ORIGINAL CONTRIBUTION

# Assessment of Dyspnea Early in Acute Heart Failure: Patient Characteristics and Response Differences Between Likert and Visual Analog Scales

Peter S. Pang, MD, Sean P. Collins, MD, Kori Sauser, MD, Adin-Cristian Andrei, PhD, Alan B. Storrow, MD, Judd E. Hollander, MD, Miguel Tavares, MD, Jindrich Spinar, MD, Cezar Macarie, MD, Dimitar Raev, MD, Richard Nowak, MD, Mihai Gheorghiade, MD, and Alexandre Mebazaa, MD

## Abstract

*Background:* Dyspnea is the most common symptom in acute heart failure (AHF), yet how to best measure it has not been well defined. Prior studies demonstrate differences in dyspnea improvement across various measurement scales, yet these studies typically enroll patients well after the emergency department (ED) phase of management.

*Objectives:* The aim of this study was to determine predictors of early dyspnea improvement for three different, commonly used dyspnea scales (i.e., five-point absolute Likert scale, 10-cm visual analog scale [VAS], or seven-point relative Likert scale).

*Methods:* This was a post hoc analysis of URGENT Dyspnea, an observational study of 776 patients in 17 countries enrolled within 1 hour of first physician encounter. Inclusion criteria were broad to reflect real-world clinical practice. Prior literature informed the a priori definition of clinically significant dyspnea improvement. Resampling-based multivariable models were created to determine patient characteristics significantly associated with dyspnea improvement.

*Results:* Of the 524 AHF patients, approximately 40% of patients did not report substantial dyspnea improvement within the first 6 hours. Baseline characteristics were similar between those who did or did not improve, although there were differences in history of heart failure, coronary artery disease, and initial systolic blood pressure. For those who did improve, patient characteristics differed across all three scales, with the exception of baseline dyspnea severity for the VAS and five-point Likert scale (c-index ranged from 0.708 to 0.831 for each scale).

*Conclusions:* Predictors of early dyspnea improvement differ from scale to scale, with the exception of baseline dyspnea. Attempts to use one scale to capture the entirety of the dyspnea symptom may be insufficient.

ACADEMIC EMERGENCY MEDICINE 2014;21:659–666 © 2014 by the Society for Academic Emergency Medicine

Signs and symptoms of heart failure prompt patients to seek medical care, with dyspnea as the most common symptom.[1,2] Yet a large proportion of patients discharged from the hospital remain symptomatic or have little weight loss. Severity of dyspnea at baseline, as well as failure to adequately relieve it during hospitalization, have been associated with worse outcomes.[3,4] Thus, improving dyspnea is important to both patients and physicians and is also an important target in acute heart failure (AHF) clinical trials.[5]

Over 5 years ago, a consensus for standardization of dyspnea assessments was proposed, yet assessment scales used today remains largely unchanged.[5] Typical assessments include a visual analog scale (VAS), a Likert scale, or both.[6] Past studies demonstrate that the VAS better captures changes in dyspnea over time compared to Likert-type scales and that patient positioning during measurement (sitting upright vs. supine) affects responses.[5,7,9] Timing of dyspnea assessment is also important, as earlier treatment is associated with greater dyspnea relief, although many patients remain symptomatic despite early therapeutic intervention.[3,4,8–10] Early dyspnea relief has also been associated with improved 30-day outcomes.[11] While the reliability of VAS and Likert scales is implied by their ease, repeated use, and ability to track changes over time, their reliability has not been studied rigorously.[5,6,12] Previous studies show differences in patient responses between Likert and VAS, suggesting poor interscale reliability.[7,9,13] Yet some of these studies have enrolled patients well after the initial emergency department (ED) phase of management, when previous work has demonstrated significant dyspnea improvement occurs early.[9]

Given these differences, we set out to determine predictors of early dyspnea improvement, according to a prespecified criterion of improvement for three different, commonly used dyspnea scales. Our hypothesis was that these predictors, as defined by baseline patient characteristics, would differ by type of scale. If they converge, then any measurement scale would suffice. If they diverge, this would suggest that each scale measures dyspnea differently, or measures different aspects of the symptom, and a unifying scale or multiple scales to accurately assess dyspnea are needed. Thus, we conducted a secondary analysis of the URGENT Dyspnea (Ularitide Global Evaluation in Acute Decompensated Heart Failure) study[9] to compare differences in proportions of responders and patient characteristics between three different measurement scales to identify early (6 hours after enrollment) dyspnea improvement.

