ORIGINAL RESEARCH

Clinical Implications of the New York Heart Association Classification

César Caraballo, MD; Nihar R. Desai, MD, MPH; Hillary Mulder, MS; Brooke Alhanti, PhD; F. Perry Wilson, MD, MS; Mona Fiuzat, PharmD; G. Michael Felker, MD; Ileana L. Piña, MD, MPH; Christopher M. O'Connor, MD; Joanne Lindenfeld, MD; James L. Januzzi, MD; Lawrence S. Cohen, MD; Tariq Ahmad, MD, MPH

Background—The New York Heart Association (NYHA) classification has served as a fundamental tool for risk stratification of heart failure (HF) and determines clinical trial eligibility and candidacy for drugs and devices. However, its ability to adequately stratify risk is unclear.

Methods and Results—To compare NYHA class with objective assessments and survival in patients with HF, we performed secondary analyses of 4 multicenter National Institutes of Health–funded HF clinical trials that included patients classified as NYHA class II or III: TOPCAT (Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist), DIG (The Effect of Digoxin on Mortality and Morbidity in Patients With Heart Failure), HF-ACTION (Efficacy and Safety of Exercise Training in Patients With Chronic Heart Failure), and GUIDE-IT (Guiding Evidence-Based Therapy Using Biomarker Intensified Treatment in Heart Failure). Twenty-month cumulative survival was compared between classes using Kaplan–Meier curves and the log rank test. NT-proBNP (N-terminal pro–B-type natriuretic peptide), Kansas City Cardiomyopathy Questionnaire (KCCQ) scores, 6-minute walk distances, left ventricular ejection fraction, and cardiopulmonary test parameters were compared using Wilcoxon rank sum tests and percentage overlap using kernel density estimations. Cumulative mortality varied significantly across NYHA classes and HF clinical trials (likelihood ratio, $P<0.001$). Mortality at 20 months for NYHA class II ranged from 7% for patients in HF-ACTION to 15% in GUIDE-IT, whereas mortality for NYHA class III ranged from 12% in TOPCAT to 26% in GUIDE-IT. There was substantial percentage overlap in values for NT-proBNP levels (79% and 69%), KCCQ scores (63% and 54%), 6-minute walk distances (63% and 54%), and left ventricular ejection fraction (88% and 83%). Similarly, there was substantial overall in values for minute ventilation–carbon dioxide production relationship (71%), maximal oxygen uptake (54%), and exercise duration (53%).

Conclusions—The NYHA system poorly discriminates HF patients across the spectrum of functional impairment. These findings raise important questions about the need for improved phenotyping of these patients to facilitate risk stratification and response to interventions. **(*J Am Heart Assoc*. 2019;8:e014240. DOI: 10.1161/JAHA.119.014240.)**

Key Words: clinical trials • heart disease • heart failure • NYHA class

A simple functional classification of heart failure (HF) patients first suggested by the New York Heart Association (NYHA) has been used clinically for almost a century.[1] It has long served as a foundational tool for risk stratification of HF and determines clinical trial eligibility and candidacy for drugs and devices. Whereas it is widely acknowledged that NYHA classification is subjective and has low reproducibility, its use is ingrained in both guidelines and contemporary practice, and it serves as a cornerstone of clinical documentation, trial enrollment, and candidacy for therapeutics in HF.[2,3] This use has implications for the success of further interventions: currently on ClinicalTrials.gov, 304 ongoing studies have the NYHA classification as an inclusion or exclusion criterion. As a result, guideline recommendations and FDA approval of invasive interventions such as cardiac resynchronization therapy, implantable pulmonary artery pressure monitoring (CardioMEMS HF System (Abbott)), and left ventricular assist devices are firmly anchored in NYHA class.[4–6]

Despite the ubiquity of the NYHA classification system in HF, its clinical implications are less clear. There is no consistent method for accurate assessment of functional

From the Section of Cardiovascular Medicine, Center for Outcomes Research Evaluation (CORE) (C.C., N.R.D., L.S.C., T.A.) and Program of Applied Translational Research (F.P.W.), Yale University School of Medicine, New Haven, CT; Duke Clinical Research Institute, Durham, NC (H.M., B.A., M.F., G.M.F.); Detroit Medical Center, Detroit, MI (I.L.P.); Inova Heart and Vascular Institute, Fairfax, VA (C.M.O.); Vanderbilt University Medical Center; Nashville, TN (J.L.); Massachusetts General Hospital and Baim Institute for Clinical Research, Boston, MA (J.L.J.).

