (Data Supplement S2). Two[56,59] of the included studies focused on history and physical examination as the primary index tests of interest.

Point estimates for variables relating to clinical history, symptoms, and physical examination findings that were reported in at least four separate patient cohorts are included in Table 1. Additional point estimates are included in Data Supplement S3. Statistical heterogeneity was very high for all pooled estimates of variables reported in more than two studies except for diabetes (I^2 = 21%). There was no single historical variable, symptom, or physical examination finding that could significantly reduce the likelihood of AHF. A history of prosthetic valve,[55] aortic or mitral valve disease,[7] and

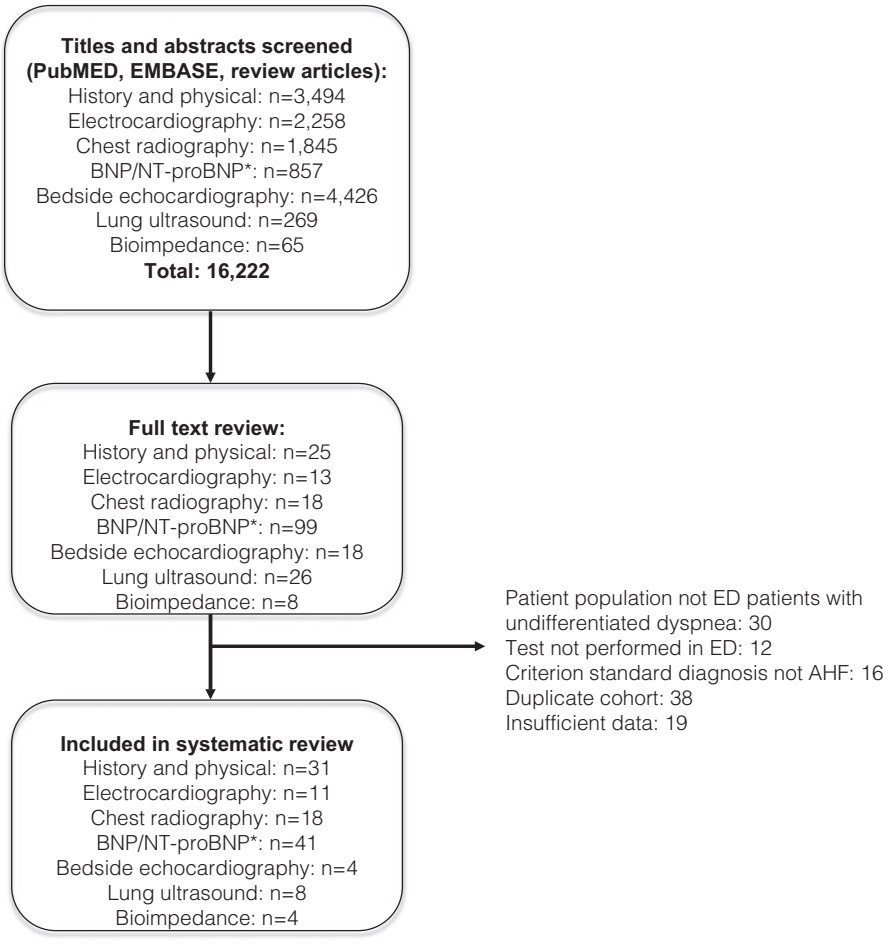

Figure 1. Summary of the selection process for the systematic review. AHF = acute heart failure; BNP = B-type natriuretic peptide; NT-proBNP = N-terminal proB-type natriuretic peptide.

hypertensive crisis[49] were found to increase the likelihood of AHF, but these point estimates were derived from a single study rather than pooled studies (Data Supplement S3). The physical examination finding with the highest positive LR was an S3 gallop. For the 14 studies that reported S3 data,[7,8,14,15,20,53,55,57,59,60,64,65] (forest plot, Data Supplement S2) positive LRs ranged from 1.6[58] to 13.0.[59]

Based on the QUADAS-2 tool and due to the patient exclusion criteria used, the risk of bias in patient selection was high for the majority of the 17 studies subjected to this analysis[7,8,12,15,16,18,35,50,51,53,54,56–60,70] (Data Supplement S2). Exclusion of patients with significant renal dysfunction, acute coronary syndrome, and other comorbidities in these studies likely resulted in a spectrum of patients that were less severely ill, leading to an overestimation of specificity for the diagnostic tests evaluated in these studies.[42] Risk of bias for the criterion standard was rated low in 65% of the included studies. The reliability for the authors' QUADAS-2 assessments was moderate[71] (κ = 0.59, 95% CI = 0.43 to 0.75).

ECG

Searches from PubMed, EMBASE, and references of review articles identified 2,258 citations, 13 of which were selected for full text review. Of these, three[15,16,65] were selected for inclusion in the systematic review. Searches dedicated to other index tests in this review identified an additional eight studies.[7,12,19,20,51,58,60,62] ECG variables were grouped with baseline patient characteristic data in each of the included studies; electrocardiogram was not the index test of primary interest in any of these studies. All but two studies[15,16] excluded patients with acute coronary ischemia or myocardial infarction. None of the abnormal electrographic findings substantially increased or decreased the likelihood of AHF (Table 2). The QUADAS-2 criteria could be applied to seven of the included studies (Data Supplement S2). Agreement between the two investigators was moderate (κ = 0.55, 95% CI = 0.29 to 0.81).

Chest Radiography

From the 1,845 citations selected from PubMed, EMBASE, and bibliographic review of key review articles, 18 were selected for full text review. However, only one study[66] from full-text review was ultimately included; the other 17 studies[7,8,12,14–16,18–21,23,54,57,58,60,64,72] included in the systematic review were identified from searches designed for the other index tests (Data Supplement S2). Chest radiography was not the primary index test of interest in any of the included

Table 1
Pooled Test Performance Characteristics for History and Physical Examination Findings

	No. of Studies	No. of Patients	% AHF (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	LR+ (95% CI)	LR– (95% CI)
Symptoms							
Orthopnea[7,8,14,16,19,21,36,48,49,53–55,58,60,65]	15	5,430	45.5 (44.2–46.9)	52.1 (50.1–54.0)	70.5 (68.8–72.1)	1.9 (1.4–2.5)	0.74 (0.64–0.85)
PND[7,8,14,35,48,51,53,59,64]	9	2,216	44.8 (42.8–46.9)	46.2 (43.7–48.6)	73.9 (71.9–75.9)	1.6 (1.2–2.1)	0.79 (0.71–0.88)
Dyspnea at rest[20,51,55,61]	4	2,038	37.9 (35.9–40.0)	54.6 (51.2–58.0)	49.6 (46.9–52.3)	1.1 (0.9–1.4)	0.88 (0.74–1.04)
Absence of productive cough[7,8,12,36,49,59,62]	7	2,414	43.0 (41.0–45.0)	82.0 (79.6–84.4)	25.8 (23.5–28.2)	1.13 (1.02–1.26)	0.6 (0.5–0.8)
History							
CRI[25,49,55,57,59,61]	6	3,009	42.8 (41.0–44.6)	32.0 (29.4–34.6)	91.4 (90.0–92.7)	3.4 (2.7–4.5)	0.75 (0.71–0.80)
Arrhythmia[7,12,18,55,62]	5	3,469	40.2 (38.6–41.9)	38.0 (36.1–40.0)	85.1 (83.9–86.2)	2.7 (2.2–3.4)	0.75 (0.68–0.83)
CHF[7,8,14,15,19–21,23,25,35,36,48–50,55,57–61,63,65]	22	8,493	46.0 (44.9–47.0)	55.5 (53.9–57.1)	80.2 (79.0–81.3)	2.7 (2.0–3.7)	0.58 (0.49–0.68)
Renal failure[15,18,36,48,50]	5	2,840	40.9 (39.1–42.7)	15.1 (13.1–17.3)	95.1 (94–96.1)	2.3 (1.3–3.9)	0.9 (0.73–1.11)
MI, history of[7,15,19,48,49,52,54,55,65]	9	4,208	40.5 (39.1–42.0)	31.8 (29.7–33.9)	87.1 (85.8–88.3)	2.1 (1.8–2.5)	0.82 (0.76–0.89)
AFIB[36,49,52–54,65]	6	1,935	51.9 (49.8–54.2)	30.2 (27.4–33.2)	85.3 (82.8–87.5)	2.1 (1.6–2.9)	0.82 (0.71–0.93)
CAD[7,14,18,20,21,25,49,55,57–61,63]	14	4,983	42.9 (41.5–44.3)	46.6 (44.5–48.7)	76.2 (74.6–77.7)	2.0 (1.7–2.4)	0.71 (0.64–0.79)
Hyperlipidemia[8,49,53,55,68]	5	2,923	39.8 (38.1–41.6)	33.8 (31.1–36.6)	75.3 (73.2–77.3)	1.6 (1.3–1.9)	0.85 (0.82–0.90)
DM[8,16,18,19,21,23,25,49,50,52–55,57,59–61,64,65]	19	7,707	47.3 (46.2–48.4)	28.8 (27.4–30.4)	81.7 (80.4–82.8)	1.5 (1.3–1.7)	0.89 (0.84–0.94)
HTN[7,8,12,14–16,18,19,21,23,25,36,48–50,53–55,57,58,60,61,63–65]	25	10,137	45.6 (44.6–46.6)	66.9 (65.5–68.3)	50.7 (49.4–52.1)	1.3 (1.3–1.4)	0.62 (0.53–0.73)
No history of COPD[7,8,15,18,20,21,23,25,36,48,50,53,55,57–59,61,63]	18	8,053	42.8 (41.7–43.9)	78.9 (77.4–80.3)	34.1 (32.6–35.6)	1.22 (1.11–1.36)	0.7 (0.6–0.8)
Examination findings							
S3[7,8,14,15,20,53–55,57–60,64,65]	14	5,900	45.2 (44.0–46.5)	12.7 (11.5–14.0)	97.7 (97.2–98.2)	4.0 (2.7–5.9)	0.91 (0.88–0.95)
JVD[7,8,12,14–16,18,19,21,25,36,48,51,53–55,57–61,64,65]	23	8,012	47.8 (46.7–48.9)	37.2 (35.7–38.7)	87.0 (85.9–88.0)	2.8 (1.7–4.5)	0.76 (0.69–0.84)
Hepatojugular reflex[56,59,61,65]	4	1,209	60.4 (57.6–63.1)	14.1 (11.9–16.6)	93.4 (91.2–95.2)	2.2 (1.3–3.7)	0.91 (0.88–0.94)
Leg edema[7,8,10,12,14,15,16,18,19–21,23,25,48,49,51,53–55,57–62,65]	26	9,626	47.2 (46.2–48.2)	51.9 (50.5- 53.4)	75.2 (74.0–76.4)	1.9 (1.6–2.3)	0.68 (0.61–0.75)
Murmur[7,12,51,54,55,58,62,65]	8	4,004	45.3 (43.8–46.8)	27.8 (25.8–29.9)	83.2 (81.6–84.8)	1.9 (0.9–3.9)	0.93 (0.79–1.08)
Rales[7,8,10,12,15,18–21,23,25,36,48,51,53–55,58–61,65]	22	8,775	48.2 (47.1–49.2)	62.3 (60.8–63.7)	68.1 (66.7–69.4)	1.8 (1.5–2.1)	0.60 (0.51–0.69)
Wheezing[7,8,12,15,20,23,36,48,53,55,58,59,65]	13	6,970	44.2 (43.0–45.3)	22.3 (20.9–23.8)	64.0 (62.5–65.4)	0.6 (0.5–0.8)	1.19 (1.10–1.30)
Absent fever[7,23,36,49,59,62,63]	7	3,197	43.6 (41.9–45.3)	92.4 (90.9–93.8)	20.6 (18.8–22.5)	1.14 (1.02–1.27)	0.4 (0.3–0.6)

AFIB = atrial fibrillation; CAD = coronary artery disease; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; CRI = chronic renal insufficiency; DM = diabetes mellitus; HTN = hypertension; JVD = jugular venous distension; LR+ = positive likelihood ratio; LR– = negative likelihood ratio; MI = myocardial infarction; PND = paroxysmal nocturnal dyspnea.