## METHODS

### Study Design
The URGENT Dyspnea study design has been previously described.[9,14] Briefly, 776 patients from 35 institutions in 17 countries over a 7-month period (January 2007 through August 2007) were enrolled in an observational cohort study. The main study was designed to better understand the association of early standard therapy with dyspnea relief. Institutional review board or ethics committee approval was obtained from every site prior to participation in the URGENT Dyspnea study.

### Study Setting and Population
Patients aged 18 years or older with dyspnea secondary to a presumptive diagnosis of AHF were enrolled within

1 hour of first physician evaluation. Inclusion criteria were intentionally kept broad to better reflect "real-world" conditions, with the understanding that a proportion of dyspneic patients enrolled within 1 hour of presentation may not have AHF. Patients unable to consent or unable to self-assess dyspnea were excluded. Patients were treated with standard AHF management at the discretion of their health care teams; demographic, clinical, and treatment data were collected. The study protocol did not provide specific AHF treatment recommendations. The individual site investigator determined the final diagnosis of AHF based on full access to all available clinical data.

### Study Protocol

This study was designed to determine predictors of early dyspnea improvement ("early" defined as baseline to 6 hours), for three distinct dyspnea scales. Thus, three separate analyses were performed, one for each scale; only patients with dyspnea improvement per their respective scale were included in each analysis. Early dyspnea improvement was defined a priori (see Table 1), as follows: 1) any one point improvement on a five-point Likert scale,[15] 2) a three-point improvement on a 10-cm VAS (based on prior research demonstrating a change of 2.1 cm on a 100-mm VAS as clinically significant),[15] and 3) moderate or marked improvement in dyspnea on a seven-point Likert scale (no baseline scores for this scale).[13] Patients with minimal dyspnea at baseline who could not meet the criteria for improvement were excluded. Baseline dyspnea was defined as the self-reported patient score for both the five-point Likert scale and the VAS, as there was no baseline dyspnea assessment for the seven-point Likert scale because it only reports change in symptoms. Severity of dyspnea was assessed by the patient in the upright position, defined as head of bed at least 60 degrees from horizontal.

### Data Analysis

Multivariable models were created to identify baseline factors associated with dyspnea improvement. To overcome the well-known shortcomings of many automatic variable selection methods (i.e., instability, sensitivity to outliers, or lack of reproducibility), we have used a simplified version of a highly cited and widely employed resampling-based machine learning ensemble method

called *bagging*.[16] This model-averaging method avoids model overfitting, improves model stability and accuracy, and provides better protection against outliers and/or extreme observations. Therefore, bagging provides a principled and robust multivariable model-building strategy for dyspnea improvement.

Three separate models were created, one for each method of dyspnea assessment. Comparisons were not made between models, as this would imply attempts to find which model had the best fit to the same data. Previous analyses have described the effect of early therapy on dyspnea relief.[9] Rather, three separate models were created using the technique below for each model. Each model was created from the unique data set derived using the simplified bagging methodology.

For each dyspnea improvement definition, a two-stage process was created: first, the original data set was resampled (with replacement) 50,000 times. Within each such sample, stepwise variable selection with 0.05/0.1 entry/exit criteria occurred in the logistic regression model for dyspnea improvement. In a voting-like fashion, the list of factors (predictors) significant at two-sided 5% level across the 50,000 resamples was created. In stage two, the short list of predictors that were significant in at least 10,000 of the 50,000 resamples was compiled. Only these predictors were then included in the final logistic regression model, which was created based solely on the original data. Results are shown in Table 2. A full list of covariates considered for inclusion is provided in Data Supplement S1 (available as supporting information in the online version of this paper); elements including demographics, presenting signs and symptoms, diagnostic data, and medical history were considered. We were unable to include site of enrollment due to small numbers of patients from some sites, inducing instability in the models. The full list of enrollment by site is included in Data Supplement S2 (available as supporting information in the online version of this paper). Specifically in regard to troponin testing, as multiple countries using multiple troponin I assays were included, we used a single arbitrary cut point of 0.04 across all patients. Model performance was quantified using the area under the receiver's operating characteristics curve (c-index). Model appropriateness based on raw and standardized Pearson and deviance residuals indicated no discernible influential data points, multicollinearity prob-