Correspondence to: Tariq Ahmad, MD, MPH, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06517. E-mail: tariq.ahmad@yale.edu

Received August 13, 2019; accepted October 25, 2019.

Clinical Perspective

What Is New?

- The New York Heart Association (NYHA) functional classification serves as a fundamental descriptor of heart failure and is used clinically and to determine trial eligibility.
- Using data from published trials, we found that NYHA class II versus class III is an unreliable predictor of adverse outcomes in heart failure and poorly discriminates among patients across the spectrum of functional impairment.

What Are the Clinical Implications?

- Continued usage of NYHA class in guidelines and trials, for US Food and Drug Administration approval of therapies, and for clinical decision making may hinder efforts to bring precision medicine to the bedside of heart failure patients.

class, and its relations with objective measures of HF (eg, NT-proBNP [N-terminal pro–B-type natriuretic peptide]) are unknown.[7] Consequently, we sought to examine the association of NYHA class with adverse outcomes and objective measures of HF in previously published landmark clinical trials.

Methods

Data from the following National Institutes of Health–funded HF clinical trials were used to examine the association of NYHA functional class with survival: TOPCAT (Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist), DIG (The Effect of Digoxin on Mortality and Morbidity in Patients With Heart Failure), HF-ACTION (Efficacy and Safety of Exercise Training in Patients With Chronic Heart Failure), and GUIDE-IT (Guiding Evidence-Based Therapy Using Biomarker Intensified Treatment in Heart Failure).[8–11] TOPCAT and DIG trial data were obtained from the publicly available database BioLINCC, whereas HF-ACTION and GUIDE-IT data were obtained from the Duke Clinical Research Institute (DCRI).[12] TOPCAT patients from Russia and the Republic of Georgia were excluded because of concerns that these patients were misclassified as having HF.[13] NYHA classes II and III were used, given the low number of patients classified as class I or IV in all trials. Kaplan–Meier failure curves were created to illustrate time to all-cause mortality up to 20 months from randomization; Kaplan–Meier failure rates at 20 months were reported, and pairwise comparisons were performed comparing all class II and class III rates across trials. Distributions of the following commonly used objective HF variables were overlaid according to NYHA class from HF-ACTION and GUIDE-IT: left ventricular ejection fraction, 6-minute walk distance, NT-proBNP, and Kansas City Cardiomyopathy Questionnaire (KCCQ) overall score. Finally, distributions of variables representing the gold standard for functional status in HF—cardiopulmonary exercise testing—were plotted according to NYHA class in HF-ACTION. Median values with the 25th percentile (first quartile) and 75th percentile (third quartile) are presented in the plots. The Wilcoxon rank sum test was used to evaluate differences in distributions. The percentage overlap between classes II and III was calculated by estimating the overlapping area of the 2 kernel density estimations for each objective measure. Two-tailed $P<0.05$ was considered statistically significant. All analyses were carried out using SAS v9.4 (SAS Institute) and R v3.4.2 (R Foundation for Statistical Computing). The institutional review boards at Yale University School of Medicine and DCRI approved the study and waived the requirement for informed consent. Because of the sensitive nature of the data collected for this study, requests to access the data set from qualified researchers trained in human subject confidentiality protocols may be sent to Brooke Alhanti at DCRI (brooke.alhanti@duke.edu).

