
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

Andrew S. Liteplo, MD, RDMS, Keith A. Marill, MD, Tomas Villen, MD, Robert M. Miller, MD, Alice F. Murray, MBChB, Peter E. Croft, BS, Roberta Capp, MD, and Vicki E. Noble, MD, RDMS

Abstract

Objectives: Sonographic thoracic B-lines and N-terminal pro-brain-type natriuretic peptide (NT-ProBNP) have been shown to help differentiate between congestive heart failure (CHF) and chronic obstructive pulmonary disease (COPD). The authors hypothesized that ultrasound (US) could be used to predict CHF and that it would provide additional predictive information when combined with NT-ProBNP. They also sought to determine optimal two- and eight-zone scanning protocols when different thresholds for a positive scan were used.

Methods: This was a prospective, observational study of a convenience sample of adult patients presenting to the emergency department (ED) with shortness of breath. Each patient had an eight-zone thoracic US performed by one of five sonographers, and serum NT-ProBNP levels were measured. Chart review by two physicians blinded to the US results served as the criterion standard. The operating characteristics of two- and eight-zone thoracic US alone, compared to, and combined with NT-ProBNP test results for predicting CHF were calculated using both dichotomous and interval likelihood ratios (LRs).

Results: One-hundred patients were enrolled. Six were excluded because of incomplete data. Results of 94 patients were analyzed. A positive eight-zone US, defined as at least two positive zones on each side, had a positive likelihood ratio (LR+) of 3.88 (99% confidence interval [CI] = 1.55 to 9.73) and a negative likelihood ratio (LR-) of 0.5 (95% CI = 0.30 to 0.82), while the NT-ProBNP demonstrated a LR+ of 2.3 (95% CI = 1.41 to 3.76) and LR- of 0.24 (95% CI = 0.09 to 0.66). Using interval LR for the eight-zone US test alone, the LR for a totally positive test (all eight zones positive) was infinite and for a totally negative test (no zones positive) was 0.22 (95% CI = 0.06 to 0.80). For two-zone US, interval LR were 4.73 (95% CI = 2.10 to 10.63) when inferior lateral zones were positive bilaterally and 0.3 (95% CI = 0.13 to 0.71) when these were negative. These changed to 8.04 (95% CI = 1.76 to 37.33) and 0.11 (95% CI = 0.02 to 0.69), respectively, when congruent with NT-ProBNP.

Conclusions: Bedside thoracic US for B-lines can be a useful test for diagnosing CHF. Predictive accuracy is greatly improved when studies are totally positive or totally negative. A two-zone protocol performs similarly to an eight-zone protocol. Thoracic US can be used alone or can provide additional predictive power to NT-ProBNP in the immediate evaluation of dyspneic patients presenting to the ED.

ACADEMIC EMERGENCY MEDICINE 2009; 16:201-210 © 2009 by the Society for Academic Emergency Medicine

Keywords: ultrasound, emergency, dyspnea, congestive heart failure, NT-ProBNP, comet tails

From the Department of Emergency Medicine, Division of Emergency Ultrasound, Massachusetts General Hospital (ASL, TV, KAM, RMM, AFM, PEC, RC, VEN), Boston, MA.

Dr. Villen is currently with Servicio de Urgencias, Hospital Infanta Sofia, Madrid, Spain.

Dr. Miller is currently with the North Shore Medical Center, Salem, MA.

Dr. Murray is currently with The New Royal Infirmary of Edinburgh, Edinburgh, Scotland.

Dr. Croft is currently with the Ohio State University College of Medicine, Columbus, OH.

Received August 26, 2008; revision received November 6, 2008; accepted November 8, 2008.

Presented at the American Institute of Ultrasound in Medicine convention, San Diego, CA, March 15, 2008.

Address for correspondence and reprints: Andrew S. Liteplo, MD; e-mail: aliteplo@partners.org.

Dyspnea is a common presentation in the emergency department (ED). Emergency physicians (EPs) often need to make rapid diagnoses and treatment plans with limited clinical information. Acute decompensated congestive heart failure (CHF) is particularly challenging, as clinical, radiographic, and laboratory test parameters have variable diagnostic value.¹⁻¹⁰ In addition, many patients carry dual diagnoses of CHF and chronic obstructive pulmonary disease (COPD). Distinguishing between the two is often difficult,¹¹ and always essential, because management is different and incorrect treatment may be associated with detrimental cardiovascular effects.^{12,13} Therefore, rapid bedside tests that help to make this distinction are extremely useful to EPs and could also be helpful to out-of-hospital emergency care providers, intensive care specialists, and other health care providers as well.

Brain-type natriuretic peptide (BNP) and N-terminal pro-brain-type natriuretic peptide (NT-ProBNP) have recently emerged as useful laboratory biomarkers in determining the cause of a patient's dyspnea.¹⁴⁻¹⁷ However, there are limitations in that NT-ProBNP levels can be increased in many non-CHF disease processes, such as acute coronary syndromes, right heart strain or failure, pulmonary embolism, critical illness, renal failure, atrial fibrillation, and advanced age, and can be decreased in patients with a high body mass index.¹⁸⁻²³

Thoracic ultrasound (US) is emerging as an increasingly helpful tool for evaluating a whole spectrum of thoracic pathology. It is unique as it largely relies on the identification of sonographic artifacts for diagnosis. There are two very distinct sonographic imaging artifacts in the evaluation of extravascular lung water or pulmonary edema. A-lines suggest aerated or hyperaerated lungs and are seen in patients with normal lungs, asthma, or COPD; B-lines suggest thickened interstitia or fluid-filled alveoli and are seen most commonly in patients with CHF.²⁴ Identification of these artifacts has been helpful in differentiating between COPD and CHF. Lichtenstein et al.^{25,26} first described diffuse B-lines as an US sign of interstitial edema in 1997. Further work in the cardiology literature correlated US findings with wedge pressure and extravascular lung water.²⁷⁻²⁹ Volpicelli et al.³⁰ simplified the scanning technique and studied diffuse B-lines in emergency patients, finding a high sensitivity and specificity for alveolar interstitial syndrome. However, the broad definition of alveolar interstitial syndrome, and the fact that some USs were performed up to 48 hours after presentation, limited the applicability of their findings.

To our knowledge, no study has compared the diagnostic utility of thoracic US to NT-ProBNP levels. Our primary goals were to determine the optimal protocol and test threshold for the US test to diagnose CHF, to compare the diagnostic efficiency of US with NT-ProBNP levels in diagnosing CHF, and to determine if US adds incremental diagnostic information when combined with NT-ProBNP. The outcomes used to measure and compare efficacy were the positive, negative, and interval likelihood ratios (LRs).