Table 1
Dyspnea Assessment Instruments and Definitions of Improvement

| Type of Scale | Scale Intervals | Definition of Clinically Significant Improvement |
|---|---|---|
| Five-point Likert scale | 1 = not short of breath, 2 = mildly short of breath, 3 = moderately short of breath, 4 = severely short of breath, 5 = very severely short of breath. | ≥1-point change |
| 10-cm VAS | 1-cm increments, anchored by "I am not breathless at all" to "I am the most breathless I have ever been." | ≥3-cm change |
| Seven-point Likert scale | Patients were asked to compare how they felt when they were first asked regarding dyspnea. Scores ranged from "Markedly worse, moderately worse, mildly worse, no change, mildly better, moderately better, markedly better." | Only patients who responded moderately or markedly improved |

VAS = visual analog scale.

Table 2
Patient Characteristics Stratified by Five-point Likert Scale Change

| Variable | Not Improved (n = 142) | Improved (n = 299) | p-value |
|---|---|---|---|
| Age (yr), mean (±SD) | 66.1 ± 15.5 | 68.8 ± 14.0 | 0.07 |
| Sex, female | 63 (44) | 134 (45) | 0.93 |
| Race | | | 0.012 |
| Black | 34 (24) | 36 (12) | |
| White | 107 (75) | 258 (86) | |
| Heart failure history | | | 0.002 |
| Chronic heart failure | 102 (73) | 166 (57) | |
| New onset heart failure | 38 (27) | 125 (43) | |
| Most recent known LVEF, median (IQR) (n = 227) | 38.0 (10–75) | 40.0 (10–80) | 0.33 |
| Past medical history | | | |
| CAD | 74 (52) | 112 (37) | 0.004 |
| Prior MI | 37 (26) | 72 (24) | 0.65 |
| Primary cardiomyopathy | 28 (20) | 30 (10) | 0.005 |
| Obesity | 37 (26) | 75 (25) | 0.83 |
| Peripheral vascular disease | 9 (6) | 30 (10) | 0.20 |
| Asthma/COPD | 29 (20) | 46 (15) | 0.19 |
| Diabetes | | | |
| Insulin-dependent | 27 (19) | 43 (14) | 0.21 |
| Non–insulin-dependent | 27 (19) | 61 (20) | 0.73 |
| Renal insufficiency | 38 (27) | 78 (26) | 0.88 |
| Anemia (Hb < 12 g/dL) | 28 (20) | 36 (12) | 0.032 |
| Hypertension | 103 (73) | 227 (76) | 0.44 |
| Prehospital treatment | | | 0.015 |
| Yes | 69 (49) | 178 (60) | |
| No | 63 (44) | 91 (30) | |
| Unknown | 10 (7) | 30 (10) | |
| Signs and symptoms prior to triage | | | |
| Onset of dyspnea before triage | | | 0.11 |
| 1 to 3 days | 26 (18) | 45 (15) | |
| 3 to 5 days | 13 (9) | 28 (9) | |
| Past 24 hours | 41 (29) | 123 (41) | |
| > 5 days | 61 (43) | 99 (33) | |
| Weight gain | 49 (35) | 91 (30) | 0.39 |
| Fatigue | 102 (72) | 167 (56) | 0.001 |
| No preceding signs | 17 (12) | 57 (19) | 0.06 |
| Initial vital signs, mean (±SD) | | | |
| Systolic BP (mm Hg) | 137.3 ± 33.7 | 146.7 ± 33.9 | 0.007 |
| Diastolic BP (mm Hg) | 78.7 ± 19.6 | 82.4 ± 19.1 | 0.06 |
| Heart rate (beats/min) | 89.4 ± 22.4 | 92.4 ± 24.6 | 0.21 |
| Respiratory rate (breaths/min) | 21.5 ± 7.0 | 23.0 ± 7.0 | 0.041 |
| Oxygen saturation | 94.3 ± 5.3 | 92.3 ± 5.9 | 0.002 |
| Physical examination findings | | | |
| Abnormal heart sounds (e.g., murmur, gallop) | 48 (34) | 71 (24) | 0.026 |
| Jugular venous distention | 70 (49) | 117 (39) | 0.044 |
| Rales | 90 (63) | 204 (68) | 0.31 |
| Wheezing | 22 (15) | 38 (13) | 0.43 |
| Peripheral edema | 85 (60) | 147 (49) | 0.036 |
| Other signs: none of the above | 10 (7) | 36 (12) | 0.11 |