Table 2
Pooled Test Performance Characteristics for Chest Radiograph and Electrocardiogram Findings

	No. of Studies	No. of Patients	% AHF (95% CI)	Sensitivity, % (95%CI)	Specificity, % (95%CI)	LR+ (95% CI)	LR− (95% CI)
Electrocardiogram							
Ischemic changes[15,51]	2	1,138	42.6 (39.8–45.5)	34.0 (29.8–38.4)	84.2 (81.2–86.9)	2.9 (1.2–7.1)	0.78 (0.73–0.84)
T-wave inversion[65]	1	709	69.4 (65.9–72.7)	10.0 (7.5–13.0)	95.85 (92.3–98.1)	2.4 (1.2–4.8)	0.94 (0.90–0.98)
Atrial fibrillation[19,20,36,58,60,65]	6	2,242	55.8 (53.7–57.8)	20.5 (18.3–22.9)	89.9 (87.9–91.7)	2.2 (1.4–3.5)	0.88 (0.85–0.91)
ST-depression[58,65]	2	1,024	60.8 (57.8–63.8)	5.6 (3.9–7.7)	96.5 (94.2–98.1)	2.0 (1.0–3.8)	0.97 (0.95–1.00)
Normal sinus rhythm[8,12,62]	3	1,207	39.6 (36.9–42.4)	55.4 (50.9–60.0)	17.8 (15.1–20.8)	0.7 (0.5–0.9)	2.88 (1.26–6.57)
ST-elevation[58]	1	219	61.2 (54.6–67.4)	5.2 (2.1–10.5)	91.8 (83.8–96.6)	0.6 (0.2–1.7)	1.03 (0.96–1.11)
Chest radiograph							
Kerley B-lines[36,72]	2	814	46.8 (43.4–50.2)	9.2 (6.5–12.5)	98.8 (97.3–99.6)	6.5 (2.6–16.2)	0.88 (0.69–1.13)
Interstitial edema[15,66,72]	3	2,001	48.3 (46.2–50.5)	31.1 (28.2–34.2)	95.1 (93.6–96.3)	6.4 (3.4–12.2)	0.73 (0.68–0.78)
Cephalization[8,57,64,66,72]	5	1,338	54.0 (51.3–56.6)	44.7 (41.1–48.4)	94.6 (92.6–96.3)	5.6 (2.9–10.4)	0.53 (0.39–0.72)
Alveolar edema[15,66,72]	3	2,001	48.3 (46.2–50.5)	5.7 (4.7–6.9)	98.9 (98.4–99.3)	5.3 (3.3–8.5)	0.95 (0.94–0.97)
Pulmonary edema*[7,8,12,14,16,18–21,23,36,54,57,58,64]	15	4,393	46.6 (45.1–48.1)	56.9 (54.7–59.1)	89.2 (87.9–90.4)	4.8 (3.6–6.4)	0.48 (0.39–0.58)
Pleural effusion[12,20,58,60,72]	5	1,326	55.1 (52.4–57.8)	16.3 (13.7–19.2)	92.8 (90.4–94.7)	2.4 (1.6–3.6)	0.89 (0.80–0.99)
Enlarged cardiac silhouette[8,12,15,18,20,21,54,58,60,64–66]	12	3,515	51.7 (49.4–52.7)	74.7 (72.9–76.5)	61.7 (59.4–63.9)	2.3 (1.6–3.4)	0.43 (0.36–0.51)

LR+ = positive likelihood ratio; LR− = negative likelihood ratio.
*Refers to generalized pulmonary edema in studies that did not report specifically on both "alveolar edema" and "interstitial edema."

primary studies; the diagnostic accuracy of chest radiography in AHF was the focus of two secondary analyses.[66,72] Of the 18 included studies (N = 5,328 patients), eight studies[7,8,15,16,54,60,64,72] excluded patients with comorbid renal insufficiency or failure. Three of the included studies[14,66,72] evaluated interstitial and alveolar edema as separate radiographic variables, while others reported pulmonary edema as an inclusive variable. It was unclear who interpreted radiographs in the majority of included studies. Retrospective review of radiology reports served as the source of data in four studies.[14,15,66,72] Interobserver agreement in radiographic interpretation was not reported in any of the included studies.

The radiographic variable most commonly reported by the included studies was pulmonary edema. For the purpose of pooling results, the terms "condensation," "pulmonary congestion," "edema," pulmonary venous hypertension," "interstitial and/or alveolar edema," and "interstitial edema" (when alveolar edema was not an included study variable) were equated with "pulmonary edema." The positive LR associated with pulmonary edema was 4.8 (95% CI = 3.6 to 6.4; Table 2; forest plot, Data Supplement S2). Pulmonary edema, however, was an insensitive finding (sensitivity = 56.9%; 95% CI = 54.7 to 59.1%). The presence of enlarged cardiac silhouette and pleural effusions were less helpful findings.

QUADAS-2 analysis could be performed on 11[7,8,12,15,16,18,23,54,58,60,72] of the included studies (Data Supplement S2). Agreement between the two investigators applying the QUADAS-2 criteria was fair (κ = 0.40, 95% CI 0.24 to 0.56). Patient selection was at risk for decreased applicability in 73% of the studies because they excluded subgroups of dyspneic patients based on comorbidities. Risk of bias for the criterion standard diagnosis was rated high in each of the included studies because radiographic results were consistently incorporated into final diagnoses.

BNP/NT-proBNP

A total of 41[7,8,10,11,14,16–21,25,35,46,50,51,54,55,58,60–62,64,67,70,73–88] unique studies were included in this review: 35 were identified from the systematic review by Hill et al.;[29] four[25,49,79,83] from our 2012–2015 PubMED and EMBASE searches and two[18,78] from searches performed for other index tests as part of this systematic review (Data Supplement S2). Of the 41 included studies, 20 evaluated BNP,[11,14,16,18,19,21,25,35,50,51,54,55,58,60,74,76,78,84–86] 14 evaluated NT-proBNP,[7,8,10,17,20,49,61,62,64,70,79,81,83,88] and seven evaluated both BNP and NT-proBNP.[46,67,73,75,77,80,82] With the exception of three studies,[18,60,81] NPs were the primary index tests of interest. Three of the included studies were funded by companies that manufactured the NP assay under investigation.[7,35,83]

The main inclusion criterion for most studies was a chief complaint of dyspnea, although a few studies broadened this criterion to include patients with lower extremity edema[70] or signs or symptoms of AHF (dyspnea, edema, fatigue;[14,85] Data Supplement S3). A common exclusion criterion was renal insufficiency[8,16,25,35,51,60,62,64,70,74,78,88] or renal failure.[55,61,79] The criterion standard diagnosis of AHF was based on

clinical judgment alone or fulfillment of clinical criteria after independent chart review of hospitalization records. Blinding to all NP results was reported in 29 (69%) of the included studies.

Test characteristics were pooled according to assay manufacturers, given the systematic differences that have been shown among commercially available NP assays.[89] At the most commonly reported cutoff value of 100 pg/mL, BNP demonstrated high sensitivity and a negative LR less than 0.2 (forest plot, Data Supplement S2). The pooled sensitivity and specificity among the 19 studies[14,19,35,50,51,54,55,58,60,67,73–78,82,84,85] (9,143 patients) that used the Triage Biosite assay was 93.5% (95% CI = 92.6 to 94.2%) and 52.9% (95% CI = 51.6 to 54.2%), respectively. Specificity improved at a cutoff value of 500 pg/mL to 89.8% (95% CI = 88.5 to 91.1%), but at the cost of reduced sensitivity to 67.7% (95% CI = 65.5 to 69.9) (Table 3). Ten studies[7,20,46,64,73,75,77,81,83,87] using the Elecsys Roche immunoassay for NT-proBNP provided test characteristic data for the cutoff value of 300 pg/mL. The pooled negative LR at this cutoff was 0.09 (95% CI = 0.03 to 0.34; forest plot, Data Supplement S2). At a significantly higher cutoff value of 1550 pg/mL, specificity improved, but only to 72.9% (95% CI = 70.6 to 75.0%). Diagnostic performance data for other NP assays used in the included studies of this systematic review are shown in Table 3.

We were able to obtain patient-level data from the authors of five studies that evaluated BNP only,[14,16,51,76,78] of two studies that evaluated NT-proBNP only,[81,83] and of two studies that evaluated both BNP and NT-proBNP.[75,77] Data derived from common commercially available NP assays were pooled to calculate interval LRs. Data from the six unique studies[14,51,75–78] from which we were able to obtain patient-level BNP data (N = 2,423 patients) and from the five studies[46,75,77,81,83] that shared patient-level NT-proBNP data (N = 2,013 patients) were pooled. Interval LRs are shown in Table 4. Interval LRs of BNP concentrations at the low end of "positive" results as defined by a dichotomous cutoff of 100 pg/mL were substantially lower (100–200 pg/mL; 0.29, 95% CI = 0.23 to 0.38) than those associated with higher BNP concentrations (1000–1500 pg/mL; 7.12, 95% CI = 4.53 to 11.18). High NT-proBNP concentrations (150,000–300,000 pg/mL) were associated with an interval LR of 2.93 (95% CI = 1.95 to 4.39). The area under the curve using a summary receiver operating characteristic (ROC) curve based on patient-level results was 0.86 (95% CI = 0.83 to 0.86) for BNP and 0.76 (95% CI = 0.74 to 0.78) for NT-proBNP (Data Supplement S2).