Results

Cumulative mortality varied significantly across NYHA class and clinical trial, ranging from ≈7% to ≈25% (overall likelihood ratio, $P<0.001$; Figure 1). Those who were NYHA class III in TOPCAT had survival similar to those characterized as NYHA class II in the GUIDE-IT and DIG trials. Mortality at 20 months for NYHA class II was 7.0% for HF-ACTION, 8.1% for TOPCAT, 14.3% for DIG, and 15.0% for GUIDE-IT. Mortality for NYHA class III was 12.1% for TOPCAT, 13.6% for HF-ACTION, 24.3% for DIG, and 26.5% for GUIDE-IT.

Distributions for objective HF variables assessed in GUIDE-IT and HF-ACTION, stratified by NYHA class, are shown in Figure 2. Numbers of patients classified as NYHA classes II and III, respectively, were as follows: GUIDE-IT, n=447 and n=358; HF-ACTION, n=1477 and n=831. The percentage of overlap among patients who were classified as NYHA classes II and III, respectively, was as follows: NT-proBNP levels, 79% and 69%; KCCQ, 63% and 54%; 6-minute walk distances, 63% and 54%; and left ventricular ejection fraction, 88% and 83%. At a population level, however, we noted statistically significant differences in median levels of NT-proBNP, KCCQ score, and 6-minute walk distance (all $P<0.001$) but not left ventricular ejection fraction ($P=0.76$).

In addition, we assessed the overlap in distributions of variables that reflect the gold standard measurement for maximal functional capacity in HF—cardiopulmonary exercise testing—according to NYHA classification in HF-ACTION. As shown in Figure 3, although there were statistically significant differences in median levels of minute ventilation–carbon dioxide production relationship, maximal oxygen uptake, and exercise duration ($P<0.001$), we noted substantial overlap in

these measures between NYHA class II versus III (71%, 54%, and 53%, respectively).

Discussion

This analysis of 4 landmark HF trials demonstrates that the NYHA system poorly differentiates patients across the spectrum of functional impairment. In this report, we examined both the macro- and microimplications of the NYHA classification system across the spectrum of HF and found that it is an unreliable predictor of survival and a poor discriminator of functional impairment in HF. A heterogeneity of risk is strikingly clear in similar NYHA classifications across studies from lower risk (eg, HF-ACTION) to higher risk (eg, GUIDE IT) and across trials including patients with HF reduced and preserved ejection fraction, implying that the prognostic value of NYHA classification is largely dependent on the baseline risk of the patient in which it is assessed. This suggestion is contrary to the general assumption that the NYHA classification is an accurate measure of mortality risk and is consistent across studies for patients of a similar class.

Whereas the heterogeneity of NYHA class across trials is a recognized consequence of differences in inclusion and exclusion criteria of the studies assessed and the heterogeneity of risk in these studies, use of NYHA symptom severity by regulatory bodies does not necessarily take this limitation into consideration. Once clinical trials are completed, therapies may be approved for specific NYHA classes and suggested based on post hoc analyses of the data.[14] For example, the CardioMEMS HF System is presently approved for NYHA class III, and a clinical trial with expected enrollment of 3600 is ongoing to extend its approval to NYHA class II patients (NCT03387813). Our findings suggest it is time to revisit the use of NYHA class to guide enrollment into trials and approval for therapy on its basis.