Table 2
Continued

| Variable | Not Improved (n = 142) | Improved (n = 299) | p-value |
|---|---|---|---|
| Laboratory results, mean (±SD) | | | |
| Serum sodium (mmol/L) | 137.1 ± 12.5 | 135.9 ± 16.5 | 0.44 |
| Serum potassium | 4.0 ± 0.9 | 3.9 ± 0.9 | 0.16 |
| Creatinine | 1.5 ± 1.4 | 1.5 ± 1.1 | 0.58 |
| Urea | 43.9 ± 51.2 | 27.2 ± 35.5 | <0.001 |
| BNP, median (IQR) (n = 98) | 866.5 (40–5,019) | 938.5 (61–4,357) | 0.65 |
| NTpro BNP, median (IQR) (n = 89) | 7,281 (902–34,019) | 6,302.5 (395–30,000) | 0.27 |
| Troponin > 0.04 ng/mL (n = 171) | 42 (30) | 129 (43) | 0.006 |
| CXR and ECG Results | | | |
| CXR | | | |
| Normal | 4 (3) | 5 (2) | 0.43 |
| Interstitial edema | 77 (54) | 122 (41) | 0.008 |
| Pulmonary edema | 18 (13) | 60 (20) | 0.06 |
| First ECG | | | |
| Normal | 10 (7) | 21 (7) | 0.99 |
| Sinus rhythm | 76 (54) | 174 (58) | 0.35 |
| Baseline medications | | | |
| Beta-blockers | 88 (62) | 143 (48) | 0.005 |
| ACE inhibitors | 74 (52) | 139 (46) | 0.27 |
| Angiotension II receptor blockers | 16 (11) | 38 (13) | 0.67 |
| Aldosterone antagonists | 26 (18) | 26 (9) | 0.003 |
| Diuretics | 105 (74) | 183 (61) | 0.009 |

Data are reported as n (%) unless otherwise noted.
ACE = angiotensin-converting enzyme; BNP = brain natriuretic peptide; BP = blood pressure; CAD = coronary artery disease; COPD = congestive obstructive pulmonary disease; ECG = electrocardiogram; Hb = hemoglobin; LVEF = left ventricular ejection fraction; MI = myocardial infarction; NTproBNP = N-terminal pro-brain natriuretic peptide.

lems, or nonlinear covariate functional forms. Statistical significance was established at the two-sided 5% alpha level and no adjustments for multiplicity were made.

## RESULTS

Of the 776 patients enrolled within 1 hour of first physician assessment, 524 had AHF at the 6-hour evaluation, 79 did not have AHF, and in 173 the diagnosis was uncertain. Differences in baseline characteristics between those with AHF, those without AHF, and those with indeterminate diagnoses have been previously reported.[9] Figure 1 describes the patient flow for each of the separate dyspnea assessments, reporting those patients able to improve per our prespecified criteria.

Overall, the majority of patients reported improvement on at least one scale by 6 hours. However, a sizable number of patients did not improve. Specifically, per the five-point Likert scale, 34.1% of patients (132 of 387) did not have a one-point improvement by 6 hours; by the seven-point Likert scale, 38.7% of patients (184 of 476) did not have moderate or marked improvement; and by the VAS, 57.2% of patients (178 of 311) did not have a three-point improvement.
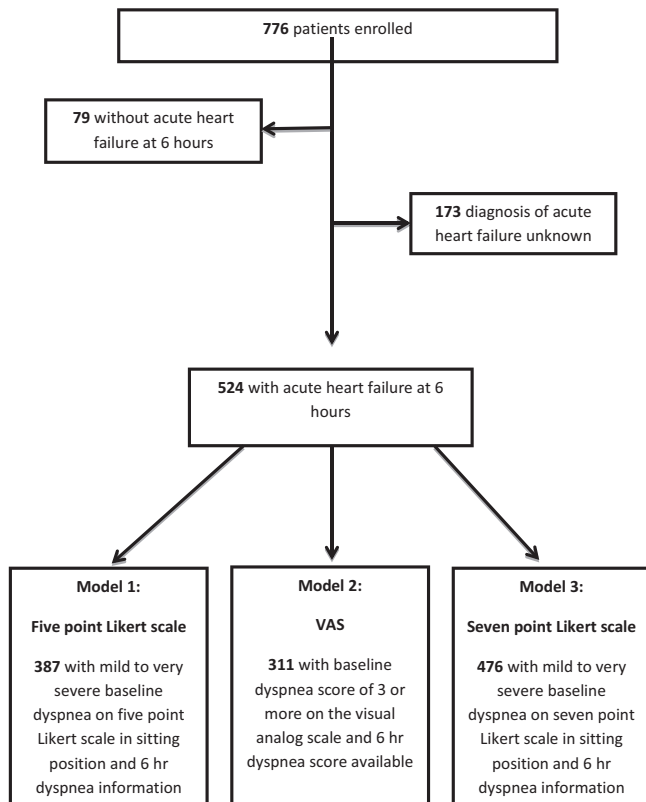
**Figure 1.** Study flow diagram of the patient groups obtained using the hour 6 dyspnea symptoms improvement definitions. AHF = acute heart failure; Dx = diagnosis.