METHODS

Study Design

In this prospective, blinded observational study we enrolled a convenience sample of patients presenting to the ED with undifferentiated dyspnea. The study was approved by the Human Research Committee Internal Review Board of Massachusetts General Hospital.

Study Setting and Population

The study was performed between December 2006 and June 2007 in the ED of an urban academic Level 1 trauma center and tertiary care facility with an annual ED census of 80,000 visits. The ED is a primary teaching site of an emergency medicine residency. EPs frequently use bedside US in their daily practice.

We included adult patients age 18 years and older who presented with shortness of breath and in whom an NT-ProBNP level was already being sent by treating physicians as part of the diagnostic workup. We chose this as an inclusion criterion because we felt that it is this group of patients for whom the diagnosis is not clinically clear, so that the potential utility of bedside thoracic US would be the most helpful. A convenience sample of patients was enrolled and only when an investigator was present (most frequently during daytime hours on weekdays). The investigator would screen for patients by asking treating physicians in the ED if any NT-ProBNP levels had been ordered as part of the diagnostic workup of a dyspneic patient. If a patient was identified, he or she was approached for written, informed consent. Exclusion criteria were inability to consent or not having an NT-ProBNP level sent. Patients with fever or respiratory distress were not excluded.

Study Protocol

Each patient had an eight-zone thoracic US performed and NT-ProBNP level checked. These were compared to a 'criterion standard' diagnosis of CHF, based on a consensus of chart review analysis. After informed written consent was obtained, baseline demographic and clinical data were gathered, from the chart, from the patient, or by asking the treating physicians. An US was performed by whichever study member was available at that time as soon as possible within the patient's ED visit. All study members were EPs with specialized training in US or departmental researchers who were medical students who had undergone didactic training (30 minutes), and hands-on instruction (2 hours) on thoracic US. Each student then demonstrated proficiency in performing and interpreting a minimum of five scans under direct attending supervision. There is a precedent for using this training protocol and literature support for its efficacy.³¹ Patients were enrolled, and an US scan was performed by a medical student, an attending EP, or a medical student who performed the US scan with direct attending EP oversight at the bedside. The study member interpreted scans in real time. Sonographers were blinded to the NT-ProBNP.

Ultrasound images were saved to a hard drive, and all available scans were later reviewed by a single

Registry of Diagnostic Medical Sonographers (RDMS)-credentialed study member blinded to clinical parameters and NT-ProBNP results, who ultimately determined positivity for each of the eight zones.

Patients' charts were reviewed at a later date, after the hospital course was completed. Each chart was independently reviewed by two EPs blinded to the US results. They determined if the patient's dyspnea on presentation was related to CHF or not (CHF+ or CHF-). Reviewers had available the entire electronic medical record, which includes all lab results (including NT-ProBNP), radiographic results, echocardiography results, admission notes, intensive care unit transfer notes, consultations, and discharge summaries, but not handwritten daily progress notes. In the cases of attending disagreement, a third EP independently reviewed the charts and served as the tie-breaker. There was no standard abstraction form, and the reviewers were not instructed on how to arrive at their conclusions. Reviewers put most emphasis on the discharge diagnosis and summary, which gave a clear etiology in the vast majority of cases. They referred to cardiology consults, echo results, chest x-ray, lab results, and other information as needed when the diagnosis was unclear. This expert physician-determined consensus of chart review analysis served as our criterion standard and is similar to the criterion standard used in other CHF research.^{14,16,17}

Measurement and US Protocol. NT-ProBNP levels were checked on all patients (F. Hoffmann-La Roche Ltd, Basel, Switzerland). Normal levels for this NT-ProBNP assay are age-dependent: age <50 years, 0–450 pg/mL; age 50–75 years, 0–900 pg/mL; and age >75 years, 0–1800 pg/mL.³² Each patient was categorized as having a normal or elevated level (NT-ProBNP- or NT-ProBNP+).

The two primary findings on thoracic US are A-lines and B-lines. A-lines were defined as hyperechoic horizontal lines parallel to the pleural line occurring at regular intervals below the pleural line. These are

artifacts caused by reverberation between the probe and pleura and are commonly found in normal or hyperaerated (emphysematous) lungs. B-lines (also called comet tail artifacts or lung comets), on the other hand, are hyperechoic reverberation artifacts that originate at the pleural line and extend radially from the probe to the edge of the screen, perpendicular to the pleural line. They move across the screen with pleural sliding and do not fade as they extend. B-lines are artifacts indicative of interstitial thickening. When interstitia and alveoli become edematous with fluid, as in CHF, B-lines become more prominent, numerous, and diffuse (Figure 1). Single B-lines can be a normal finding. Multiple B-lines have also been described in pneumonia, acute respiratory distress syndrome, and pulmonary fibrosis.³³

We followed a scanning protocol described by Volpicelli et al.³⁰ in which eight zones of the lungs are scanned (Figure 2). A 2–5 MHz 60-mm broadband curved array probe (Sonosite Titan or Micromaxx, Sonosite Inc., Bothell, WA) on abdominal settings was placed in sagittal and coronal orientations on the chest wall, perpendicular to the ribs, and the thoracic space was visualized to a depth of 18 cm. These are the scanning parameters that have been used in prior research in this field. Patients were scanned in their position of comfort.

The goal of the US was to identify the presence or absence of three or more B-lines in each of the eight zones. A zone was considered positive if multiple (at least three) B-lines were identified in an intercostal space. We defined each possible combination of number of positive zones on the right and number of positive zones on the left (right number–left number) by the maximum amount of zones present bilaterally, as summarized in Figure 3. A “4-B” scan is one in which all four zones are positive bilaterally (4–4). A “3-B” scan is one in which at least three zones are positive bilaterally (4–3, 3–3, 3–4). A “2-B” scan is when two or more zones are positive bilaterally (4–2, 3–2, 2–2, 2–3, 2–4). A “1-B” scan is when one or more zones are positive bilaterally