According to the QUADAS-2 analysis performed by Hill et al.,[29] the only category rated as high risk of bias in the majority of included studies was patient selection. In four[25,49,64,79]of the five studies published after Hill et al.'s[29] systematic review (that evaluated more than one index test), patients with renal disease were excluded. By eliminating patients with this common comorbidity from the sample population, there are likely to be fewer false-positive NP results; spectrum bias in these studies results in inflated specificity.[42] Two studies[18,83] not included in Hill et al.'s[29] systematic review evaluated a single index

Table 3
Pooled Test Performance Characteristics for Natriuretic Peptides

Assay	Cutoff (pg/mL)	N	n	% AHF (95% CI)	Sensitivity % (95%CI)	Specificity % (95%CI)	LR+ (95% CI)	LR− (95% CI)
BNP								
Triage, Biosite	100[14,19,35,50,51,55,58,60,66,67,73–78,82,84,85]	19	9,143	44.7 (43.7–45.8)	93.5 (92.6–94.2)	52.9 (51.6–54.2)	2.2 (1.8–2.7)	0.11 (0.07–0.16)
	200[11,14,19,51,66,73,75–78,84]	11	3,279	50.4 (48.7–52.1)	85.9 (84.2–87.6)	72.2 (69.9–74.4)	3.1 (2.3–4.0)	0.18 (0.12–0.27)
	500[14,51,58,74–78]	8	3,915	46.7 (45.1–48.3)	67.7 (65.5–69.9)	89.8 (88.5–91.1)	9.1 (4.1–20.2)	0.34 (0.26–0.45)
AxSym, Abbott	100[18,21,36,46]	4	684	52.3 (48.6–56.1)	93.3 (90.2–95.7)	53.1 (47.5–58.6)	1.9 (1.5–2.4)	0.15 (0.08–0.29)
iSTAT, Abbott	100[85,86]	2	585	42.6 (38.6–46.6)	94.4 (90.7–96.9)	64.6 (59.2–69.7)	3.0 (1.2–7.4)	0.05 (0.02–1.23)
NT-proBNP								
Elecsys, Roche diagnostic	300[7,20,46,64,73,75,77,81,83,87]	10	3,498	45.0 (43.4–46.7)	90.4 (88.9–91.8)	38.2 (36.0–40.4)	1.8 (1.4–2.2)	0.09 (0.03–0.34)
	1,000[7,46,62,73,75,77,81,83]	8	2,988	44.8 (43.0–46.6)	84.8 (82.8–86.7)	65.5 (63.2–67.8)	2.7 (1.9–3.9)	0.20 (0.12–0.33)
	1,550[10,46,61,75,77,79–81,83]	9	3,043	37.3 (35.6–39.0)	75.5 (73.4–77.9)	72.9 (70.6–75.0)	3.1 (2.3–4.3)	0.32 (0.20–0.51)
Dimension, Dade Behring	300[70]	1	401	30.4 (26.0–35.2)	95.9 (90.7–98.6)	48.0 (42.0–54.1)	1.9 (1.6–2.1)	0.09 (0.04–0.20)

AHF = acute heart failure; BNP = B-type natriuretic peptide; LR+ = positive likelihood ratio; LR− = negative likelihood ratio; N = number of studies; n = number of patients; NT-proBNP = N-terminal proB-type natriuretic peptide.

Table 4
Interval LRs of BNP and NT-proBNP Values

BNP Value (pg/mL)	Interval LR	*N* (%)	NT-proBNP (pg/mL)	Interval LR	*N* (%)
0–100	0.14 (0.12–0.18)	617 (28)	0–100	0.09 (0.05–0.17)	150 (7.5)
100–200	0.29 (0.23–0.38)	308 (14)	100–300	0.23 (0.16–0.33)	205 (10.2)
200–300	0.89 (0.67–1.17)	188 (9)	300–600	0.28 (0.20–0.39)	212 (10.5)
300–400	1.34 (0.98–1.83)	148 (7)	600–900	0.63 (0.46–0.87)	151 (7.5)
400–500	2.05 (1.47–2.84)	148 (7)	900–1,500	0.84 (0.67–1.06)	249 (12.4)
500–600	3.50 (2.30–5.35)	115 (5)	1,500–3,000	1.49 (1.19–1.86)	273 (13.6)
600–800	4.13 (3.01–5.68)	218 (10)	3,000–5,000	2.36 (1.81–3.08)	225 (11.2)
800–1,000	5.00 (3.21–7.89)	130 (6)	5,000–10,000	2.48 (1.91–3.21)	239 (11.9)
1,000–1,500	7.12 (4.53–11.18)	160 (7)	10,000–15,000	2.84 (1.90–4.23)	112 (5.6)
1,500–2,500	8.33 (4.60–15.12)	105 (5)	15,000–30,000	2.93 (1.95–4.39)	111 (5.5)
2,500–5,001	8.91 (4.09–19.43)	65 (3)	30,000–20,0000	3.30 (2.05–5.31)	86 (4.3)
		2,202 (100)			2,013 (100)

BNP = B-type natriuretic peptide; LR = likelihood ratio; NT-proBNP = N-terminal proB-type natriuretic peptide.

test. The QUADAS-2 evaluation of these two studies is presented in Data Supplement S3.

Lung US

A total of 268 citations were identified by the PubMed and EMBASE searches, 26 of which were selected for full-text review. One ED-based study[53] was excluded because cardiologists rather than EPs performed and interpreted the lung US (Data Supplement S2). Studies that took place in European countries[23,24] in which emergency medicine was emerging as a new specialty at the time of patient enrollment were included if the physicians performing and interpreting lung US were considered EPs in their countries. A total of eight studies[23,24,48,62,63,81,90,91] (N = 1,918 patients) were selected for inclusion in this review. All studies included adults presenting to the ED with dyspnea. Study characteristics are described in Data Supplement S3. Russell et al.[63] selected only those patients in whom dyspnea was truly undifferentiated; patients in whom a diagnosis of AHF seemed clinically obvious were excluded. Training was limited to didactic and workshop sessions for most studies; in two of the studies[48,63] sonographers were fellowship-trained in emergency US. Study characteristics are presented in Data Supplement S3.

A positive lung US was defined in every study by the presence of at least three B lines in two bilateral lung zones. In six of the included studies,[23,24,48,62,63,81] the US protocol was based on scanning eight thoracic lung zones (four anterior and four lateral, as described by Volpicelli[92]). Two studies[90,91] modified this protocol to interrogate six anterior-lateral thoracic lung zones.

Diffuse pulmonary edema identified on lung US proved to be a diagnostic variable with discriminatory value (positive LR 7.4, 95% CI = 4.2 to 12.8; negative LR 0.16, 95% CI = 0.05 to 0.51). Statistical heterogeneity was high for these pooled estimates (I^2 = 78% and I^2 = 99%, respectively). Positive LRs among the eight included studies ranged from 2.8[63] to 19[24] (forest plot, Data Supplemental 2). Pleural effusions visualized on lung US were less helpful in diagnosing or excluding AHF (positive LR 2.0, 95% CI = 1.4 to 2.8).

The risk for bias in the criterion standard diagnosis was rated as high in the four studies[23,24,63,90] that could be evaluated by QUADAS-2 (Data Supplement S3) because standard clinical criteria were not applied to the criterion standard diagnosis in these studies. In only one of these studies[24] were lung US results incorporated into the criterion standard diagnosis. Risk of decreased applicability in the domain of patient selection was rated as high in three studies that excluded patients with comorbid conditions[24] or patients requiring mechanical ventilation.[23,63] Agreement between the two investigators applying the QUADAS-2 criteria was high[71] (κ = 0.93, 95% CI = 0.79 to 1.0).

Beside Echocardiography

After reviewing the full-text version of 18 studies, four[20,21,48,63] (N = 675patients) were selected for inclusion in this review (Data Supplement S3). Three ED-based studies[18,19,53] were excluded from this review because cardiologists rather than EPs performed or interpreted the sonograms (Data Supplement S2). EPs fellowship-trained in US performed the US examinations in the studies by Anderson et al.[48] and Russell et al.[63] The sonographers in the studies by Nazerian et al.,[20] Russell et al.,[63] and Wang et al[21] were blinded to clinical data. Inter-observer agreement was reported only by Wang et al.[21] (0.889). In all four included studies[20,21,48,63] the physicians who made the final diagnosis were blinded to the echocardiograms performed as index tests. Three studies[20,21,63] analyzed patients with suboptimal imaging as false negatives. To maintain consistency, these patients were not included in pooled estimates in our review.

A summary of the echocardiographic test characteristics analyzed in the four included studies[20,21,48,63] are reported in Table 5. Ejection fraction (EF) was determined by visual estimation in the studies by Anderson et al.,[48] Nazerian et al.,[20] and Russell et al.[63] Wang et al.[21] measured EF based on left ventricular dimensions at end-diastole and end-systole but did not report the test characteristics of EF as an index test for AHF in their study. Elevated left ventricular end-diastolic dimension, defined as > 28.6 mm/mm^2 by Wang et al.,[21] did not improve diagnostic accuracy when compared to visual estimation of EF. Only one study, by Nazerian et al.,[20] evaluated diastolic function, which found that a

Table 5
Pooled Test Performance Characteristics for Lung US and Beside Echocardiography Findings

	N	*n*	% AHF (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	LR+ (95% CI)	LR− (95% CI)
Lung US							
Positive B-line scan*[23, 24, 48, 62, 63, 81, 90, 91]	8	1914	48.2 (46.0–50.5)	85.3 (82.8–87.5)	92.7 (90.9–94.3)	7.4 (4.2–12.8)	0.16 (0.05–0.51)
Pleural effusion(s)[63,90]	2	155	40.7 (33.2–48.5)	63.5 (50.4–75.3)	71.7 (61.4–80.6)	2.0 (1.4–2.8)	0.49 (0.22–1.10)
Bedside echocardiography							
Restrictive mitral pattern*[20]	1	125	43.2 (34.9 -52.0)	81.5 (68.6–90.7)	90.1 (80.7–95.9)	8.3 (4.0–16.9)	0.21 (0.12–0.36)
Reduced EF[20,48,63]	3	325	41.2 (36.0–46.7)	80.6 (72.9–86.9)	80.6 (74.3–86.0)	4.1 (2.4–7.2)	0.24 (0.17–0.35)
Increased LV end-diastolic dimension†,[21]	1	84	58.3 (47.7–68.3)	79.6 (65.7–89.7)	68.6 (50.7–83.1)	2.5 (1.5–4.2)	0.30 (0.16–0.54)

Lung ultrasound: *defined as ≥ 2 bilateral lung zones with ≥ 3 B-lines per intercostal space.
Bedside echocardiography: *defined as E/A ratio > 2 or E/A between 1 and 2 and deceleration time (DT) < 130 msec; DT < 130 msec alone if atrial fibrillation. †Defined as LVEDD > 28.6 mm/mm2.
AHF = acute heart failure; EF = ejection fraction; LV = left ventricular; LR = likelihood ratio; *N* = number of studies; *n* = number of patients; US = ultrasound.

restrictive pattern on pulsed Doppler analysis of mitral inflow most accurately predicted AHF with positive and negative LRs of 8.3 (95% CI = 4.0 to 16.9) and 0.21 (95% CI = 0.12 to 0.36), respectively.