Baseline characteristics are seen in Table 2. The five-point Likert scale was chosen for baseline characteristics in Table 2 as it had the best discrimination by c-statistic. Patients with dyspnea improvement were similar to patients who did not improve with the following exceptions: improved patients were more likely to have new-onset HF, higher systolic blood pressures, and troponin I levels > 0.04 ng/mL. They were also less likely to be chronically treated with beta-blocker therapy, diuretics, and aldosterone antagonists. Table 3 shows the treatments patients received within the first 6 hours of management.

Three separate multivariable models identifying patient characteristics associated with dyspnea improvement are shown in Tables 4A–4C, representing each method of dyspnea assessment. The discrimination ability of each model was relatively high (c-index = 0.708 to 0.831). Severity of dyspnea at baseline was common across all three models as a predictor of improvement; the less severe the dyspnea, the less likely patients would report improvement. No other clinical, laboratory, or treatment factors were common to all three models. Overall, there were more differences in patient characteristics associated with dyspnea improvement per each model of dyspnea assessment than there were similarities.

## DISCUSSION

In this observational study of AHF patients enrolled within 1 hour of first physician contact, most patients report dyspnea improvement within 6 hours of initial management, irrespective of the assessment scale used. However, a sizable proportion of patients did not report improvement according to our predetermined definitions of clinical significance. For those who did improve, substantial variation in patient characteristics associated with dyspnea improvement between measurement scales were observed. These differences support previous studies suggesting different scales measure dyspnea differently.[7,9] No patient characteristic was associated with greater or lesser dyspnea improvement in all three of our models (Figure 2). Severity of dyspnea at baseline was associated with improvement for both the five-point Likert and the VAS, although only for those patients with the most severe dyspnea as measured by the Likert scale. Of note, the seven-point Likert scale did not have a baseline dyspnea assessment.

Table 3
Initial Treatment Within the First 6 Hours

| Variable | Not Improved ($n$ = 142) | Improved ($n$ = 299) | p-value |
|---|---|---|---|
| CPAP at 6 hours | | | 0.13 |
|   Yes | 2 (1) | 12 (4) | |
|   No | 134 (99) | 262 (96) | |
| Loop diuretic | 132 (93) | 243 (81) | 0.001 |
| Nitroglycerin | 29 (20) | 86 (29) | 0.06 |
| Dobutamine | 6 (4) | 10 (3) | 0.64 |
| Dopamine | 5 (4) | 7 (2) | 0.48 |

Data are reported as number (%). Groups are stratified based on 5 point Likert Scale.
CPAP = continuous positive airway pressure.

Table 4A
Patient Characteristics Associated With Dyspnea Improvement by Five-point Likert Scale

| Variables | AOR | (95% CI) | p-value |
|---|---|---|---|
| Baseline dyspnea | | | |
|   Very severely SOB | 33.85 | (9.43–121.52) | <0.001 |
|   Severely SOB | 7.61 | (3.74–15.49) | 0.45 |
|   Moderately SOB | 5.79 | (3.07–10.89) | 0.77 |
|   Mildly SOB | REF | REF | REF |
| Oxygen saturation | 0.97 | (0.92–1.03) | 0.29 |
| Preceding sign: fatigue | 0.55 | (0.30–0.99) | 0.045 |
| Jugular venous distention | 0.48 | (0.28–0.83) | 0.009 |
| New onset of heart failure | 0.99 | (0.55–1.77) | 0.97 |
| Coronary artery disease | 0.77 | (0.45–1.33) | 0.35 |
| Chronic therapy: aspirin | 0.60 | (0.35–1.04) | 0.07 |
| Chest X-ray: pulmonary edema | 1.82 | (0.88–3.75) | 0.11 |
| First ECG | | | |
|   Paced rhythm | 0.36 | (0.13–0.96) | 0.041 |
|   Other abnormal rhythm | 0.47 | (0.21–1.08) | 0.076 |
| IV loop diuretic use | 0.41 | (0.14–1.18) | 0.10 |
| Transdermal nitroglycerin in the past 6 hours | 0.35 | (0.14–1.18) | 0.048 |

$N$ = 387; c-index = 0.831.
AOR = adjusted odds ratio; ECG = electrocardiogram; SOB = short of breath.