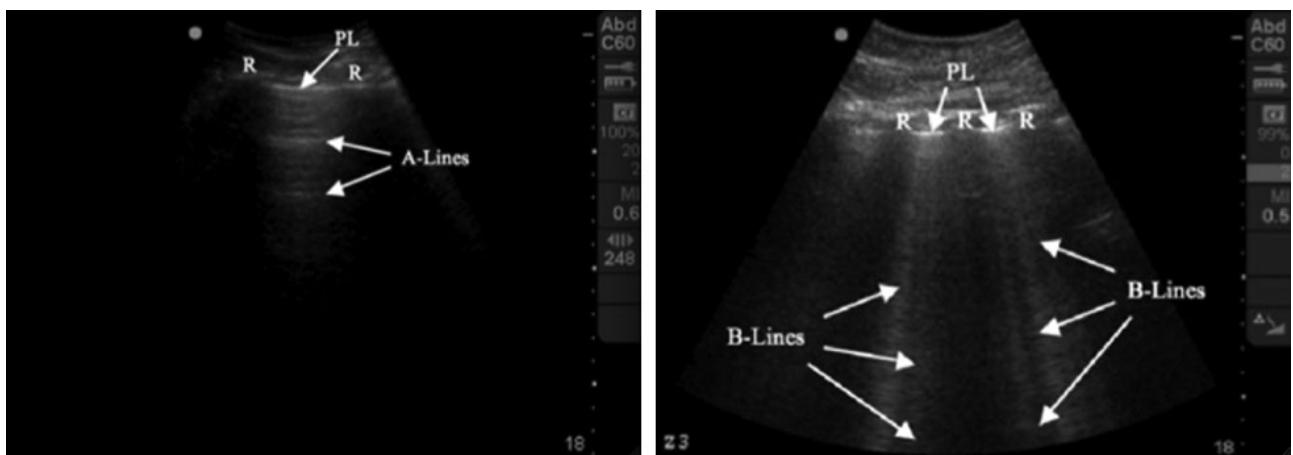


Figure 1. Ultrasound images demonstrating A-lines and B-lines. On the left, A-lines appear as hyperechoic lines between the ribs (R) and parallel to the pleural line (PL). On the right, B-lines appear as hyperechoic lines originating at the pleural line and extending radially to the bottom of the screen, perpendicular to the PL. Note that these are different from the anechoic rib shadow seen posteriorly to the ribs.

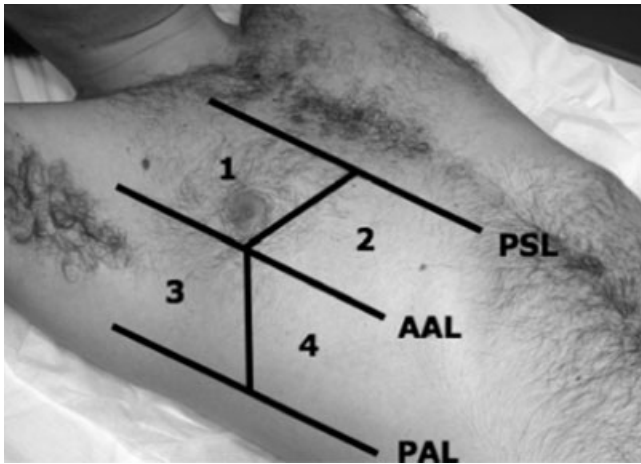


Figure 2. The right anterior chest wall divided into four zones. PSL is the parasternal line, AAL is the anterior axillary line, and PAL is the posterior axillary line. Zones 5 through 8 are the contralateral equivalents (also numbered medially to laterally).

		Left						
		0	1	2	3	4		
Right	4	4-0	4-1	4-2	4-3	4-4	4-B	Totally Positive
	3	3-0	3-1	3-2	3-3	3-4	3-B	
	2	2-0	2-1	2-2	2-3	2-4	2-B	Positive
	1	1-0	1-1	1-2	1-3	1-4	1-B	
	0	0-0	0-1	0-2	0-3	0-4	1-U	Anything Positive (does not include 0-B)
0-B		Totally Negative						

Figure 3. Numbers indicate the total number of positive zones on a side. Each combination of zones falls into a single definition, seen on the right and below. They are 4-B, 3-B, 2-B, 1-B, 1-U, and 0-B. Five thresholds for positivity are defined: totally positive, very positive, positive, minimally positive, and anything positive. Any scan that has the minimal number of positive zones or more is positive, and the others were considered negative.

(4-1, 3-1, 2-1, 1-1, 1-2, 1-3, 1-4). A “1-U” scan is when at least one zone is positive unilaterally, but the other side is negative (4-0, 3-0, 2-0, 1-0, 0-1, 0-2, 0-3, 0-4). Finally, a “0-B” scan is when there are no positive zones bilaterally (0-0).

In the protocol of Volpicelli et al.,³⁰ these classifications were grouped into only two dichotomous outcomes. A positive study was one in which at least two zones on each side were positive; 4-B, 3-B, and 2-B scans would all be considered positive. A negative study was one in which these conditions were not met; 1-B, 1-U, and 0-B are negative.

Data Analysis

All data, including the results of the scans, were input into a Microsoft Excel (Microsoft Corp., Redmond,

WA) spreadsheet. These were analyzed using Excel 2004 (Version 11.3.5) and EPIDAT (Version 3.1, Pan American Health Organization, Washington, DC). Cohen’s kappa with associated 95% confidence interval (CI) was used to assess interobserver agreement.

Ultrasound results were analyzed using a dichotomous positive or negative result or generalized to multiple test outcomes using interval LRs. The use of multiple interval LRs allows the clinician to gain maximal predictive value from a diagnostic test by, for example, addressing the difference between an US where all of the zones are positive or only some of the zones are positive.^{34,35} For dichotomous analysis, a minimum threshold of zones for positivity was set, and any scan that met these criteria or more was considered a positive test. We define these thresholds for positivity as such: totally positive includes only 4-B; very positive includes 4-B and 3-B; positive includes 4-B, 3-B, and 2-B (as defined by Volpicelli et al.); minimally positive includes 4-B, 3-B, 2-B, and 1-B; and anything positive includes 4-B, 3-B, 2-B, 1-B, and 1-U, which is all scans except ones that are totally negative (0-B). These thresholds are demonstrated in Figure 3. No adjustment was made for clustering on investigators performing the scans because the majority were students with similar scanning training and limited medical knowledge.

Using all eight US zones, the threshold for a positive test was varied, and a receiver operating characteristic (ROC) curve was constructed. To gain some sense for the dependence of the US test on the scanner, the ROC curve was also produced for those scans performed solely by medical students and for those performed by or in the presence of an ED attending. Next, multiple interval LR outcomes were calculated. A simplification of the US exam using a single zone on each side was explored with dichotomous outcomes and ROC curves. Interval LR outcomes were calculated for the best-performing two-zone US exam. The test characteristics of the dichotomous NT-ProBNP test were determined and compared to the US test using LRs and the ROC curve. Finally, the additional benefit of combining the eight- or two-zone US and the NT-ProBNP tests was assessed. Only dichotomous LRs were calculated when combining the eight-zone US and NT-ProBNP tests because interval LRs would have led to an unwieldy 12 different outcome categories.