The QUADAS-2 criteria could only be applied to the study by Russell et al.,[63] because these criteria do not apply to studies that compare multiple index tests. Risk of bias in patient selection was rated as high due to their exclusion of patients with causes of dyspnea presumed to be obvious (a patient with known heart failure who was not compliant with taking medications) and those who had received prior treatment (Data Supplement S3).

Bioimpedance

Four studies[24,25,60,78] that evaluated the diagnostic accuracy of bioimpedance on ED patients (*N* = 1,039 patients) were selected from the 65 studies that were screened (Data Supplement S2). One study[60] evaluated conventional bioelectrical impedance analysis, while three[24,25,78] used bioelectrical impedance vector analysis (BIVA) in which resistance and reactance are plotted as a bivariate vector on a nomogram. Characteristics of the included studies are listed in Data Supplement S3. One[78] of the four studies was designed as a case-control study in which study groups were divided based on BNP value, clinical evidence of AHF, and respiratory symptoms; however, for our pooled analyses the author provided us with the final criterion standard diagnoses of AHF/not AHF and the associated hydration index values for each patient.

Segmental resistance measures thoracic fluid status using sensing and output electrodes placed on the anterior thigh and suprasternal notch.[93] Segmental resistance values lower than the cutoff determined by ROC curve intercept offered the highest positive LR (10.6, 95% CI = 5.8 to 19.2; Table 6). BIVA, which accounts for age, sex, and body mass index, offered lower positive LRs (Table 6).

QUADAS-2 criteria could be applied to the two[24,60] included studies that evaluated a single index test (Data Supplement S3). Risk of bias in patient selection was rated as high in these studies because they excluded patients with comorbid conditions that caused ascites and peripheral edema.

Test–Treatment Threshold Estimates

Test–treatment threshold models contextualize a diagnostic test in a clinical setting by linking the discriminatory value of a test with the benefits and risks associated with a given treatment. Evidence regarding the clinical benefits and risks of pharmacologic treatments for AHF, however, is very limited.[94] No pharmacologic therapy for the treatment of acute AHF has been given a class I/level of evidence A recommendation from the American Heart Association or Heart Failure Society of America.[13,95] Randomized controlled trials investigating the effect of loop diuretics on clinically relevant outcomes are lacking. In the DOSE (Diuretic Optimization Strategies Evaluation in Acute Heart Failure) trial,[96] high-dose intravenous diuretic therapy was associated with a higher risk of increasing serum creatinine within 72 hours of treatment initiation compared with low-dose diuretic therapy, but this increase was not observed at 60 days.

A recent Cochrane review on nitrate vasodilator therapy showed no difference in outcomes compared with alternative interventions, but this review reflects a paucity of available evidence.[97] In the Vasodilation in the Management of Acute CHF (VMAC) trial[98] comparing intravenous nitroglycerin with nesiritide and with placebo, 10% of the 216 patients receiving intravenous nitroglycerin experienced symptomatic hypotension; mean doses 3 hours after treatment initiation were 42 μg/min in catheterized patients and 29 μg/min in noncatheterized patients. These studies, however, may not provide an accurate estimation of the benefits and risks of diuretics and nitrates when administered early in the course of managing a patient in the ED. Some of the patients included in these studies were enrolled well after their initial presentation and early management. Doses studied in the VMAC trial[98] may be lower than those administered to hypertensive patients with AHF in the ED. A test–treatment threshold model based on these numbers might be less relevant to the EP.

The estimated benefits and risks of common pharmacologic interventions such as intravenous diuretics or nitrate vasodilator therapy that we chose to apply to a test–treatment threshold model, therefore, are hypothetical and used for illustrative purposes. Mathematically, the test–treatment threshold is based on the relative costs of treating patients without disease C (i.e., hypotension from vasodilator and diuretic therapy) and failing to treat those with disease B. Higher estimated risk of heart failure treatment C relative to treatment benefit B results in a higher treatment threshold. If the risk of treating dyspneic patients without underlying AHF were estimated to be 0.45 and the risk of failing to treat AHF patients expeditiously were 0.19, then the treatment threshold, or posterior probability of disease (P_{TT}), at which the costs of these two risks (B and C) would be balanced is 70%. This assumes that the cost of NP testing (including harm to the patient) is negligible. Figure 2 displays the posterior probabilities yielded by different diagnostic tests when starting at a pretest probability of 46%. Those diagnostic tests with positive LRs large enough to yield a posterior probability greater than 70% would guide the Bayesian clinician to treat for AHF.

Applying the Pauker and Kassirer[45] threshold model to a diagnostic test with a dichotomous outcome also allows one to start with a posterior probability (treatment threshold; P_{TT}) and using the positive and negative LRs associated with the test, calculate the range of pretest probabilities for which the diagnostic test has the potential to change the decision to treat for AHF. This range is defined by the limits of the no treat–test threshold and the test–treatment threshold. Starting with a posterior treatment threshold probability of 70% and using the positive and negative LRs of lung US, the no treat–test threshold is 24% and the test–treat threshold is 94% (Figure 3). When the pretest probability lies between these thresholds, lung US has the potential to affect management. Different values for estimated risks and benefits of administering AHF treatment can be

Table 6
Pooled Test Performance Characteristics for Bioimpedance Variables

	No. of Studies	No. of Patients	% AHF (95% CI)	Sensitivity, % (95%CI)	Specificity, % (95%CI)	LR+ (95% CI)	LR− (95% CI)
Segmental BIA (≤54 Ohms) 60	1	292	58.9 (53.0–64.6)	88.4 (82.6–92.8)	91.7 (85.2–95.9)	10.6 (5.8–19.2)	0.13 (0.08–0.19)
Whole body BIA (≤441 Ohms) 60	1	292	58.9 (53.0–64.6)	65.1 (57.5–72.2)	90.0 (83.2–94.7)	6.5 (3.8–11.3)	0.39 (0.31–0.48)
BIVA (Z(Xc) – 1SD) 24	1	315	53.7 (48.1–59.1)	69.2 (61.7–76.1)	78.8 (71.2–85.1)	3.3 (2.4–4.5)	0.44 (0.37–0.54)
BIVA, HI (73.4%) 25,78	2	422	69.7 (65.1–73.9)	81.6 (77.8–85.1)	66.1 (60.1–71.6)	2.0 (1.2–3.3)	0.34 (0.12–0.65)

BIA = bioelectrical impedance analysis; BIVA = bioelectrical impedance vector analysis; HI = hydration index; Z(Xc) – 1SD = 1 standard deviation below mean Z-score vector (reactance).

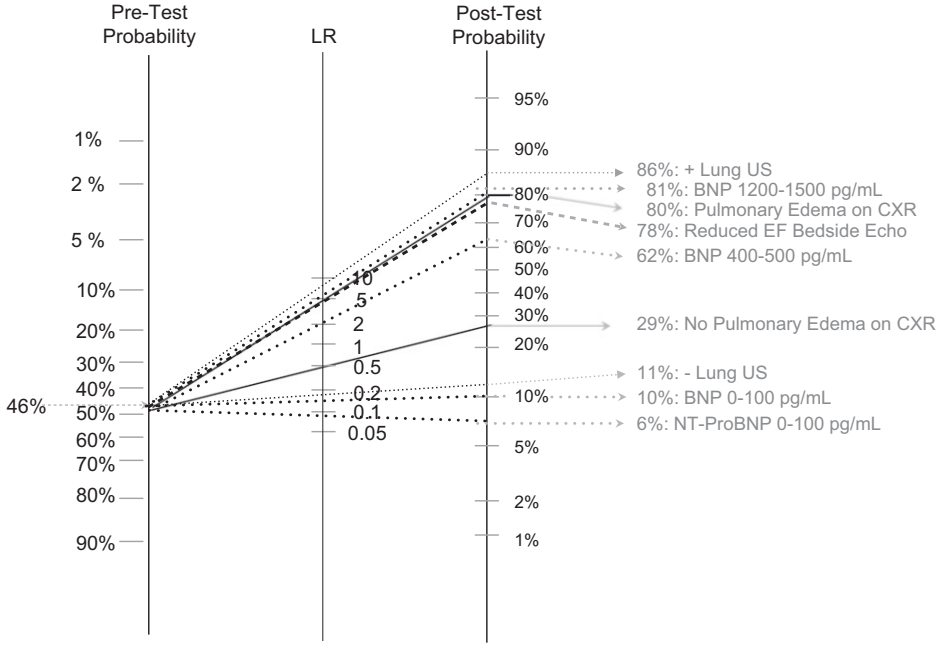

Figure 2. Posterior probabilities yielded by different diagnostic tests when starting at a pretest probability of 46%. BNP = B-type natriuretic peptide; CXR = chest radiograph; LR = likelihood ratio; NT-proBNP = N-terminal proB-type natriuretic peptide.

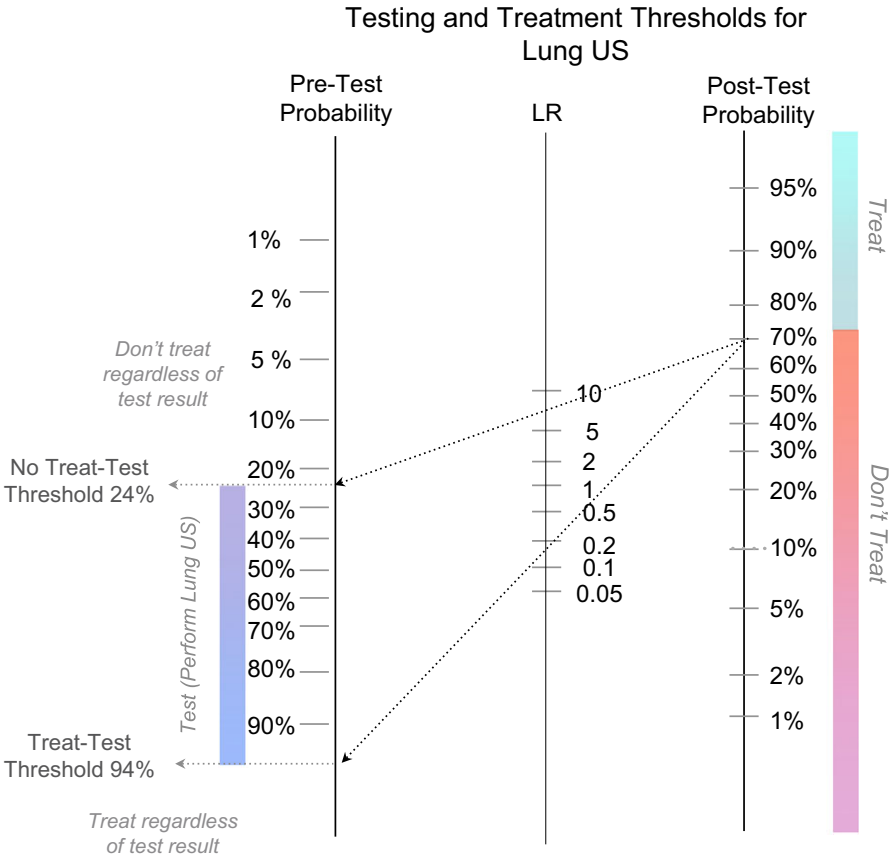

Figure 3. Testing and treatment thresholds for lung US. LR = likelihood ratio; US = ultrasound.

incorporated into the Pauker and Kassirer model[45] using a Microsoft Excel calculator (Data Supplement S4).