Table 4B
Patient Characteristics Associated with Dyspnea Improvement by VAS

| Model Scale | Variables | AOR | 95% CI | p-value |
|---|---|---|---|---|
| VAS | Baseline dyspnea* | 1.55 | (1.36–1.77) | <0.001 |
| 7-point Likert | Jugular venous distention | 0.46 | (0.28–0.77) | 0.003 |

*N* = 311; c-index = 0.756
*For each one-point change by the VAS. Patients who could not improve three points at baseline were excluded.
AOR = adjusted odds ratio; VAS = visual analog scale.

Table 4C
Patient Characteristics Associated With Dyspnea Improvement by Seven-point Likert Scale

| Variables | AOR | (95% CI) | p-value |
|---|---|---|---|
| Dyspnea onset | | | |
| Past 24 hours | 2.00 | (1.12–3.56) | 0.31 |
| 1 to 3 days | 1.45 | (0.74–2.83) | 0.66 |
| 3 to 5 days | 2.35 | (1.41–3.94) | 0.044 |
| >5 days | REF | REF | REF |
| Peripheral edema | 0.76 | (0.49–1.19) | 0.23 |
| New onset of heart failure | 1.75 | (1.12-.75) | 0.014 |
| Primary cardiomyopathy | 0.51 | (0.29–0.89) | 0.019 |
| Peripheral vascular disease | 3.40 | (1.40–8.27) | 0.007 |
| Hypertension | 1.51 | (0.95–2.39) | 0.08 |
| Chest radiograph: normal | 0.37 | (0.10–1.40) | 0.14 |
| First ECG: other abnormal rhythm | 0.38 | (0.18–0.78) | 0.009 |

*N* = 476; c-index = 0.708.
AOR = adjusted odds ratio; ECG = electrocardiogram.



*only for patients with 'very severe' shortness of breath at baseline per 5-point Likert scale
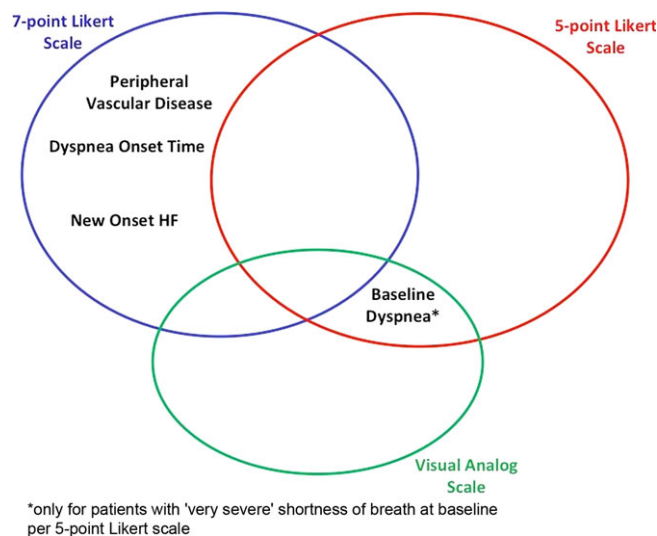
**Figure 2.** Patient characteristics associated with dyspnea improvement for each assessment model. HF = heart failure.

Dyspnea, or the sense of breathlessness, discomfits patients, driving them to seek emergent care. Although most patients improve during hospitalization, dyspnea remains an unmet need for a sizable proportion of patients despite early treatment. This lack of early benefit has also been observed in a large clinical trial, ASCEND-HF (Acute Study of Clinical Effectiveness of Nesiritide in Decompensated Heart Failure), which used a seven-point Likert scale.[11] However, ASCEND-HF enrolled patients up to 24 hours after presentation, in contrast to our study, which enrolled patients within 1 hour of first physician evaluation. While other studies using *physician* assessment suggests nearly all patients improve,[17] this study indicates that *patient*-assessed dyspnea is not sufficiently relieved in many patients, consistent with clinical trials where patients are enrolled well after the ED phase of management, yet still have dyspnea.[8,18] Future trials targeting dyspnea may need to enroll patients earlier to capture the majority of those who experience early improvement.[8,13] Patients who do not improve significantly with early management may represent a refractory or "hard-to-treat" group.