A total of 28 sensitivity and specificity measurements and 45 LRs were calculated. An adjustment of the analysis for multiple testing was considered to address the increased chance of a Type I or false-positive statistically significant result. However, because the various tests were highly correlated, standardized adjustments such as the Bonferroni correction might be overly conservative and lead to increased Type II or false-negative errors. Assessing and adjusting for the degree of correlation would be difficult. Ultimately, no specific adjustment for multiple testing was made, but conservative 99% CIs are presented for all of the test data as recommended by Campbell et al.³⁶ to decrease the likelihood of a Type I error.

StatXact (Statxact 3, Cytel Software, Cambridge, MA) was used to compute exact 99% CIs of sensitivities and

specificities, and the log method was used to calculate 99% CIs for LRs.³⁷ The area under the curve (AUC) and associated nonparametric 99% CI for ROC curves were computed using SPSS 15 (SPSS Inc., Chicago, IL). No power calculation was performed prior to initiation of the study.

RESULTS

A total of 100 patients were enrolled. US studies were feasible 100% of the time and always took less than 5 minutes to perform. The study investigators always attempted to enroll patients as close to their time of arrival as possible. The distribution of US scanners and readers is depicted in Table 1. The median time after patient arrival to the ED to NT-ProBNP being drawn was 79 minutes (interquartile range [IQR] = 105 minutes) and to US was 96 minutes (IQR = 117 minutes). All US studies were performed in the first 12 hours after arrival, with 95% occurring in the first 6 hours. One patient did not have complete US interpretations available for review. Five patients did not have an NT-ProBNP level. Ninety-four patients had complete data sets that were available for review. Forty patients were CHF+ (34 NT-ProBNP+, 6 NT-ProBNP-), and 54 patients were CHF- (20 NT-ProBNP+, 34 NT-ProBNP-).

All available US images were reviewed by an RDMS-credentialed physician as described previously, and this overview result was compared to the interpretation documented by the study member who obtained the scans. The reviewer agreed with the initial zone assessment in 92.4% of the cases, and the overall positivity or negativity of the study in 97.7%. The Cohen kappa index for interobserver reliability was 0.82 (95% CI = 0.78 to 0.87). Eight patients did not have images available for review. We included them in the analysis as we felt that the sonographers' interpretations agreed well with the reviewers'. Overall, 64% of patients were enrolled by medical students only, 14% by EPs only, and for 22% of patients both were present. Of the two independent reviewers who served as criterion standards for CHF, the overall Cohen kappa index for interreviewer reliability was 0.87 (95%

Table 1
Distribution of Enrollment by US Scanner and Reader

US Scanner	US Reader	Total Number of Patients	Scans Done with Faculty Supervision
MS-1	Att-2	49	7
MS-1	MS-1	1	0
MS-2	Att-2	13	11
MS-3	Att-2	11	3
MS-3	MS-3	7	0
Att-1	Att-2	3	0
Att-2	Att-2	10	0
Total		94	21

All faculty supervision provided by Att-2.
Att = attending; MS = medical student; US = ultrasound.

Table 2
Patient Demographics

Demographic	Number (%)	
Average age, years (\pm SD)	74 (\pm 14) (range, 18–94)	
Male	55 (59)	
Medical history		
CHF	46 (49)	
Hypertension	37 (40)	
COPD	29 (31)	
Atrial fibrillation	23 (26)	
Diabetes	21 (22)	
Chronic renal insufficiency	17 (18)	
Asthma	6 (6)	
Diagnosis	ED (n = 94)	Hospital (n = 88)
Shortness of breath	26 (28)	2 (2)
CHF	20 (21)	25 (28)
Chest pain, NOS	11 (12)	11 (13)
Pneumonia	10 (11)	17 (19)
COPD	9 (10)	9 (10)
Atrial fibrillation	3 (3)	4 (5)
Asthma	1 (1)	0 (0)
ACS	0 (0)	6 (7)
Other	11 (12)	19 (22)
Admitted patients	88 (94)	

ACS = acute coronary syndrome (includes unstable angina, coronary artery disease, non-ST elevation myocardial infarction); CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; ED = emergency department; NOS = not otherwise specified.

CI = 0.76 to 0.97). A third reviewer was only necessary in 6 of 94 cases. Patient demographics are summarized in Table 2.

Positive US as a Predictor of CHF. In this study, we used the methods and definitions of a positive US as defined by Volpicelli et al.³⁰ Their definition of a positive scan requires two or more positive zones bilaterally of eight zones measured. The operating characteristics of positive US findings for the diagnosis of CHF based on the criterion standard consensus diagnosis are presented in Table 3.

Modified Thresholds of US Positivity. Physiologically, it is reasonable to postulate that more positive zones bilaterally would indicate a greater likelihood of CHF. The data were analyzed using dichotomous and interval LRs. Eight-zone USs were given one of six possible classifications based on the number of positive zones per side (Figure 3). The operating characteristics of dichotomous LRs were determined for the five resulting test thresholds and plotted on an ROC curve (Figure 4). This approach used a dichotomous outcome of positive or negative, as the threshold for positivity was varied from anything positive to totally positive (Table 3). To assess for the influence of the scanner, Figure 5 splits the ROC curve into patients scanned by medical students only and patients scanned with an attending present. Finally, an interval LR was calculated for each of the six possible individual outcomes (Table 4).