DISCUSSION

The diagnosis of AHF in the adult ED population remains challenging. This systematic review demonstrates that clinical elements such as past medical history, presenting symptoms, and physical examination findings, on their own, cannot be relied on for excluding or establishing the diagnosis. It also demonstrates that symptoms commonly sought in a clinical history such as orthopnea, paroxysmal nocturnal dyspnea, and weight gain fail to distinguish patients with AHF. S3 gallop, the physical examination finding most suggestive of AHF, is only 13% sensitive. Rales and peripheral edema are even less suggestive of AHF. This review did not evaluate the diagnostic accuracy of historical elements, symptoms, or examination findings in combination. Clinical gestalt, based on an aggregate effect of the history and physical examination, likely outperforms these diagnostic elements in isolation and plays an important role in determining the pretest probability of AHF. However, greater awareness of the limitations associated with individual variables may help to avoid diagnostic overconfidence and some of the biases inherent in a heuristic approach to formulating an initial diagnosis.

Our review shows that the ECG, often available within minutes of patient arrival, does little to alter the probability of AHF. However, none of the studies included in this review investigated QRS amplitude, a parameter that has been shown to attenuate with worsening heart failure.[99,100] Chest radiography is considered a fundamental component of the ED workup for AHF but radiographic signs of pulmonary edema and vascular redistribution are often absent in AHF patients. The poor sensitivity of CXR findings for diagnosing pulmonary edema has been previously described.[26,101] Even in the presence of severely elevated pulmonary capillary wedge pressures in patients with heart failure, radiographic pulmonary congestion is absent 39% of the time.[102]

The data in this systematic review are consistent with that of prior studies demonstrating that at recommended respective cut-points of 100 and 300 pg/mL,[103] BNP and NT-proBNP testing are most useful for excluding AHF.[29,104] Calculation of interval LRs from patient-level NP results, however, is unique to this review and helps provide a more clinically intelligible interpretation of NP test performance. This review highlights the shortcomings of dichotomizing continuous variables into binary outcomes above and below a single cutoff point. The positive LR associated with a BNP of 150 pg/mL, when applying pooled binary data relating to the single cutoff value of 100 pg/mL (Triage, Biosite, Table 3), is 2.2 (95% CI = 1.8 to 2.7). Based on this value, a diagnosis of AHF would be favored because it is statistically grouped with substantially higher BNP values in AHF patients. Using interval LR data, however, a

BNP value of 150 pg/mL is associated with an LR of 0.29 (95% CI = 0.23 to 0.38) and thus favors an alternative diagnosis (Table 4). Only above a cut-point of 800 pg/mL for BNP does the interval LR substantially increase the posttest probability of AHF. The interval LRs for NT-proBNP only modestly favor the diagnosis of AHF even at extremely elevated levels.

This review focuses on the diagnostic accuracy of NPs. The utility of these biomarkers, however, is based on whether or not their use results in clinical benefit.[105] Several randomized controlled trials that have evaluated the clinical performance of NPs have failed to demonstrate differences in patient-centered clinical outcomes.[106–108] One possible explanation is that NPs add little value beyond the clinical judgment of an EP. When the diagnosis of AHF is relatively certain, clinical judgment has been shown to be more accurate than the BNP result.[9] In the subset of dyspneic patients for whom the diagnosis of AHF is uncertain, the performance of NPs is suboptimal.[109] NPs may prove to have greater clinical utility when used with a Bayesian approach and interpreted as a continuous rather than as a dichotomous variable.[87]

Given the limitations of NP values above recommended "rule-out" cut-points, point-of-care lung US can play a potentially significant role in the ED diagnosis of AHF in the ED. The diagnostic performance of lung US in this review was superior to any other diagnostic test that was studied in more than one cohort of patients. With a pooled positive LR of 7.4 (95% CI = 4.2 to 12.8) and negative LR of 0.16 (95% CI = 0.05 to 0.51), lung US demonstrates potential to both rule in and exclude the diagnosis of AHF. Statistical heterogeneity among these studies was substantial, and this should be taken into account while interpreting these summary estimates. Spectrum bias resulting from the exclusion of patients with alternative causes of dyspnea[62] and peripheral edema[24] should also be considered.

Perhaps the most robust data in our pooled lung US sample comes from the multicenter study by Pivetta et al.[23] (N = 1,005). Patient exclusions were limited to those with initially obvious causes of dyspnea (traumatic pneumothorax) and intubated patients. Lung US in this study was performed by 62 EPs across community and academic hospitals, and the criterion standard diagnosis of AHF at hospital discharge was determined with high inter-rater agreement (κ = 0.93). The positive and negative LRs determined in this study were 14.0 (95% CI = 10.2 to 19.3) and 0.10 (95% CI = 0.08 to 0.14), respectively. Incorporation of lung US into the classification of AHF in this study led to a net reclassification improvement of 19% (95% CI = 14.6 to 23.6%). The authors of this study report that the vast majority of lung US examinations were performed within 40 minutes of ED presentation. The feasibility of lung US to rapidly identify pulmonary edema in real time shortly after ED presentation and before therapeutic intervention may increase the sensitivity of this test.

In a recent meta-analysis, Al Deeb et al.[22] evaluated the test performance of lung US in diagnosing AHF. The pooled positive and negative LRs from this meta-analysis were 12.4 (95% CI = 5.7 to 26.8) and 0.06 (95% CI = 0.02 to 0.22), respectively. However, only three of the seven studies[62,81,90] included in their analysis took place in the ED. Two[110,111] of the seven included studies took place in the intensive care unit (ICU) where spectrum bias might increase the sensitivity of pathologic B-lines. While the prevalence of other causes of diffuse sonographic B-lines such as interstitial lung disease, multifocal pneumonia, and severe acute respiratory distress syndrome might decrease the specificity of this pattern for AHF in the ICU,[112] higher estimates of specificity in these studies may be related to study design. The test characteristics of B-lines in one ICU study were evaluated in predefined groups of patients with pulmonary edema, COPD, and no cardiopulmonary disorder.[111] In the other ICU study, sonography was performed by experts in thoracic US who used sonographic artifacts to classify other causes of dyspnea including pneumonia, venous thromboembolism, and pneumothorax.[110]

Echocardiography is considered key to the diagnosis of chronic heart failure, but limited availability and echocardiographic experience among EPs have limited its role in the diagnostic evaluation of dyspnea in the ED setting. Echocardiography is more technically challenging and complex than lung US, but perhaps the simplest echocardiographic assessment relating to heart failure is the visual estimation of EF. EF data was based on visual estimation in the three included studies[20,48,63] that reported this variable. Visual estimation of EF by both cardiologists[112] and EPs[113] has been shown to correlate well with quantitative assessments of EF. However, use of reduced EF alone as an echocardiographic variable for predicting AHF would result in the failure to detect the 50% of heart failure patients with preserved EF.[114] Further, many patients also have a prior EF available in the medical record, which may limit the amount of new information provided by an EF recorded in the ED.

Pulsed Doppler analysis of mitral inflow enables the evaluation of relative velocities during early and late diastolic filling and provides a surrogate measure of elevated left ventricular filling pressures. A restrictive pattern of diastolic filling in the single included study was associated with a positive LR of 8.3 (95% CI = 4.0 to 16.9) for the diagnosis of AHF. However, mitral inflow analysis cannot reliably be applied to patients who are tachycardic or have permanent pacing or mitral valve prostheses, and standard definitions cannot be applied to patients in atrial fibrillation.[115] Accurate mitral inflow analysis is also highly dependent on preload, leading to an increasingly common use of combined approaches that incorporate tissue Doppler imaging techniques.[116] Perhaps most importantly, acquiring mitral inflow and tissue Doppler data may be beyond the scope of many EPs who do not have formal fellowship training. Given this, further studies investigating the ability of a restrictive pattern to predict AHF and the ability of EPs with minimal echocardiographic training to accurately perform and interpret Doppler analysis are needed.

Bioimpedance analysis has emerged as a potential noninvasive assessment of pulmonary congestion and peripheral edema, based on the theory that intrathoracic fluid decreases resistance to an applied electrical current.[117] Both thoracic and whole body impedance

devices have been evaluated. While many implantable devices incorporate a data channel to measure thoracic impedance, EPs rarely have the output from the device while the patient is being evaluated in the ED. Therefore, this review focuses on measurements taken by external monitoring devices. Barriers to routine adoption thus far have included the time that it takes to acquire the data and requirement for the patient to remain relatively motionless for up to several minutes.[118]

Implications for Future Research

Given the suboptimal accuracy with which AHF is discriminated from other causes of dyspnea in the ED, clinical decision aids based on multiple logistic regression modeling may have a role in assisting with the clinical diagnosis of AHF. Useful models will likely require several diagnostic components and should incorporate rather than take the place of clinical gestalt. One such model based on age, pretest probability, and NT-proBNP intervals has demonstrated diagnostic accuracy in both internally derived and externally validated cohorts.[87] Lung US and bedside echocardiography are diagnostic tests to consider including in different decision models and NPs are likely best utilized to exclude the diagnosis of AHF. We could foresee other biomarkers and diagnostic tests quantifying pathophysiology and assisting with the diagnosis of AHF by excluding other causes such as infection, acute kidney injury, or obstructive pulmonary disease. The clinical utility of these new diagnostic adjuncts would need to be evaluated alongside current diagnostic modalities.