For each scale, we found unique clinical factors associated with dyspnea improvement, suggesting that individual scales may capture different aspects of the patient's reported symptoms. It was not our intent to determine whether one scale is better than another, but to better understand the similarities or differences in regards to mapping patient characteristics. While investigators prefer a single, "best" scale, our study suggests that use of multiple current scales may be required, as a single scale may not accurately capture all AHF patients. Based on our data, investigators might consider use of a single scale if targeting a certain subgroup of AHF with the patient characteristics most associated with improvement or lack of improvement to test whether interventions lead to changes in dyspnea severity. More importantly, our findings suggest a comprehensive, validated patient-reported outcome of dyspnea improvement remains a critical unmet need in AHF research.

Despite the prominence of dyspnea, trialists debate its significance as a target.[19] Perhaps such a relentless focus on dyspnea has hindered other opportunities to improve outcomes. Acute trials have, in general, not defined or addressed other items important to patients, such as functional capacity as primary endpoints. Patients present with symptoms, however, and continue to have significant symptoms despite early therapy. For some HF patients, symptom improvement would be a welcome trade for life-years. The less severe the symptoms, the less willing they are to trade survival time.[20] Furthermore, failure to adequately relieve early symptoms in clinical trials has been associated with worse 30-day outcomes.[4,11] However, these trials enrolled patients relatively late in comparison to this analysis. Further, these results assume that current dyspnea assessments are correct. Although we defined "clinically significant" dyspnea for our study, the minimal clinical important difference has not been well studied, with a study of 74 patients being the largest to date.[15] It may be the scale or use of the scale that is the issue, not the symptom.

Further study of dyspnea and how to measure it is warranted, given its prominence as a presenting symptom and its association with worse outcomes. Standardization of assessment remains important as divergent or

different results based on type of dyspnea measurement scale used raises questions about how to interpret study results.

## LIMITATIONS

This was a post hoc analysis, limited by the inability to account for unmeasured confounders. Thus, our findings are hypothesis-generating only. The inclusion and exclusion criteria were broad, raising the possibility that non-AHF patients were included. However, diagnostic accuracy in the ED setting is relatively high, and natriuretic peptide levels provide additional support.[21] Compared to other AHF clinical trials, the number of patients with AHF is relatively small. However, those trials enrolled patients much later than this study. Although the dyspnea scales used in this study have been commonly used in AHF clinical trials, they have not been rigorously validated according to established standards for a patient-reported outcome. While the scales do meet the face validity criterion, they have largely been adapted from other settings, raising the possibility that imprecision may have confounded our results. The definition of improvement significantly affects the results, with different thresholds likely to yield differing associated patient characteristics. However, we used the definitions of dyspnea improvement most commonly reported by other AHF trials and studies. The protocol did not mandate a specific order or randomize the order of the three dyspnea assessment scales. While that would have been ideal, practically this would have required significantly more resources, as this process would have been repeated several times. While the majority of patients received intravenous diuretic therapy, the lack of detailed doses for both diuretics and vasodilators raises the possibility that undertreatment may have an untoward effect. While diuretics and nitrates do improve dyspnea in our analysis, the fact that dyspnea improved in patients who did not receive diuretics suggests potential confounding by intention. As this was not a randomized trial design originally, physicians may have chosen certain treatments for reasons not captured by our study design.

## CONCLUSIONS

Patients seek acute medical care when they feel short of breath. Most acute heart failure patients report early relief from dyspnea after initial therapy, but a sizable number do not. Greater baseline dyspnea severity predicts early dyspnea improvement in the two scales where baseline dyspnea was assessed. Variations in patient characteristics associated with dyspnea improvement were observed, depending on the type of measurement instrument used. This, highlights the need to better understand the symptom most responsible for AHF hospital presentation, and to standardize its assessment.