Table 3
Operating Characteristics of US and NT-ProBNP for CHF

Threshold for a Positive Test	Sensitivity	Specificity	LR+	LR-
Totally positive	0.23 (0.08, 0.43)	1.00 (0.91, 1.00)	Infinite	0.78 (0.62, 0.97)
Very positive	0.43 (0.23, 0.64)	0.96 (0.84, 0.99)	11.48 (1.78, 74.22)	0.60 (0.42, 0.86)
Positive	0.58 (0.36, 0.77)	0.85 (0.69, 0.95)	3.88 (1.55, 9.73)	0.50 (0.30, 0.82)
Minimally positive	0.78 (0.57, 0.92)	0.72 (0.54, 0.86)	2.79 (1.51, 5.15)	0.31 (0.14, 0.69)
Anything positive	0.90 (0.72, 0.98)	0.46 (0.29, 0.64)	1.68 (1.18, 2.40)	0.22 (0.06, 0.80)
NT-ProBNP	0.85 (0.65, 0.96)	0.63 (0.45, 0.79)	2.30 (1.41, 3.76)	0.24 (0.09, 0.66)
Positive and NT-ProBNP	0.53 (0.32, 0.73)	0.91 (0.76, 0.98)	5.67 (1.75, 18.35)	0.52 (0.33, 0.81)
Positive or NT-ProBNP	0.90 (0.72, 0.98)	0.57 (0.39, 0.74)	2.11 (1.37, 3.25)	0.17 (0.05, 0.61)
Minimally Positive and NT-ProBNP	0.68 (0.46, 0.85)	0.87 (0.71, 0.96)	5.21 (1.99, 13.61)	0.37 (0.20, 0.68)
Minimally Positive or NT-ProBNP	0.95 (0.79, 0.99)	0.48 (0.31, 0.66)	1.83 (1.28, 2.61)	0.10 (0.02, 0.62)
Zones 1 and 5 (anterior/superior)	0.40 (0.21, 0.61)	0.89 (0.74, 0.97)	3.64 (1.19, 11.16)	0.68 (0.48, 0.97)
Zones 2 and 6 (anterior/inferior)	0.40 (0.21, 0.61)	0.96 (0.84, 0.99)	10.80 (1.66, 70.28)	0.62 (0.44, 0.87)
Zones 3 and 7 (lateral/superior)	0.50 (0.29, 0.71)	0.94 (0.81, 0.99)	9.00 (1.98, 40.99)	0.53 (0.35, 0.81)
Zones 4 and 8 (lateral/inferior)	0.53 (0.32, 0.73)	0.89 (0.74, 0.97)	4.73 (1.62, 13.85)	0.53 (0.34, 0.83)

All data reported as *n* (99% CIs), *n* = 94 (40 CHF+, 54 CHF-). CHF = congestive heart failure; CI = confidence interval; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; NT-ProBNP = N-terminal pro brain-type natriuretic peptide; US = ultrasound.

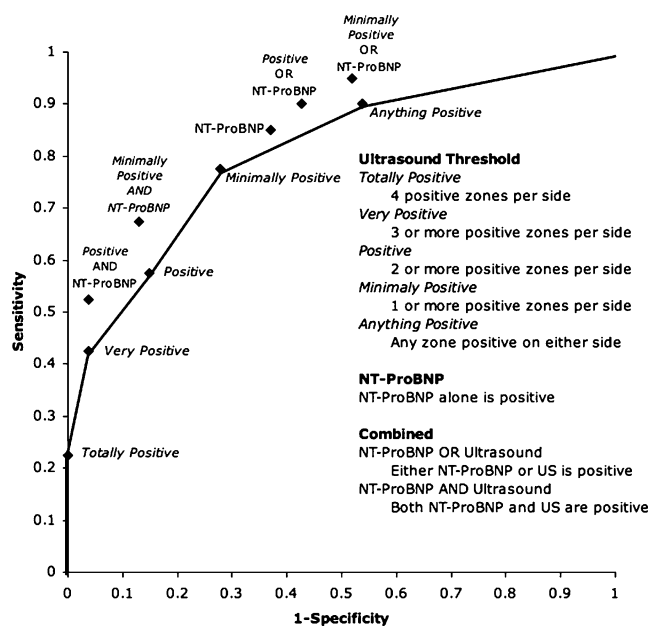


Figure 4. ROC curve of sonographic B-lines and NT-ProBNP in the detection of CHF. CHF = congestive heart failure; NT-Pro-BNP = N-terminal pro-brain-type natriuretic peptide; ROC = receiver operating characteristic.

The performance of individual pairs of symmetric bilateral zones as a test for CHF was also analyzed (Table 3). ROC curves were drawn for zone pairs 1 and 5, 2 and 6, 3 and 7, and 4 and 8, and the AUC was calculated for each pair. Although there was no statistically significant difference between zone pairs, Zones 4 and 8 (AUC = 0.78; 99% CI = 0.65 to 0.91) most closely approximated the eight-zone protocol performance (AUC = 0.81; 99% CI = 0.70 to 0.93). Interval LRs were calculated for the two-zone (Zones 4 and 8) US test (Table 4).

NT-ProBNP as a Predictor of CHF. The performance of NT-ProBNP as a test for CHF was determined

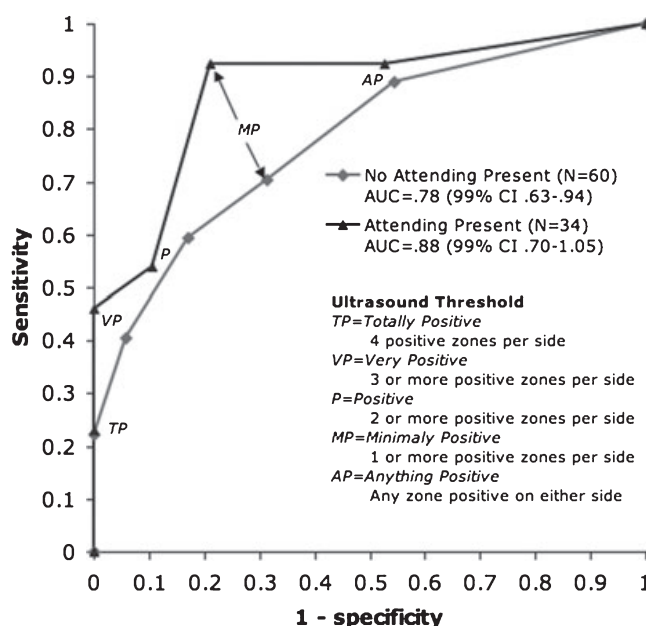


Figure 5. The ROC curve for ultrasound with dichotomous thresholds, analyzing the effect of the presence of an attending physician. When an attending scanned alone or with a medical student, there was a statistically insignificant trend toward better performance than when medical students scanned alone. ROC = receiver operating characteristic.

(Table 3). The performance of NT-ProBNP when combined with the eight-zone US test with a threshold of positive or minimally positive for positivity was also calculated. The two tests were combined in a fashion that required both to be positive (e.g., NT-ProBNP and US) or either to be positive (e.g., NT-ProBNP or US; Table 3). Finally, the NT-ProBNP test was combined with the two-zone (Zones 4 and 8) US test. To investigate the maximal predictive power when the two tests are congruent (both positive or negative), the LR interval approach was taken to yield six possible test outcomes and associated LRs (Table 4).