Future diagnostic studies should adhere to the STARD reporting guidelines to make more explicit the potential for bias.[47] Ideally, diagnostic studies for AHF would avoid overt spectrum bias by including patients with comorbidities such as renal failure. Reporting test characteristics as they pertain to discrete intervals of biomarker values in future studies would be more informative than reporting data relating to a single cutoff value.[87] As shown in this review, dichotomization of continuous variables can produce distorted and exaggerated LRs. Dichotomous results from cutoff values in future biomarker and bioimpedance studies may be reported for expediency. Results based on cutoff values that are determined a priori rather than retrospectively derived from intercepts of ROC curves are likely to be more generalizable. Diagnostic test results that are based on interpreter judgment should be reported with statistical data reflecting inter-rater reliability. Until a better criterion standard for the diagnosis of AHF becomes available, a clinical diagnosis made by two reviewers blinded to the index test will have to serve as the criterion standard.

LIMITATIONS

The most important limitations of this systematic review apply to the criterion standard diagnosis of AHF. The criterion standard for AHF in the included studies was a clinical diagnosis made by two or more clinicians after medical record review. In the absence of a better criterion standard for the diagnosis of AHF, a clinical diagnosis that combines a subjective assessment of a patient with some combination of objective data points must be relied on as a criterion standard. Estimations of the diagnostic accuracy of an index test for AHF are only as valid as the criterion standard diagnoses are accurate. The credibility of the criterion standard diagnosis is likely to vary among the studies included in this review. Incorporation of echocardiographic data, for example, into the diagnostic workup and criterion standard diagnosis among the included studies was variable.

Although our search strategies for the electronic databases used in this systematic review were constructed by an experienced health sciences librarian using accepted principles of health information science, it is still possible that they may have missed some studies eligible for inclusion. However, the comprehensive nature of our literature search makes it unlikely that we have excluded any important studies whose findings would significantly alter the conclusions of this review. In addition, there is the possibility of language bias in the findings of this review because non-English publications were excluded. However, this type of exclusion has not been shown to significantly affect the results of meta-analyses.[119]

Data relating to all diagnostic tests reported by a study were included in the meta-analysis, even if some of these diagnostic tests were not the primary index test evaluated by that study. We analyzed multiple tests derived from a single study as independent diagnostic variables to aggregate results according to diagnostic category. Coming from the same sample of patients, which has a distinct composition and spectrum of disease, it is unlikely that the individual variables investigated in this sample behave independently with respect to test performance. Likewise, to think of the summary LRs for each diagnostic test as stand-alone values would overlook their interdependence and derivation from the same studies. Also, data collection relating to other diagnostic tests included in a study may not have been as rigorous as that applied to the study's primary index test. Data from those diagnostic tests other than the primary index test under a study's investigation likely factored into the criterion standard diagnosis. Incorporation bias challenges the validity of these data, potentially increasing the sensitivity and specificity of the diagnostic test.[42,43]

The quality of studies that were included in the meta-analysis was variable. Differences in inclusion and exclusion criteria among the included studies put them at varying degrees of risk for spectrum bias.[42] Appraisal of pooled results should factor in the clinical as well as statistical heterogeneity among included studies.

A limitation specific to the NP analyses in this review was our lack of consideration of age as a variable affecting BNP and NT-proBNP values. Data relating to the specificity of these tests may have been more accurate if age-based cutoffs were evaluated.[7,12,120,121] Other variables known to affect NP values, such as renal function,[75,77,80,122,123] were also not factored in to our analysis of diagnostic accuracy.

Last, the test–treatment threshold model suggested in this review is a conceptual model. It is limited by the

paucity of available data on which to base estimates for the clinical benefits and risks of treatment. However, it may be useful for the EP to apply this model to patient disposition; the posterior treatment threshold can be thought of as a threshold above which the physician, confident that AHF is the underlying cause of dyspnea, can stop testing. Below this threshold, the physician would be compelled to seek further evidence in the ED favoring the diagnosis of AHF.

CONCLUSIONS

Elements of clinical history, symptoms, physical examination, chest radiography, and electrocardiography, on their own, lack discriminatory value in making or excluding the diagnosis of acute heart failure in ED patients. B-type natriuretic peptide and N-terminal proB-type natriuretic peptide are most valuable in ruling out acute heart failure when values are lower than the suggested cutoff points of 100 and 300 pg/mL, respectively. Values above these thresholds may be less helpful for establishing the diagnosis than previously described. Lung ultrasound appears to have the best combination of test characteristics with the presence or absence of diffuse B-lines providing reliable information to confirm or exclude the ED diagnosis of acute heart failure. Other diagnostic modalities such as bioimpedance may have future utility in the ED setting, although more study is needed.

We thank Christopher Stewart, senior assistant librarian at SUNY Downstate Medical Center, for his help in conducting the literature searches for this systematic review. We thank the following authors for providing us with patient-level natriuretic peptide data: Tommy Chung, Camille Chenevier-Gobeaux, Joël Coste, Christopher R. deFilippi, Salvatore Di Somma, Andrew S. Liteplo, Simcha R. Meisel, and Thomas Mueller.

References

1. Gheorghiade M, Zannad F, Sopko G, et al. Acute heart failure syndromes: current state and framework for future research. Circulation 2005;112:3958–68.
2. Adams KF Jr, Fonarow GC, Emerman CL, et al. Characteristics and outcomes of patients hospitalized for heart failure in the United States: rationale, design, and preliminary observations from the first 100,000 cases in the Acute Decompensated Heart Failure National Registry (ADHERE). Am Heart J 2005;149:209–16.
3. Fonarow GC, Abraham WT, Albert NM, et al. Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF): rationale and design. Am Heart J 2004;148:43–51.
4. Storrow AB, Jenkins CA, Self WH, et al. The burden of acute heart failure on U.S. emergency departments. JACC. Heart Failure 2014;2:269–77.
5. Lien CT, Gillespie ND, Struthers AD, McMurdo ME. Heart failure in frail elderly patients: diagnostic difficulties, co-morbidities, polypharmacy and treatment dilemmas. Eur J Heart Failure 2002;4:91–8.
6. Hawkins NM, Petrie MC, Jhund PS, Chalmers GW, Dunn FG, McMurray JJ. Heart failure and chronic obstructive pulmonary disease: diagnostic pitfalls and epidemiology. Eur J Heart Failure 2009;11:130–9.
7. Januzzi JL Jr, Camargo CA Jr, Anwaruddin S, et al. The N-terminal Pro-BNP investigation of dyspnea in the emergency department (PRIDE) study. Am J Cardiol 2005;95:948–54.
8. Bayes-Genis A, Santalo-Bel M, Zapico-Muniz E, et al. N-terminal probrain natriuretic peptide (NT-proBNP) in the emergency diagnosis and in-hospital monitoring of patients with dyspnoea and ventricular dysfunction. Eur J Heart Failure 2004;6:301–8.
9. McCullough PA, Nowak RM, McCord J, et al. B-type natriuretic peptide and clinical judgment in emergency diagnosis of heart failure: analysis from Breathing Not Properly (BNP) Multinational Study. Circulation 2002;106:416–22.
10. Ray P, Arthaud M, Birolleau S, et al. Comparison of brain natriuretic peptide and probrain natriuretic peptide in the diagnosis of cardiogenic pulmonary edema in patients aged 65 and older. J Am Geriatr Soc 2005;53:643–8.
11. Villacorta H, Duarte A, Duarte NM, et al. The role of B-type natriuretic peptide in the diagnosis of congestive heart failure in patients presenting to an emergency department with dyspnea. Arq Bras Cardiol 2002;79:569–72.
12. Berdague P, Caffin PY, Barazer I, et al. Use of N-terminal prohormone brain natriuretic peptide assay for etiologic diagnosis of acute dyspnea in elderly patients. Am Heart J 2006;151:690–8.
13. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013;62:e147–239.
14. Collins SP, Lindsell CJ, Peacock WF, et al. The combined utility of an S3 heart sound and B-type natriuretic peptide levels in emergency department patients with dyspnea. J Cardiac Failure 2006;12:286–92.
15. Collins SP, Peacock WF, Lindsell CJ, et al. S3 detection as a diagnostic and prognostic aid in emergency department patients with acute dyspnea. Ann Emerg Med 2009;53:748–57.
16. Lokuge A, Lam L, Cameron P, et al. B-type natriuretic peptide testing and the accuracy of heart failure diagnosis in the emergency department. Circ Heart Failure 2010;3:104–10.
17. Robaei D, Koe L, Bais R, Gould I, Stewart T, Tofler GH. Effect of NT-proBNP testing on diagnostic certainty in patients admitted to the emergency department with possible heart failure. Ann Clin Biochem 2011;48:212–7.
18. Huang CH, Tsai MS, Hsieh CC, Wang TD, Chang WT, Chen WJ. Diagnostic accuracy of tissue Doppler echocardiography for patients with acute heart failure. Heart 2006;92:1790–4.