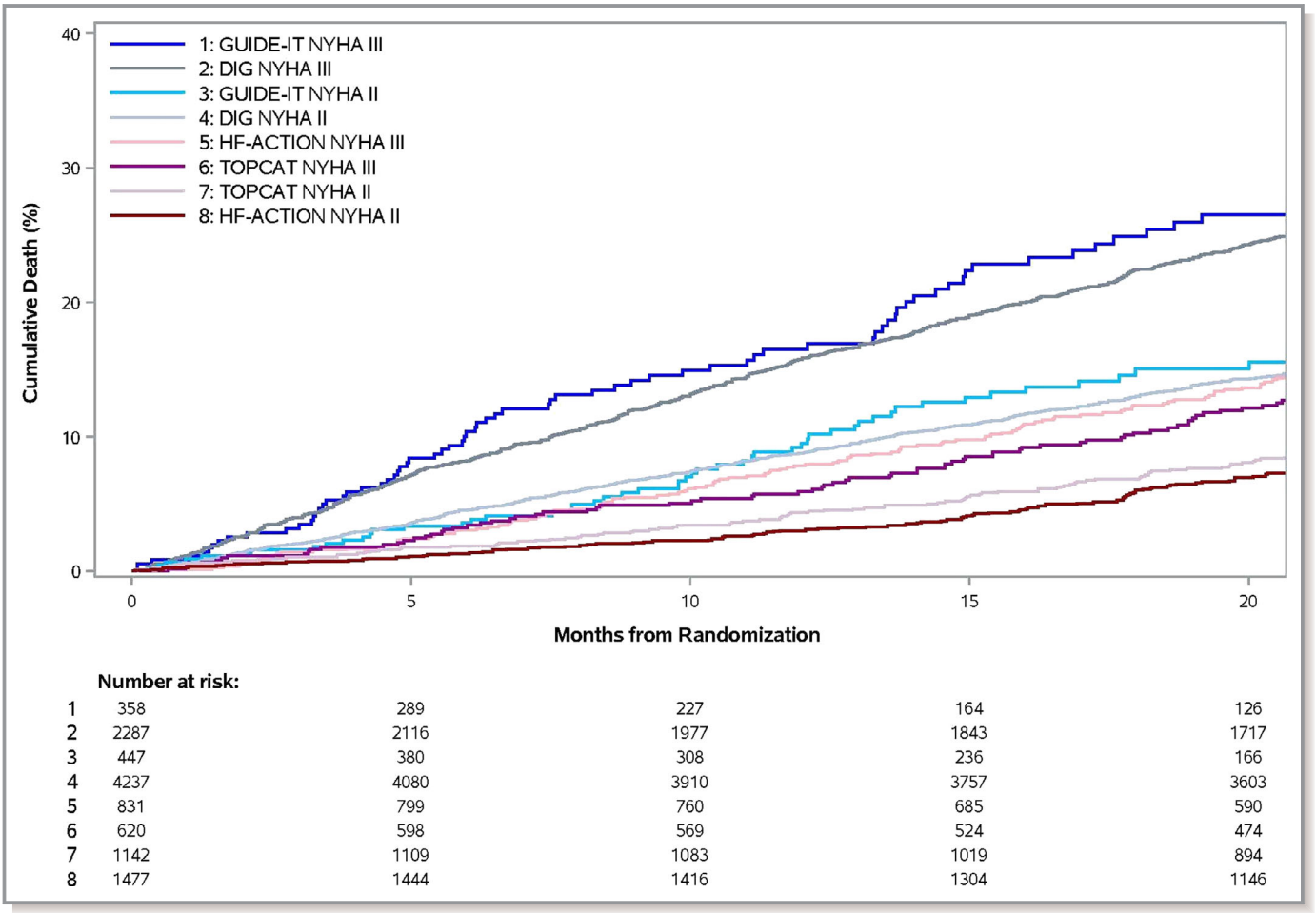

Figure 1. Kaplan–Meier curves for all-cause mortality according to clinical trial and New York Heart Association (NYHA) classification. Clinical trials shown are TOPCAT (Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist), DIG (The Effect of Digoxin on Mortality and Morbidity in Patients With Heart Failure), HF-ACTION (Efficacy and Safety of Exercise Training in Patients With Chronic Heart Failure), and GUIDE-IT (Guiding Evidence-Based Therapy Using Biomarker Intensified Treatment in Heart Failure).