Table 4
Interval LRs

	Eight-zone US		Two-zone US (Zones 4 and 8 only)		Two-zone US (Zones 4 and 8 only) and NT-ProBNP Combined				
	<i>n</i>	LR	Zone positivity	<i>n</i>	LR	Zone positivity	NT-ProBNP	<i>n</i>	LR
4-B	9	Infinite	Both (4+ 8+)	27	4.73 (1.62, 13.85)	Both (4+ 8+)	+	21	8.04 (1.76, 37.33)
3-B	10	5.4 (0.74, 39.21)	Either (4+ 8- or 4- 8+)	23	1.24 (0.48, 3.17)	Either (4+ 8- or 4- 8+)	+	16	2.25 (0.66, 7.68)
2-B	12	1.35 (0.33, 5.47)				Either (4+ 8- or 4- 8+)	+	7	0.23 (0.01, 3.62)
1-B	15	1.54 (0.45, 5.28)				Neither (4- 8-)	+	17	0.74 (0.22, 2.46)
1-U	19	0.48 (0.14, 1.66)	Neither (4- 8-)	44	0.3 (0.13, 0.71)	Neither (4- 8-)	+	17	0.74 (0.22, 2.46)
0-B	29	0.22 (0.06, 0.80)				Neither (4- 8-)	-	27	0.11 (0.02, 0.69)

99% CIs are in parentheses, *n* = 94.
CI = confidence interval; LR = likelihood ratio; NT-ProBNP = N-terminal pro brain-type natriuretic peptide; US = ultrasound.

DISCUSSION

The benefits of US are many. It is a rapid, immediate, reproducible, noninvasive, and nonirradiating tool that is increasingly available to many EPs. It can be performed at the bedside and repeated frequently and is inexpensive and portable. Finally, thoracic US can be used in settings where other techniques are not readily available, such as in remote locations, at high altitudes,³⁸ in developing parts of the world with limited radiographic capabilities, or even in the out-of-hospital setting (e.g., ambulances and helicopters).

The thoracic US technique itself is easy to learn. Medical student researchers with limited clinical and US experience were able to perform and interpret the exam with excellent correlation with an RDMS-credentialed US fellowship-trained EP. In fact, 64% of patients were enrolled by a medical student only. The study was always feasible, and patients were scanned in their position of comfort; they did not have to sit erect or lie supine for the study to be performed.

Prior research on thoracic US for pulmonary edema has primarily focused on patients in intensive care units. In a study of dyspneic patients presenting to the ED, USs were done within 24 hours of presentation and not upon arrival to the ED. This study adds to the current body of knowledge about thoracic US in the following ways: it studies the operating characteristics of B-lines when the US is performed in the ED, it compares performance of thoracic US to NT-ProBNP, and it analyzes the eight-zone scanning protocol and compares its performance to a simplified two-zone protocol.

This study was designed to investigate the use of US alone, or in combination with serum NT-ProBNP, to diagnose CHF in ED patients. When used alone, thoracic US was investigated with an eight-zone technique and a dichotomous positive or negative outcome, or six incremental outcomes. A simplified two-zone technique, which would have the advantage of being faster to perform, was also investigated in an analogous manner. Thoracic US alone using an eight-zone technique and dichotomous outcome demonstrated a high LR+ and low LR- depending on the threshold chosen for a positive test, but the test did not demonstrate both a high LR+ and a low LR- simultaneously at any single diag-

nostic test threshold (Table 3). Using the US test with six interval LR outcomes, 38 of 94 patients (40%) had a congruent US exam with all eight zones positive (4-B) or negative (0-B) and an infinite (specificity of 1.00) or 0.22 LR, respectively (Table 4). This suggests that in a significant proportion of patients, bedside US results can strongly increase or decrease the likelihood of CHF.

Except for Zones 2 and 6, which tended to perform worse, the two-zone US test demonstrated comparable test characteristics to the full eight-zone study (Table 3). The lateral zone pairs 3 and 7 and 4 and 8 tended to demonstrate higher sensitivity. These results suggest that for a two-zone test, lateral probe placement may be preferred, but in general, precise probe placement does not appear to be critical. The AUC was not significantly different for any of the two-zone tests, but Zones 4 and 8 demonstrated the highest AUC in our sample and thus were used for further two-zone analysis (Table 4).

The comparability of the eight-zone and two-zone US tests suggests that a faster and easier two-zone test may suffice for evaluation of CHF. This is because the different zones tended to be positive or negative together in the same patients. Thus, the two-zone tests demonstrated high, but not perfect, collinearity. The eight-zone test, however, has strong predictive abilities in ruling in CHF when all eight zones are positive or significantly decreasing the likelihood when all are negative (as was the case in 40% of patients), as seen when using the six-outcome interval LR approach (Table 4).

The serum NT-ProBNP test alone was more sensitive but less specific than a positive threshold of an eight-zone US. Combinations were tested with both the positive and the minimally positive thresholds and the requirement that one or both of the NT-ProBNP and US tests be positive for a positive test result. In none of the four combinations tested were the LR+ and LR- simultaneously improved when the NT-ProBNP test was combined with the dichotomous eight-zone US test (Table 3). However, in 18 cases, a US (using the positive threshold) was a better predictor of CHF than the NT-ProBNP, in 16 patients without CHF the NT-ProBNP was positive but the US negative, and in 2 patients with CHF the NT-ProBNP was negative and the US positive.

Combining the NT-ProBNP test with the two-zone US test and using six interval LR outcomes, 48 of 94 patients (51%) had a congruent outcome where the NT-ProBNP and both US zones 4 and 8 were positive or negative. The LRs for these tests were 8.04 and 0.11, respectively (Table 4). These LRs are not significantly more extreme than the positive and negative LRs for the NT-ProBNP test alone. In approximately half of the patients in this sample, the two-zone US test could be used in combination with the serum NT-ProBNP to strongly suggest or refute the diagnosis of CHF. In the other half of patients, the NT-ProBNP and US tests would not be completely congruent and would provide little predictive information (Table 4).

Five patients who were enrolled ultimately did not have NT-ProBNP levels performed. These five patients were thus excluded from the analysis. All of these patients had totally negative (0-B) USs with no positive zones and none had a diagnosis consistent with CHF. Inclusion of these patients would have improved the US test characteristics. It is possible that enrolling all dyspneic patients presenting to the ED, including those that are clinically obviously having a COPD or CHF exacerbation, would have further improved the operating characteristics. We specifically chose not to include these patients, as we felt that US would be most helpful in patients in whom there is diagnostic uncertainty. This may also explain why our sensitivities and specificities for US as a diagnostic tool for CHF are not as high as in other published studies.