19. Logeart D, Saudubray C, Beyne P, et al. Comparative value of Doppler echocardiography and B-type natriuretic peptide assay in the etiologic diagnosis of acute dyspnea. J Am Coll Cardiol 2002;40:1794–800.
20. Nazerian P, Vanni S, Zanobetti M, et al. Diagnostic accuracy of emergency Doppler echocardiography for identification of acute left ventricular heart failure in patients with acute dyspnea: comparison with Boston criteria and N-terminal prohormone brain natriuretic peptide. Acad Emerg Med 2010;17:18–26.
21. Wang HK, Tsai MS, Chang JH, Wang TD, Chen WJ, Huang CH. Cardiac ultrasound helps for differentiating the causes of acute dyspnea with available B-type natriuretic peptide tests. Am J Emerg Med 2010;28:987–93.
22. Al Deeb M, Barbic S, Featherstone R, Dankoff J, Barbic D. Point-of-care ultrasonography for the diagnosis of acute cardiogenic pulmonary edema in patients presenting with acute dyspnea: a systematic review and meta-analysis. Acad Emerg Med 2014; 21:843–52.
23. Pivetta E, Goffi A, Lupia E, et al. Lung ultrasound-implemented diagnosis of acute decompensated heart failure in the ED: a SIMEU multicenter study. Chest 2015; 148:202–10.
24. Piccoli A, Codognotto M, Cianci V, et al. Differentiation of cardiac and noncardiac dyspnea using bioelectrical impedance vector analysis (BIVA). J Cardiac Failure 2012;18:226–32.
25. Di Somma S, Lalle I, Magrini L, et al. Additive diagnostic and prognostic value of bioelectrical impedance vector analysis (BIVA) to brain natriuretic peptide 'grey-zone' in patients with acute heart failure in the emergency department. Eur Heart J Acute Cardiovasc Care 2014;3:167–75.
26. Wang CS, FitzGerald JM, Schulzer M, Mak E, Ayas NT. Does this dyspneic patient in the emergency department have congestive heart failure? JAMA 2005;294:1944–56.
27. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015;349:g7647.
28. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283:2008–12.
29. Hill SA, Booth RA, Santaguida PL, et al. Use of BNP and NT-proBNP for the diagnosis of heart failure in the emergency department: a systematic review of the evidence. Heart Failure Rev 2014;19:421–38.
30. McKee PA, Castelli WP, McNamara PM, Kannel WB. The natural history of congestive heart failure: the Framingham study. N Engl J Med 1971;285:1441–6.
31. Carlson KJ, Lee DC, Goroll AH, Leahy M, Johnson RA. An analysis of physicians' reasons for prescribing long-term digitalis therapy in outpatients. J Chron Dis 1985;38:733–9.
32. Schocken DD, Arrieta MI, Leaverton PE, Ross EA. Prevalence and mortality rate of congestive heart failure in the United States. J Am Coll Cardiol 1992;20:301–6.
33. Rosamond WD, Chang PP, Baggett C, et al. Classification of heart failure in the atherosclerosis risk in communities (ARIC) study: a comparison of diagnostic criteria. Circ Heart Failure 2012;5:152–9.
34. Fonseca C, Oliveira AG, Mota T, et al. Evaluation of the performance and concordance of clinical questionnaires for the diagnosis of heart failure in primary care. Eur J Heart Failure 2004;6(813–20):21–2.
35. Maisel AS, Krishnaswamy P, Nowak RM, et al. Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure. N Engl J Med 2002;347:161–7.
36. Collins SP, Lindsell CJ, Yealy DM, et al. A comparison of criterion standard methods to diagnose acute heart failure. Congest Heart Failure 2012;18:262–71.
37. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med Res Methodol 2006;6:31.
38. DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled Clin Trials 1986;7:177–88.
39. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005;58:882–93.
40. Brown MD, Reeves MJ. Evidence-based emergency medicine/skills for evidence-based emergency care. Interval likelihood ratios: another advantage for the evidence-based diagnostician. Ann Emerg Med 2003;42:292–7.
41. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36.
42. Kohn MA, Carpenter CR, Newman TB. Understanding the direction of bias in studies of diagnostic test accuracy. Acad Emerg Med 2013;20:1194–206.
43. Worster A, Carpenter C. Incorporation bias in studies of diagnostic tests: how to avoid being biased about bias. Can J Emerg Med 2008;10:174–5.
44. Dickstein K, Cohen-Solal A, Filippatos G, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2008: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2008 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association of the ESC (HFA) and endorsed by the European Society of Intensive Care Medicine (ESICM). Eur Heart J 2008;29:2388–442.

45. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med 1980;302:1109–17.
46. Mueller T, Gegenhuber A, Poelz W, Haltmayer M. Diagnostic accuracy of B type natriuretic peptide and amino terminal proBNP in the emergency diagnosis of heart failure. Heart 2005;91:606–12.
47. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. Clin Chem 2003;49:1–6.
48. Anderson KL, Jenq KY, Fields JM, Panebianco NL, Dean AJ. Diagnosing heart failure among acutely dyspneic patients with cardiac, inferior vena cava, and lung ultrasonography. Am J Emerg Med 2013;31:1208–14.
49. Behnes M, Brueckmann M, Lang S, et al. Connective tissue growth factor (CTGF/CCN2): diagnostic and prognostic value in acute heart failure. Clin Res Cardiol 2014;103:107–16.
50. Choi S, Park D, Lee S, Hong Y, Kim S, Lee J. Cutoff values of B-type natriuretic peptide for the diagnosis of congestive heart failure in patients with dyspnoea visiting emergency departments: a study on Korean patients visiting emergency departments. Emerg Med J 2007;24:343–7.
51. Chung T, Sindone A, Foo F, et al. Influence of history of heart failure on diagnostic performance and utility of B-type natriuretic peptide testing for acute dyspnea in the emergency department. Am Heart J 2006;152:949–55.
52. Gruson D, Ketelslegers JM, Verschuren F, Thys F. Head-to-head comparison of the prohormone proBNP1-108 with BNP and Nt-proBNP in patients admitted to emergency department. Clin Biochem 2012;45:249–52.
53. Kajimoto K, Madeen K, Nakayama T, Tsudo H, Kuroda T, Abe T. Rapid evaluation by lung-cardiac-inferior vena cava (LCI) integrated ultrasound for differentiating heart failure from pulmonary disease as the cause of acute dyspnea in the emergency setting. Cardiovasc Ultrasound 2012;10:49.
54. Knudsen CW, Riis JS, Finsen AV, et al. Diagnostic value of a rapid test for B-type natriuretic peptide in patients presenting with acute dyspnoe: effect of age and gender. Eur J Heart Failure 2004;6:55–62.
55. Maisel A, Mueller C, Nowak R, et al. Mid-region pro-hormone markers for diagnosis and prognosis in acute dyspnea: results from the BACH (Biomarkers in Acute Heart Failure) trial. J Am Coll Cardiol 2010;55:2062–76.
56. Marantz PR, Kaplan MC, Alderman MH. Clinical diagnosis of congestive heart failure in patients with acute dyspnea. Chest 1990;97:776–81.
57. Miller JB, Sen A, Strote SR, et al. Inferior vena cava assessment in the bedside diagnosis of acute heart failure. Am J Emerg Med 2012;30:778–83.
58. Morrison LK, Harrison A, Krishnaswamy P, Kazanegra R, Clopton P, Maisel A. Utility of a rapid B-natriuretic peptide assay in differentiating congestive heart failure from lung disease in patients presenting with dyspnea. J Am Coll Cardiol 2002;39:202–9.
59. Mueller C, Frana B, Rodriguez D, Laule-Kilian K, Perruchoud AP. Emergency diagnosis of congestive heart failure: impact of signs and symptoms. Can J Cardiol 2005;21:921–4.
60. Parrinello G, Paterna S, Di Pasquale P, et al. The usefulness of bioelectrical impedance analysis in differentiating dyspnea due to decompensated heart failure. J Card Failure 2008;14:676–86.
61. Potocki M, Breidthardt T, Reichlin T, et al. Comparison of midregional pro-atrial natriuretic peptide with N-terminal pro-B-type natriuretic peptide in the diagnosis of heart failure. J Intern Med 2010;267:119–29.
62. Prosen G, Klemen P, Strnad M, Grmec S. Combination of lung ultrasound (a comet-tail sign) and N-terminal pro-brain natriuretic peptide in differentiating acute heart failure from chronic obstructive pulmonary disease and asthma as cause of acute dyspnea in prehospital emergency setting. Crit Care 2011;15:R114.
63. Russell FM, Ehrman RR, Cosby K, et al. Diagnosing acute heart failure in patients with undifferentiated dyspnea: a lung and cardiac ultrasound (LuCUS) protocol. Acad Emerg Med 2015;22:182–91.
64. Shaikh K, Hanif B, Siddique AA, Shaikh MY, Khan MN. Pro-brain natriuretic peptide plasma levels, left ventricular dimensions and ejection fraction in acute dyspnoea. J Coll Physicians Surg Pak 2012;22:751–5.
65. Steg PG, Joubin L, McCord J, et al. B-type natriuretic peptide and echocardiographic determination of ejection fraction in the diagnosis of congestive heart failure in patients with acute dyspnea. Chest 2005;128:21–9.
66. Knudsen CW, Omland T, Clopton P, et al. Diagnostic value of B-type natriuretic peptide and chest radiographic findings in patients with acute dyspnea. Am J Med 2004;116:363–8.
67. Lainchbury JG, Campbell E, Frampton CM, Yandle TG, Nicholls MG, Richards AM. Brain natriuretic peptide and N-terminal brain natriuretic peptide in the diagnosis of heart failure in patients with acute shortness of breath. J Am Coll Cardiol 2003;42:728–35.
68. Shah KB, Kop WJ, Christenson RH, et al. Natriuretic peptides and echocardiography in acute dyspnoea: implication of elevated levels with normal systolic function. Eur J Heart Failure 2009;11:659–67.
69. Klemen P, Golub M, Grmec S. Combination of quantitative capnometry, N-terminal pro-brain natriuretic peptide, and clinical assessment in differentiating acute heart failure from pulmonary disease as cause of acute dyspnea in pre-hospital emergency setting: study of diagnostic accuracy. Croatian Med J 2009;50:133–42.
70. Behnes M, Brueckmann M, Ahmad-Nejad P, et al. Diagnostic performance and cost effectiveness of measurements of plasma N-terminal pro brain