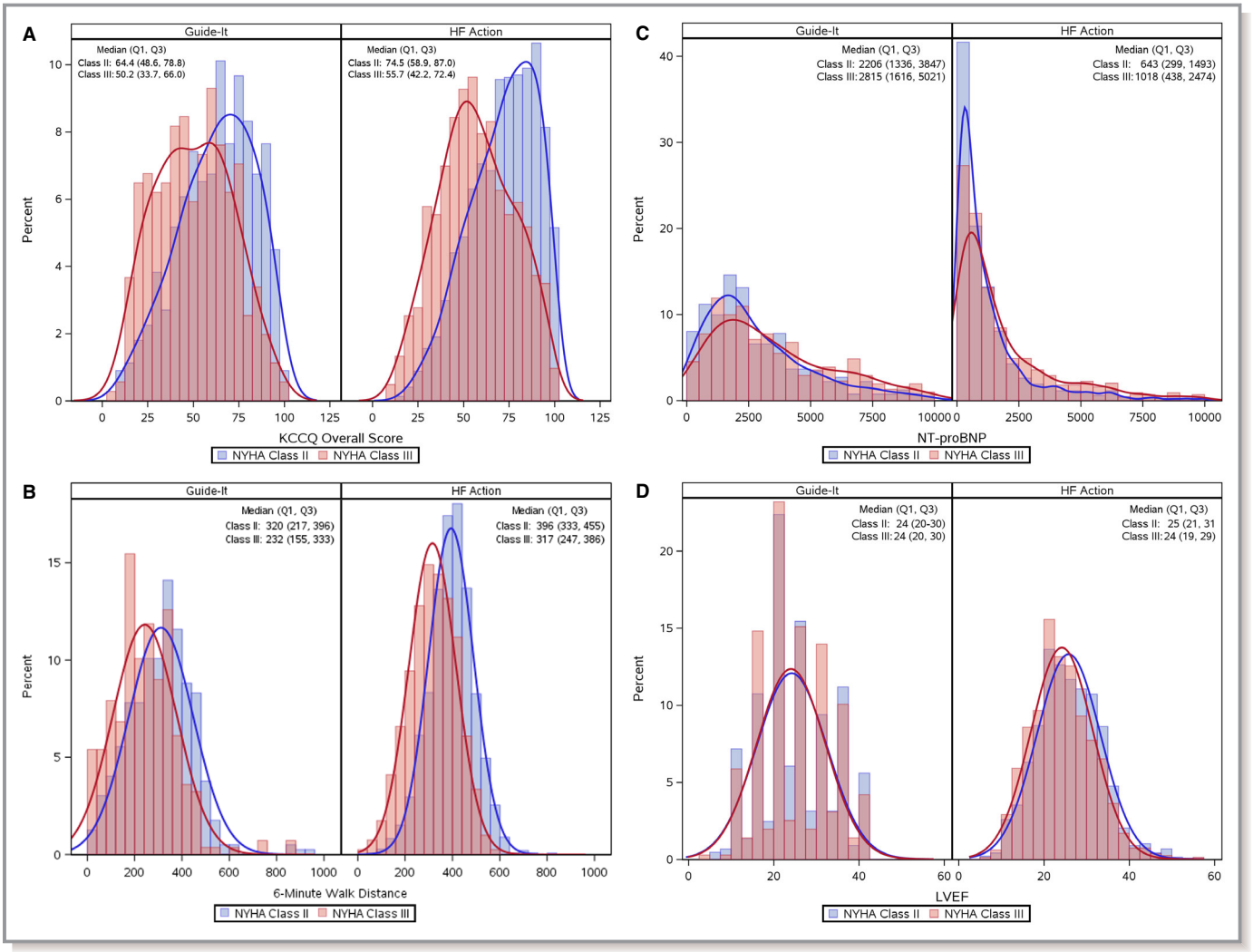

Figure 2. Distributions of objective measures of heart failure according New York Heart Association (NYHA) classes II and III (red shows overlap in values). Numbers of patients classified as NYHA classes II and III, respectively, were as follows: GUIDE-IT, n=447 and n=358); HF-ACTION, n=1477 and n=831. Values represent medians and interquartile ranges between NYHA classes II and III in GUIDE-IT and HF-ACTION clinical trials. **A**, Kansas City Cardiomyopathy Questionnaire (KCCQ) distributions. **B**, Six-minute walk distance distributions. **C**, NT-proBNP (N-terminal pro–B-type natriuretic peptide) distributions. **D**, Left ventricular ejection fraction (LVEF) distributions. GUIDE-IT indicates Guiding Evidence-Based Therapy Using Biomarker Intensified Treatment in Heart Failure; HF-ACTION, Efficacy and Safety of Exercise Training in Patients With Chronic Heart Failure; Q, quartile.