Ultrasound could be used alone to assess for CHF in dyspneic ED patients. It performs similarly (overlapping CIs) to NT-ProBNP in that the likelihood of CHF is increased when the test is positive and decreased when negative, but has the advantage of being noninvasive and immediately available. It would be especially useful in settings such as developing countries, austere environments, or out-of-hospital care where laboratory or radiographic tests are unavailable. The current results are promising and suggest there are advantages and disadvantages to using either an eight- or a two-zone protocol and using dichotomous or interval LR outcomes. The overall predictive power of the eight- and two-zone protocols is similar, and the two-zone test is faster. However, using the eight-zone test, when all zones are either positive or negative, the likelihood of CHF is greatly increased or decreased, respectively. This immediate, significant information that is available at the bedside could have a direct and positive impact on patient care.

Ultrasound could also be used in combination with NT-ProBNP to assess for CHF in the ED. One major advantage of this approach is the immediate availability of US, which can be performed at the bedside within the first moment of a patient's arrival to guide initial management before the radiographic or blood results are available. Once the NT-ProBNP test result is available, the results of the two tests can be used in combination. Combining multiple interval LR outcomes seems advantageous. In particular, if the NT-ProBNP and US results are congruent, then this will significantly alter the likelihood of CHF. The data suggest that when a dichotomous US outcome and NT-ProBNP are com-

bined, there is a modest increase in predictive abilities. However, using the two-zone US in combination with NT-ProBNP and multiple interval LR outcomes, it is suggested, but not proven, that half of the patients can be segregated into groups at very high or low risk for CHF.

LIMITATIONS

Although USs were performed as close to patient arrival as possible, it is conceivable that US results may have been altered by treatment (e.g. diuretics) either early in their ED course or in the out-of-hospital setting, decreasing the sensitivity of the test. The issue of the lack of a true criterion standard remains a difficulty in CHF studies. We chose to have two attending physicians (and three in the cases of discrepancy) review the charts as a criterion standard. This is a limitation, as analysis is subjective. However, we feel that the high kappa value between the reviewers indicates that subjective differences were minimal.

The results of NT-ProBNP in comparison to the criterion standard composite analysis involve some inherent bias as well. Reviewers had access to these data points and were able to use a positive or negative result in their ultimate determination of the etiology of shortness of breath. This means that NT-ProBNP status may not be completely independent from CHF diagnosis status and hence the test characteristics of NT-ProBNP may be positively biased in this study. Even if the chart reviewers were blinded to the NT-ProBNP values, these values would have likely affected the treating physicians' judgments and decisions and indirectly biased the chart reviewers. It seems difficult to remove this bias from a composite clinical diagnosis of CHF.

The US test did not suffer from similar bias because the US readers were blinded to other patient data and the chart reviewers were blinded to the US test results. Nevertheless, the US examiners were not explicitly blinded to other patient data at the bedside. This bias is likely minimized by the fact that the majority of patients were enrolled by medical student researchers, who would have less clinical acumen with which to influence their scans or interpretations. Despite this potential bias, the presence or absence of B-lines is an objective finding that is not generally subject to scanner bias. No interobserver reliability of US scanning was assessed, as the patients were each scanned only once.

The units of analysis in this study were the 94 patients. The medical student scanners could also be the units of analysis, and this would require adjustment for clustering on the various individuals. This approach was not formally taken because the student background and training were limited, and their technique was likely to be stereotyped and homogenous. The effect of attending physician presence during the scan was assessed; improvement in the diagnostic ability of US was not noted.

There was only a single US-trained EP reader. As a derivational study, this clearly represents an optimal scenario for EP expertise in reading USs in the ED. The generalizability of this skill in reading pulmonary USs,

and thus the study results to other EPs without advanced training, remains to be seen.

Even for patients with congruent NT-ProBNP and US test results, the LR+ and LR- were not statistically significantly different from the analogous LRs for NT-ProBNP alone or the two-zone US test alone. Our preliminary work suggests that the US test can be predictive of CHF. However, this must be confirmed in other settings and more research done with larger patient samples to examine whether the trends are significant.

Our study population was dyspneic patients presenting to the ED in whom an NT-ProBNP level was already being sent for diagnostic purposes. Generalizability of our findings may be limited, as our population is likely to have a higher proportion of patients with a diagnosis of CHF than a population of all dyspneic patients presenting to the ED.

CONCLUSIONS

Bedside thoracic US for B-lines can be used to predict CHF when a predefined threshold is used. Interval LRs are powerfully predictive when all or no zones are positive. A two-zone protocol performs similarly to a full eight-zone protocol. Thoracic US can be used alone or in conjunction with NT-ProBNP in immediate evaluation of dyspneic patients presenting to the ED. The data suggest that congruent NT-ProBNP and US results may alter the odds of CHF, compared to the NT-ProBNP test alone. Further studies are needed to evaluate US's utility, both alone and in combination with NT-ProBNP testing, to diagnose CHF in the ED and to evaluate the potential for practicing EPs to perform and interpret this test.