## References

1. Yancy CW, Lopatin M, Stevenson LW, De Marco T, Fonarow GC. Clinical presentation, management, and in-hospital outcomes of patients admitted with acute decompensated heart failure with preserved systolic function: a report from the Acute Decompensated Heart Failure National Registry (ADHERE) database. J Am Coll Cardiol 2006;47:76–84.
2. Abraham WT, Adams KF, Fonarow GC, et al. In-hospital mortality in patients with acute decompensated heart failure requiring intravenous vasoactive medications: an analysis from the Acute Decompensated Heart Failure National Registry (ADHERE). J Am Coll Cardiol 2005;46:57–64.
3. Metra M, Cleland JG, Weatherley BD, et al. Dyspnoea in patients with acute heart failure: an analysis of its clinical course, determinants, and relationship to 60-day outcomes in the PROTECT pilot study. Eur J Heart Fail 2010;12:499–507.
4. Metra M, O'Connor CM, Davison BA, et al. Early dyspnoea relief in acute heart failure: prevalence, association with mortality, and effect of rolofylline in the PROTECT Study. Eur Heart J 2011;32:1519–34.
5. Pang PS, Cleland JG, Teerlink JR, et al. A proposal to standardize dyspnoea measurement in clinical trials of acute heart failure syndromes: the need for a uniform approach. Eur Heart J 2008;29:816–24.
6. West RL, Hernandez AF, O'Connor CM, Starling RC, Califf RM. A review of dyspnea in acute heart failure syndromes. Am Heart J 2010;160:209–14.
7. Allen LA, Metra M, Milo-Cotter O, et al. Improvements in signs and symptoms during hospitalization for acute heart failure follow different patterns and depend on the measurement scales used: an international, prospective registry to evaluate the evolution of Measures of Disease Severity in Acute Heart Failure (MEASURE-AHF). J Card Fail 2008;14:777–84.
8. O'Connor CM, Starling RC, Hernandez AF, et al. Effect of nesiritide in patients with acute decompensated heart failure. N Engl J Med 2011;365:32–43.
9. Mebazaa A, Pang PS, Tavares M, et al. The impact of early standard therapy on dyspnoea in patients with acute heart failure: the URGENT-dyspnoea study. Eur Heart J 2010;31:832–41.
10. Pang PS, Konstam MA, Krasa HB, et al. Effects of tolvaptan on dyspnoea relief from the EVEREST trials. Eur Heart J 2009;30:2233–40.
11. Mentz RJ, Hernandez AF, Stebbins A, et al. Predictors of early dyspnoea relief in acute heart failure and the association with 30-day outcomes: findings from ASCEND-HF. Eur J Heart Fail 2013;15:456–64.
12. Johnson MJ, Oxberry SG, Cleland JG, Clark AL. Measurement of breathlessness in clinical trials in patients with chronic heart failure: the need for a standardized approach: a systematic review. Eur J Heart Fail 2010;12:137–47.
13. Teerlink JR, Cotter G, Davison BA, et al. Serelaxin, recombinant human relaxin-2, for treatment of acute heart failure (RELAX-AHF): a randomised, placebo-controlled trial. Lancet 2013;381:29–39.
14. Pang PS, Tavares M, Collins SP, et al. Design and rationale of the URGENT Dyspnea study: an international, multicenter, prospective study. Am J Ther 2008;15:299–303.

15. Ander DS, Aisiku IP, Ratcliff JJ, Todd KH, Gotsch K. Measuring the dyspnea of decompensated heart failure with a visual analog scale: how much improvement is meaningful? Congest Heart Fail 2004;10:188–91.

16. Brieman L. Bagging predictors. Machine Learn 1996;24:123–40.

17. Pang PS, Gheorghiade M, Dihu J, et al. Effects of tolvaptan on physician-assessed symptoms and signs in patients hospitalized with acute heart failure syndromes: analysis from the Efficacy of Vasopressin Antagonism in Heart Failure Outcome Study with Tolvaptan (EVEREST) Trials. Am Heart J 2011;161:1067–72.

18. Gheorghiade M, Konstam MA, Burnett JC Jr, et al. Short-term clinical effects of tolvaptan, an oral vasopressin antagonist, in patients hospitalized for heart failure: the EVEREST clinical status trials. JAMA 2007;297:1332–43.

19. Gheorghiade M, Ruschitzka F. Beyond dyspnoea as an endpoint in acute heart failure trials. Eur Heart J 2011;32:1442–5.

20. Stevenson LW, Hellkamp AS, Leier CV, et al. Changing preferences for survival after hospitalization with advanced heart failure. J Am Coll Cardiol 2008;52:1702–8.

21. 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.

## Supporting Information

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

**Data Supplement S1.** Baseline covariates collected.
**Data Supplement S2.** Participating countries.