Other more objective and better calibrated measures of disease severity and patient-reported symptoms such as the KCCQ, the Minnesota Living with Heart Failure Questionnaire, and biomarkers might be better suited to guide enrollment strategies and to appraise the impact of therapeutic interventions on patient symptoms.

Although the median levels of almost all objective HF parameters differed significantly between NYHA classes II and III, there was immense overlap in values. These findings, along with the longstanding recognition that the NYHA classification system has poor reproducibility, raise the question of whether our care of HF patients might be enhanced if we elevated the clinical use of disease descriptors that are more objective and precise.[3] Furthermore, these limitations make its centrality to the HF guidelines potentially inconsistent with the goal of improving patient care.

Limitations

Several limitations should be considered. First, we did not include patients classified as NYHA class I or IV because they constituted a small minority in the HF trials we assessed. Second, detailed phenotyping of HF patients that included natriuretic peptide levels and KCCQ assessments was available for only 2 of the trials, and only 1 trial had data on cardiopulmonary exercise testing, the gold standard for measure of functional status in HF. Third, we did not have data on real-world use of NYHA classification, but it would not

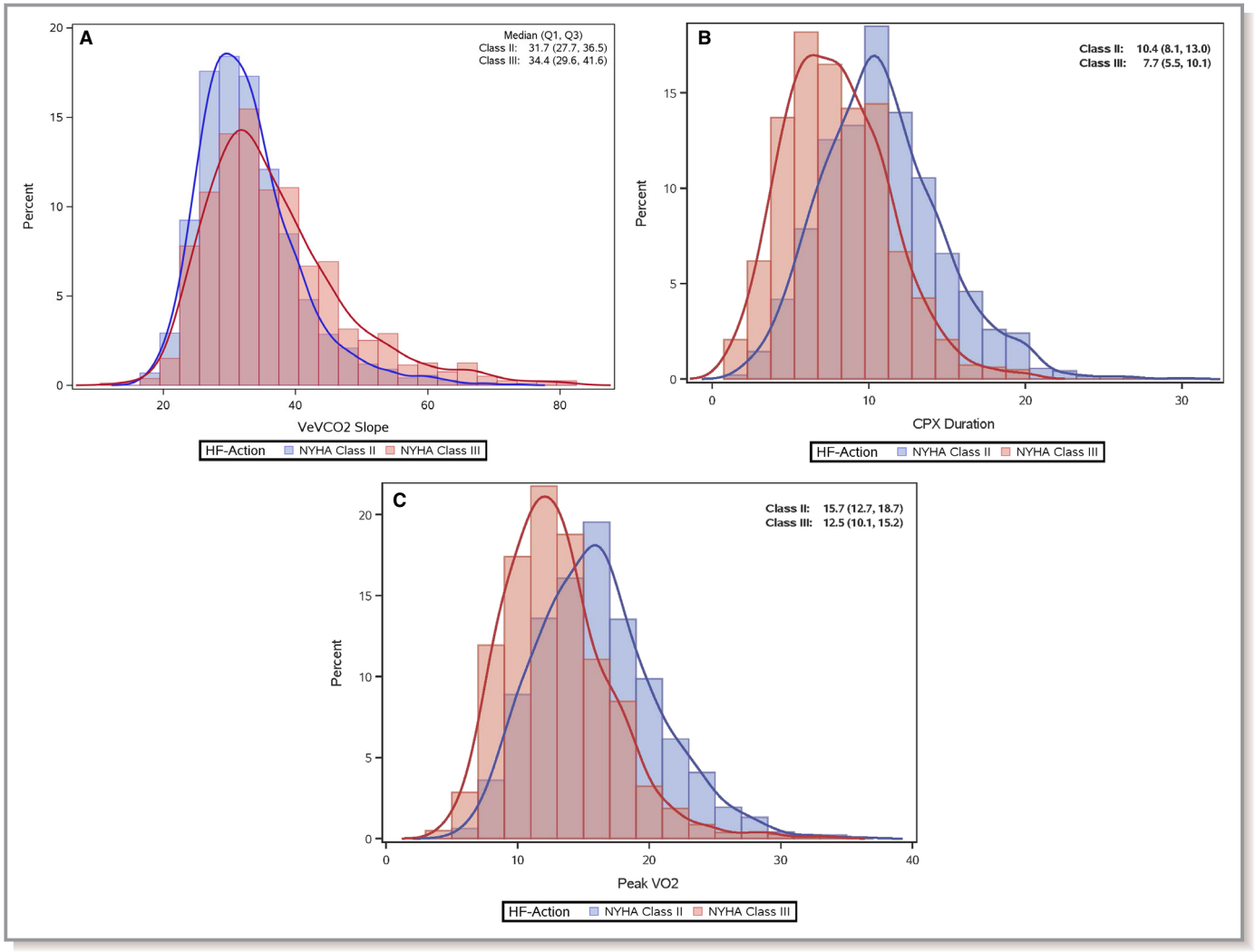

Figure 3. Distributions of cardiopulmonary exercise testing variables according New York Heart Association (NYHA) classes II and III in HF-ACTION (red shows overlap in values). Values in figure represent medians and interquartile ranges. **A**, Ventilation–carbon dioxide production relationship (VE/V_{CO_2} slope). **B**, Cardiopulmonary exercise testing (CPX) duration. **C**, Maximal oxygen uptake (peak V_{O_2}). HF-ACTION indicates Efficacy and Safety of Exercise Training in Patients With Chronic Heart Failure; Q, quartile.