References

1. Wright SP, Doughty RN, Pearl A, et al. Plasma amino-terminal pro-brain natriuretic peptide and accuracy of heart-failure diagnosis in primary care: a randomized, controlled trial. *J Am Coll Cardiol.* 2003; 42:1793-800.
2. Zehtabchi S, Brandler ES. Evidence-based emergency medicine/rational clinical examination abstract. Does this patient have congestive heart failure? *Ann Emerg Med.* 2008; 51:87-90.
3. Collins S, Storrow AB, Kirk JD, Pang PS, Diercks DB, Gheorghide M. Beyond pulmonary edema: diagnostic, risk stratification, and treatment challenges of acute heart failure management in the emergency department. *Ann Emerg Med.* 2008; 51:45-57.
4. Stevenson LW, Perloff JK. The limited reliability of physical signs for estimating hemodynamics in chronic heart failure. *JAMA.* 1989; 261:884-8.
5. Marcus GM, Gerber IL, McKeown BH, et al. Association between phonocardiographic third and fourth heart sounds and objective measures of left ventricular function. *JAMA.* 2005; 293:2238-44.
6. Lok CE, Morgan CD, Ranganathan N. The accuracy and interobserver agreement in detecting the 'gallop sounds' by cardiac auscultation. *Chest.* 1998; 114:1283-8.
7. Fonseca C, Morais H, Mota T, et al. The diagnosis of heart failure in primary care: value of symptoms and signs. *Eur J Heart Fail.* 2004; 800:821-2.
8. Fonseca C, Mota T, Morais H, et al. The value of the electrocardiogram and chest x-ray for confirming or refuting a suspected diagnosis of heart failure in the community. *Eur J Heart Fail.* 2004; 12:821-2.
9. 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.
10. Collins SP, Lindsell CJ, Storrow AB, Abraham WT, ADHERE Scientific Advisory Committee, Investigators and Study Group. Prevalence of negative chest radiography results in the emergency department patient with decompensated heart failure. *Ann Emerg Med.* 2006; 47:13-8.
11. Collins SP, Lindsell CJ, Peacock WF, Eckert DC, Askew J, Storrow AB. Clinical characteristics of emergency department heart failure patients initially diagnosed as non-heart failure. *BMC Emerg Med.* 2006; 6:11.
12. Salpeter SR, Ormiston TM, Salpeter EE. Cardiovascular effects of beta-agonists in patients with asthma and COPD: a meta-analysis. *Chest.* 2004; 125:2309-21.
13. Singer AJ, Emerman C, Char DM, et al. Bronchodilator therapy in acute decompensated heart failure patients without a history of chronic obstructive pulmonary disease. *Ann Emerg Med.* 2008; 51: 25-34.
14. Tung RH, Camargo CA Jr, Krauser D, et al. Amino-terminal pro-brain natriuretic peptide for the diagnosis of acute heart failure in patients with previous obstructive airway disease. *Ann Emerg Med.* 2006; 48:66-74.
15. McCullough PA, Hollander JE, Nowak RM, et al. Uncovering heart failure in patients with a history of pulmonary disease: rationale for the early use of B-type natriuretic peptide in the emergency department. *Acad Emerg Med.* 2003; 10:198-204.
16. 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.
17. 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.
18. Schwam E. B-type natriuretic peptide for diagnosis of heart failure in emergency department patients: a critical appraisal. *Acad Emerg Med.* 2004; 11: 686-91.
19. Morello A, Lloyd-Jones DM, Chae CU, et al. Association of atrial fibrillation and amino-terminal pro-brain natriuretic peptide concentrations in dyspneic subjects with and without acute heart failure: results from the ProBNP Investigation of Dyspnea in the Emergency Department (PRIDE) study. *Am Heart J.* 2007; 153:90-7.
20. Luchner A, Hengstenberg C, Lowel H, Riegger GA, Schunkert H, Holmer S. Effect of compensated

- renal dysfunction on approved heart failure markers: direct comparison of brain natriuretic peptide (BNP) and N-terminal pro-BNP. *Hypertension*. 2005; 46:118–23.
21. Chenevier-Gobeaux C, Claessens YE, Voyer S, Desmoulin 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.
 22. Bayes-Genis A, Lloyd-Jones DM, van Kimmenade RR, et al. Effect of body mass index on diagnostic and prognostic usefulness of amino-terminal pro-brain natriuretic peptide in patients with acute dyspnea. *Arch Intern Med*. 2007; 167:400–7.
 23. Krauser DG, Lloyd-Jones DM, Chae CU, et al. Effect of body mass index on natriuretic peptide levels in patients with acute congestive heart failure: a ProBNP investigation of dyspnea in the emergency department (PRIDE) substudy. *Am Heart J*. 2005; 149:744–50.
 24. Lichtenstein D. *General Ultrasound in the Critically Ill*. New York, NY: Springer, 2004.
 25. Lichtenstein D, Meziere G, Biderman P, Gepner A, Barre O. The comet-tail artifact. An ultrasound sign of alveolar-interstitial syndrome. *Am J Respir Crit Care Med*. 1997; 156:1640–6.
 26. 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.
 27. Picano E, Frassi F, Agricola E, Gligorova S, Gargani L, Mottola G. Ultrasound lung comets: a clinically useful sign of extravascular lung water. *J Am Soc Echocardiogr*. 2006; 19:356–63.
 28. Jambrik Z, Monti S, Coppola V, et al. Usefulness of ultrasound lung comets as a nonradiologic sign of extravascular lung water. *Am J Cardiol*. 2004; 93:1265–70.
 29. Agricola E, Bove T, Oppizzi M, et al. “Ultrasound comet-tail images”: a marker of pulmonary edema: a comparative study with wedge pressure and extravascular lung water. *Chest*. 2005; 127:1690–5.
 30. 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.
 31. Bedetti G, Gargani L, Corbisiero A, Frassi F, Poggianti E, Mottola G. Evaluation of ultrasound lung comets by hand-held echocardiography. *Cardiovasc Ultrasound*. 2006; 4:34.
 32. Januzzi JL, van Kimmenade R, Lainchbury J, Bayes-Genis A, Ordonez-Llanos J, Santalo-Bel M, Pinto YM, Richards M. NAT-ProBNP testing for diagnosis and short-term prognosis in acute destabilized heart failure: an international pooled analysis of 1256 patients: the International Collaborative of NT-ProBNP Study. *Eur Heart J*. 2006; 27:330–7.
 33. Volpicelli G, Caramello V, Cardinale L, Mussa A, Bar F, Frascisco MF. Detection of sonographic B-lines in patients with normal lung or radiographic alveolar consolidation. *Med Sci Monit*. 2008; 14:CR122–8.
 34. Dent B, Kendall RJ, Boyle AA, Atkinson PR. Emergency ultrasound of the abdominal aorta by UK emergency physicians: a prospective cohort study. *Emerg Med J*. 2007; 24:547–9.
 35. Brown MD, Reeves MJ. Interval likelihood ratios: another advantage for the evidence-based diagnostician. *Ann Emerg Med*. 2003; 42:292–7.
 36. Gallagher EJ. Evidence-based emergency medicine editorial. The problem with sensitivity and specificity. *Ann Emerg Med*. 2003; 42:298–303.
 37. Campbell MJ, Daly LE, Machin D. Special topics. In: Altman DG, Machin D, Bryant TN, Gardner MJ, eds. *Statistics with Confidence*. 2nd ed. London: BMJ Books; 2000, pp 163–7.
 38. Altman DG. Diagnostic tests. In: Altman DG, Machin D, Bryant TN, Gardner MJ, eds. *Statistics with Confidence*. 2nd ed. London: BMJ Books, 2000, pp 105–19.
 39. Fagenholz PJ, Gutman JA, Murray AF, Noble VE, Thomas SH, Harris NS. Chest ultrasonography for the diagnosis and monitoring of high-altitude pulmonary edema. *Chest*. 2007; 131:1013–8.