natriuretic peptide in patients presenting with acute dyspnea or peripheral edema. Int J Cardiol 2009;135:165–74.
71. Altman DG. Practical statistics for medical research. London: Chapman and Hall, 1991.
72. Mueller-Lenke N, Rudez J, Staub D, et al. Use of chest radiography in the emergency diagnosis of acute congestive heart failure. Heart 2006;92:695–6.
73. Alibay Y, Beauchet A, El Mahmoud R, et al. Plasma N-terminal pro-brain natriuretic peptide and brain natriuretic peptide in assessment of acute dyspnea. Biomed Pharmacother 2005;59:20–4.
74. Boldanova T, Noveanu M, Breidthardt T, et al. Impact of history of heart failure on diagnostic and prognostic value of BNP: results from the B-type Natriuretic Peptide for Acute Shortness of Breath Evaluation (BASEL) study. Int J Cardiol 2010;142:265–72.
75. Chenevier-Gobeaux C, Claessens YE, Voyer S, Desmoulins D, Ekindjian OG. Influence of renal function on N-terminal pro-brain natriuretic peptide (NT-proBNP) in patients admitted for dyspnoea in the emergency department: comparison with brain natriuretic peptide (BNP). Clin Chim Acta 2005;361:167–75.
76. Coste J, Jourdain P, Pouchot J. A gray zone assigned to inconclusive results of quantitative diagnostic tests: application to the use of brain natriuretic peptide for diagnosis of heart failure in acute dyspneic patients. Clin Chem 2006;52:2229–35.
77. de Filippi CR, Seliger SL, Maynard S, Christenson RH. Impact of renal disease on natriuretic peptide testing for diagnosing decompensated heart failure and predicting mortality. Clin Chem 2007; 53:1511–9.
78. Di Somma S, De Berardinis B, Bongiovanni C, Marino R, Ferri E, Alfei B. Use of BNP and bioimpedance to drive therapy in heart failure patients. Congest Heart Failure 2010;16(Suppl 1): S56–61.
79. Eckstein J, Potocki M, Murray K, et al. Direct comparison of mid-regional pro-atrial natriuretic peptide with N-terminal pro B-type natriuretic peptide in the diagnosis of patients with atrial fibrillation and dyspnoea. Heart 2012;98:1518–22.
80. Gorissen C, Baumgarten R, de Groot M, van Haren E, Kragten H, Leers M. Analytical and clinical performance of three natriuretic peptide tests in the emergency room. Clin Chem Lab Med 2007;45:678–84.
81. Liteplo AS, Marill KA, Villen T, et al. Emergency thoracic ultrasound in the differentiation of the etiology of shortness of breath (ETUDES): sonographic B-lines and N-terminal pro-brain-type natriuretic peptide in diagnosing congestive heart failure. Acad Emerg Med 2009;16:201–10.
82. Mebazaa A, Vanpoucke G, Thomas G, et al. Unbiased plasma proteomics for novel diagnostic biomarkers in cardiovascular disease: identification of quiescin Q6 as a candidate biomarker of acutely decompensated heart failure. Eur Heart J 2012;33:2317–24.
83. Meisel SR, Januzzi JL, Medvedovski M, et al. Pre-admission NT-proBNP improves diagnostic yield and risk stratification - the NT-proBNP for EValuation of dyspnoeic patients in the Emergency Room and hospital (BNP4EVER) study. Eur Heart J Acute Cardiovasc Care 2012;1:99–108.
84. Ray P, Arthaud M, Lefort Y, et al. Usefulness of B-type natriuretic peptide in elderly patients with acute dyspnea. Intensive Care Med 2004;30:2230–6.
85. Ro R, Thode HC Jr, Taylor M, Gulla J, Tetrault E, Singer AJ. Comparison of the diagnostic characteristics of two B-type natriuretic peptide point-of-care devices. J Emerg Med 2011;41:661–7.
86. 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 noncardiac dyspnea. Am J Cardiol 2009;104:689–94.
87. Steinhart B, Thorpe KE, Bayoumi AM, Moe G, Januzzi JL Jr, Mazer CD. Improving the diagnosis of acute heart failure using a validated prediction model. J Am Coll Cardiol 2009;54:1515–21.
88. Zaninotto M, Mion M, Altinier S, et al. NT-proBNP in the differential diagnosis of acute dyspnea in the emergency department. Clin Biochem 2005;38:1041–4.
89. Clerico A, Zaninotto M, Prontera C, et al. State of the art of BNP and NT-proBNP immunoassays: the CardioOrmoCheck study. Clin Chim Acta 2012;414:112–9.
90. Cibinel GA, Casoli G, Elia F, et al. Diagnostic accuracy and reproducibility of pleural and lung ultrasound in discriminating cardiogenic causes of acute dyspnea in the emergency department. Intern Emerg Med 2012;7:65–70.
91. Ozkan B, Unluer EE, Akyol PY, et al. Stethoscope versus point-of-care ultrasound in the differential diagnosis of dyspnea: a randomized trial. Eur J Emerg Med 2015; 22:440–3.
92. Volpicelli G, Mussa A, Garofalo G, et al. Bedside lung ultrasound in the assessment of alveolar-interstitial syndrome. Am J Emerg Med 2006;24:689–96.
93. Baumgartner RN, Chumlea WC, Roche AF. Estimation of body composition from bioelectric impedance of body segments. Am J Clin Nutr 1989;50:221–6.
94. Alexander P, Alkhawam L, Curry J, et al. Lack of evidence for intravenous vasodilators in ED patients with acute heart failure: a systematic review. Am J Emerg Med 2015;33:133–41.
95. Heart Failure Society of America, Lindenfeld J, Albert NM, et al. HFSA 2010 Comprehensive Heart Failure Practice Guideline. J Card Fail 2010;16:e1–194.
96. Felker GM, Lee KL, Bull DA, et al. Diuretic strategies in patients with acute decompensated heart failure. N Engl J Med 2011;364:797–805.
97. Wakai A, McCabe A, Kidney R, et al. Nitrates for acute heart failure syndromes. Cochrane Database Syst Rev 2013; 8:CD005151.

98. Publication Committee for the VMAC Investigators. (Vasodilation in the Management of Acute CHF). Intravenous nesiritide vs nitroglycerin for treatment of decompensated congestive heart failure: a randomized controlled trial. JAMA 2002;287:1531–40.
99. Madias JE. Superiority of the limb leads over the precordial leads on the 12-lead ECG in monitoring fluctuating fluid overload in a patient with congestive heart failure. J Electrocardiol 2007;40:395–9.
100. Kataoka H, Madias JE. Changes in the amplitude of electrocardiogram QRS complexes during follow-up of heart failure patients. J Electrocardiol 2011;44(394):e1–9.
101. Lichtenstein D, Goldstein I, Mourgeon E, Cluzel P, Grenier P, Rouby JJ. Comparative diagnostic performances of auscultation, chest radiography, and lung ultrasonography in acute respiratory distress syndrome. Anesthesiology 2004;100:9–15.
102. Chakko S, Woska D, Martinez H, et al. Clinical, radiographic, and hemodynamic correlations in chronic congestive heart failure: conflicting results may lead to inappropriate care. Am J Med 1991;90:353–9.
103. McMurray JJ, Adamopoulos S, Anker SD, et al. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. Eur. J Heart Failure 2012;14:803–69.
104. Roberts E, Ludman AJ, Dworzynski K, et al. The diagnostic accuracy of the natriuretic peptides in heart failure: systematic review and diagnostic meta-analysis in the acute care setting. BMJ 2015;350:h910.
105. Carpenter CR, Keim SM, Worster A, Rosen P; BEEM (Best Evidence in Emergency Medicine). Brain natriuretic peptide in the evaluation of emergency department dyspnea: is there a role? J Emerg Med 2012;42:197–205.
106. Rutten JH, Steyerberg EW, Boomsma F, et al. N-terminal pro-brain natriuretic peptide testing in the emergency department: beneficial effects on hospitalization, costs, and outcome. Am Heart J 2008;156:71–7.
107. Schneider HG, Lam L, Lokuge A, et al. B-type natriuretic peptide testing, clinical outcomes, and health services use in emergency department patients with dyspnea: a randomized trial. Ann Intern Med 2009;150:365–71.
108. Trinquart L, Ray P, Riou B, Teixeira A. Natriuretic peptide testing in EDs for managing acute dyspnea: a meta-analysis. Am J Emerg Med 2011;29:757–67.
109. Hohl CM, Mitelman BY, Wyer P, Lang E. Should emergency physicians use B-type natriuretic peptide testing in patients with unexplained dyspnea? Can J Emerg Med 2003;5:162–5.
110. Lichtenstein DA, Meziere GA. Relevance of lung ultrasound in the diagnosis of acute respiratory failure: the BLUE protocol. Chest 2008;134:117–25.
111. Lichtenstein D, Meziere G. A lung ultrasound sign allowing bedside distinction between pulmonary edema and COPD: the comet-tail artifact. Intensive Care Med 1998;24:1331–4.
112. Gudmundsson P, Rydberg E, Winter R, Willenheimer R. Visually estimated left ventricular ejection fraction by echocardiography is closely correlated with formal quantitative methods. Int J Cardiol 2005;101:209–12.
113. Unluer EE, Karagoz A, Akoglu H, Bayata S. Visual estimation of bedside echocardiographic ejection fraction by emergency physicians. West J Emerg Med 2014;15:221–6.
114. Owan TE, Hodge DO, Herges RM, Jacobsen SJ, Roger VL, Redfield MM. Trends in prevalence and outcome of heart failure with preserved ejection fraction. N Engl J Med 2006;355:251–9.
115. Nagueh SF, Appleton CP, Gillebert TC, et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography. J Am Soc Echocardiogr 2009;22:107–33.
116. Ommen SR, Nishimura RA, Appleton CP, et al. Clinical utility of Doppler echocardiography and tissue Doppler imaging in the estimation of left ventricular filling pressures: a comparative simultaneous Doppler-catheterization study. Circulation 2000;102:1788–94.
117. Summers RL, Shoemaker WC, Peacock WF, Ander DS, Coleman TG. Bench to bedside: electrophysiologic and clinical principles of noninvasive hemodynamic monitoring using impedance cardiography. Acad Emerg Med 2003;10:669–80.
118. Tang WH, Tong W. Measuring impedance in congestive heart failure: current options and clinical applications. Am Heart J 2009;157:402–11.
119. Moher D, Pham B, Lawson ML, Klassen TP. The inclusion of reports of randomised trials published in languages other than English in systematic reviews. Health Technol Assess 2003;7:1–90.
120. Shaikh K, Ahmad M. Diagnostic significance of NT-proBNP estimation in patients with acute dyspnea. J Coll Physicians Surg Pak 2011;21:584–8.
121. Chenevier-Gobeaux C, Delerme S, Allo JC, et al. B-type natriuretic peptides for the diagnosis of congestive heart failure in dyspneic oldest-old patients. Clin Biochem 2008;41:1049–54.
122. Chenevier-Gobeaux C, Guerin S, Andre S, et al. Midregional pro-atrial natriuretic peptide for the diagnosis of cardiac-related dyspnea according to renal function in the emergency department: a comparison with B-type natriuretic peptide (BNP) and N-terminal proBNP. Clin Chem 2010;56:1708–17.
123. Takase H, Dohi Y. Kidney function crucially affects B-type natriuretic peptide (BNP), N-terminal proBNP and their relationship. Eur J Clin Invest 2014;44:303–8.

Supporting Information

The following supporting information is available in the online version of this paper:

Data Supplement S1. AHF systematic review data extraction form.
Data Supplement S2. Figures.
Data Supplement S3. Tables.
Data Supplement S4. Test treatment calculator.

Announcing Usus – A community website on usage

Usus (Latin for usage) is a new, independent, community-run website (http://www.usus.org.uk/) for all those interested in the usage of online content. It is designed to support a productive conversation among librarians, publishers, aggregators, and repository managers so that we can all get the best possible usage reports for our electronic resources.

The Usus website provides:

- A source of hints and tips on solving known problems
- A list of vendors with problems that are affecting the credibility and/or usefulness of the COUNTER reports
- A collection point for suggestions for new COUNTER usage reports and metrics

Supervisory Board

The Usus Supervisory Board will ensure that the website is editorially independent and will serve the needs of the community. Chaired by Anne Osterman, Deputy Director of VIVA (the Virtual Library of Virginia), the members of the Supervisory Board are:

Anne Osterman, VIVA, USA (Chair)
Simon Bevan, Cranfield University, UK
Melissa Blaney, ACS Publications, USA
Anna Creech, University of Richmond, USA
Lorraine Estelle, JISC, UK
Oliver Pesch, EBSCO, USA
Kari Schmidt, Montgomery College, USA
Mark Tullos, ProQuest, USA

Travel Award

Thank you to Project COUNTER for the financial support needed to get Usus off the ground. COUNTER has also offered to provide a travel award worth £1,000/$1,500 to a librarian who contributes the best opinion piece for the News & Opinions section of the Usus site. The award can be used to travel to the Charleston Conference, UKSG Conference, or Electronic Resources & Libraries Conference. Please send your submissions of 1,000 words or less to usus.stats@gmail.com by December 31, 2014.

For more information, please contact Usus at: usus.stats@gmail.com