be expected to change our conclusions in a meaningful manner. Fourth, we did not limit our analysis to HF patients with reduced ejection fraction trials for the survival analyses. Prior studies have shown that prognosis is similar in HF patients with reduced or preserved ejection fraction. Fifth, the trials covered a long period of time, and therapies for HF have improved; however, the "newest" clinical trial—GUIDE-IT—had patients who did the worst within similar categories of NYHA class, supporting our hypothesis.

Conclusions

The NYHA classification system is an unreliable predictor of adverse outcomes in HF and poorly discriminates among patients across the spectrum of functional impairment. Its continued usage in guidelines, clinical trials, for US Food and Drug Administration approval of therapies, and for clinical decision making may hinder our progress toward bringing precision medicine to the bedside of HF patients.

Disclosures

None.

References

1. White PD, Myers MM. The classification of cardiac diagnosis. *JAMA*. 1921;77:1414–1415.
2. Goldman L, Hashimoto B, Cook EF, Loscalzo A. Comparative reproducibility and validity of systems for assessing cardiovascular functional class: advantages of a new specific activity scale. *Circulation*. 1981;64:1227–1234.
3. Raphael C, Briscoe C, Davies J, Ian Whinnett Z, Manisty C, Sutton R, Mayet J, Francis DP. Limitations of the New York Heart Association functional classification system and self-reported walking distances in chronic heart failure. *Heart*. 2007;93:476–482.

4. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr, Drazner MH, Fonarow GC, Geraci SA, Horwich T, Januzzi JL, Johnson MR, Kasper EK, Levy WC, Masoudi FA, McBride PE, McMurray JJ, Mitchell JE, Peterson PN, Riegel B, Sam F, Stevenson LW, Tang WH, Tsai EJ, Wilkoff BL. 2013 ACCF/AHA guideline for the management of heart failure: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2013;128:1810–1852.
5. Abraham WT, Adamson PB, Bourge RC, Aaron MF, Costanzo MR, Stevenson LW, Strickland W, Neelagaru S, Raval N, Krueger S, Weiner S, Shavelle D, Jeffries B, Yadav JS; CHAMPION Trial Study Group. Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: a randomised controlled trial. *Lancet*. 2011;377:658–666.
6. Rogers JG, Aaronson KD, Boyle AJ, Russell SD, Milano CA, Pagani FD, Edwards BS, Park S, John R, Conte JV, Farrar DJ, Slaughter MS; HeartMate III. Continuous flow left ventricular assist device improves functional capacity and quality of life of advanced heart failure patients. *J Am Coll Cardiol*. 2010;55:1826–1834.
7. Baggish AL, van Kimmenade RR, Pinto Y, Richards AM, Lainchbury J, Bayes-Genis A, Santalo M, Ordonez-Llanos J, Januzzi JL. New York Heart Association class versus amino-terminal pro-B type natriuretic peptide for acute heart failure prognosis. *Biomarkers*. 2010;15:307–314.
8. O'Connor CM, Whellan DJ, Lee KL, Keteyian SJ, Cooper LS, Ellis SJ, Leifer ES, Kraus WE, Kitzman DW, Blumenthal JA, Rendall DS, Miller NH, Fleg JL, Schulman KA, McKelvie RS, Zannad F, Pina IL. Efficacy and safety of exercise training in patients with chronic heart failure: HF-ACTION randomized controlled trial. *JAMA*. 2009;301:1439–1450.
9. Pitt B, Pfeffer MA, Assmann SF, Boineau R, Anand IS, Claggett B, Clausell N, Desai AS, Diaz R, Fleg JL, Gordeev I, Harty B, Heitner JF, Kenwood CT, Lewis EF, O'Meara E, Probstfield JL, Shaburishvili T, Shah SJ, Solomon SD, Sweitzer NK, Yang S, McKinlay SM; TOPCAT Investigators. Spironolactone for heart failure with preserved ejection fraction. *N Engl J Med*. 2014;370:1383–1392.
10. Felker GM, Anstrom KJ, Adams KF, Ezekowitz JA, Fiuzat M, Houston-Miller N, Januzzi JL Jr, Mark DB, Pina IL, Passmore G, Whellan DJ, Yang H, Cooper LS, Leifer ES, Desvigne-Nickens P, O'Connor CM. Effect of natriuretic peptide-guided therapy on hospitalization or cardiovascular mortality in high-risk patients with heart failure and reduced ejection fraction: a randomized clinical trial. *JAMA*. 2017;318:713–720.
11. Digitalis Investigation G. The effect of digoxin on mortality and morbidity in patients with heart failure. *N Engl J Med*. 1997;336:525–533.
12. Giffen CA, Wagner EL, Adams JT, Hitchcock DM, Welniak LA, Brennan SP, Carroll LE. Providing researchers with online access to NHLBI biospecimen collections: the results of the first six years of the NHLBI BioLINCC program. *PLoS One*. 2017;12:e0178141.
13. Pfeffer MA, Claggett B, Assmann SF, Boineau R, Anand IS, Clausell N, Desai AS, Diaz R, Fleg JL, Gordeev I, Heitner JF, Lewis EF, O'Meara E, Rouleau JL, Probstfield JL, Shaburishvili T, Shah SJ, Solomon SD, Sweitzer NK, McKinlay SM, Pitt B. Regional variation in patients and outcomes in the Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist (TOPCAT) trial. *Circulation*. 2015;131:34–42.
14. Providencia R, Boveda S, Defaye P, Segal O, Algalarrondo V, Sadoul N, Lambiase P, Piot O, Klug D, Perier MC, Bouzeman A, Barra S, Bories MC, Gras D, Fauchier L, Bordachar P, Babuty D, Deharo JC, Leclercq C, Marijon E; DAI-PP Investigators. Outcome of primary prevention implantable cardioverter defibrillator therapy according to New York Heart Association functional classification. *Am J Cardiol*. 2016;118:1225–1232.