ORIGINAL ARTICLE

Wide Variation in Clinicians' Assessment of New York Heart Association/World Health Organization Functional Class in Patients With Pulmonary Arterial Hypertension

DARREN B. TAICHMAN, MD, PhD; MICHAEL D. McGOON, MD; MICHAEL O. HARHAY, MBe; CHRIS ARCHER-CHICKO, MSN; JEFFREY S. SAGER, MD; MEENA MURUGAPPAN, MD; MURALI M. CHAKINALI, MD; HAROLD I. PALEVSKY, MD; AND ROBERT GALLOP, PhD

OBJECTIVE: **To assess interrater reliability of the New York Heart Association/World Health Organization functional classification as applied by clinicians (defined as both physicians and nurses in this article) to patients with pulmonary arterial hypertension (PAH).**

PATIENTS AND METHODS: **Between March 16 and August 31, 2007, a survey that described 10 hypothetical patients was completed by physicians and nurses attending a conference on PAH. Results were subsequently validated with physicians and nurses who were contacted online through the Pulmonary Hypertension Association. Respondents were asked to assign each patient's functional class as they would normally in clinical practice.**

RESULTS: **The functional class evaluations were completed by 113 clinicians, 87 (77%) of whom had participated in PAH trials; 106 (94%) reported using functional class when determining therapy. Clinicians reported a broad range of factors they considered when evaluating functional class, and their assessments of functional class varied widely. The intraclass correlation coefficient was 0.58 for the initial patient survey and 0.62 for the online survey. At best, one patient was ranked as either class II (by 60 clinicians [53%]) or class III (by 53 [47%]). Clinicians' rankings spanned at least 3 functional classes for each of the other patients. Equally divergent rankings were observed among nurses and physicians. Cluster analysis identified clinicians' tendencies toward "higher" or "lower" functional class rankings. Of the 113 clinicians, 101 (89%) thought that the patients described resembled those seen in their practices.**

CONCLUSION: **Despite the wide use of the New York Heart Association/World Health Organization functional class in clinical care and as a research tool, interrater agreement may be inadequate. Efforts to promote a uniform approach to evaluating functional class might help to standardize PAH care and research.**

Mayo Clin Proc. 2009;84(7):586-592

ICC = intraclass correlation coefficient; NYHA = New York Heart Association; PAH = pulmonary arterial hypertension; REVEAL = Registry to Evaluate Early and Long-Term PAH Disease Management; WHO = World Health Organization

A modified New York Heart Association (NYHA) functional classification system was adopted by the World Health Organization (WHO) in 1998 to facilitate evaluation of patients with pulmonary arterial hypertension (PAH).[1] Since then, the NYHA/WHO functional class has been used not only for routine evaluation of a patient's status when providing clinical care but also as an enrollment criterion and end point in clinical trials of therapy for PAH. Indications approved by the US Food and Drug Association for the labeling of PAH-specific therapies

stipulate specific NYHA/WHO functional classes.[2] Consensus guidelines of major clinical societies include the NYHA/WHO functional class in evaluation of patients and recommend therapies on the basis of a patient's functional class.[3-5] Accordingly, access to some therapies is restricted by insurance payers in part by the NYHA/WHO functional class assigned to the patient.

The NYHA functional classification system was originally adopted in 1928 as a means of assessing the effect of cardiac disease on patients seen in clinical practice.[6] Its use has expanded from an aid for clinical assessment and diagnosis as originally proposed to its wide adoption as a research tool in studies of ischemic and other left-sided heart disorders.[7] Assessment of interrater reliability (the degree to which individual clinicians will independently assign the same functional class to patients) of the NYHA functional classification has been limited in these disease states[7,8] and nonexistent for the application of the modified NYHA/WHO functional classification of patients with PAH. Given the use of functional class not only in describing the participants enrolled in clinical trials and their response to investigational agents but also in guiding application of these results to the care of individual patients, we performed an initial evaluation of the agreement among clinicians in assessing functional class in patients with PAH.

© 2009 Mayo Foundation for Medical Education and Research

PATIENTS AND METHODS

CLINICIANS' EVALUATIONS OF NYHA/WHO FUNCTIONAL CLASS

We evaluated agreement between clinicians in their assessment of NYHA/WHO functional class by presenting information about 10 hypothetical patients with PAH (eAppendix online linked to this article). The patients described were selected to represent a range of PAH severity. Because the goal was to assess whether clinicians approached functional class in the same manner, no "gold standard" or "correct" method of determining functional class was assumed; a "correct" functional class ranking for the patients was not provided. The presence or absence of symptoms with activities of daily living, time since diagnosis, hemodynamic values, 6-minute walk distance, PAH-specific medications used, and the patient's occupation were included in varying combinations. Four physicians and a nurse practitioner with expertise in evaluation and treatment of PAH were individually asked to review the patients presented to ensure they were similar to those seen in daily practice and that all 4 NYHA/WHO functional classes were represented according to each of these clinician's approaches to assessment (which were not specified). Two of the physicians and the nurse practitioner were involved in PAH-related research, one physician in an academic but non-PAH research practice and one physician in a specialized referral private practice. No attempt was made to assess agreement in functional class rankings among these practitioners.

Between March 16 and August 31, 2007, these descriptions of 10 patients were presented to physicians and nurses who provide care to patients with PAH. Clinicians were instructed to read each of the patient descriptions and then "assign a NYHA/WHO class as you would in your own practice." Accordingly, no guide was provided as to how judgments might be made, nor was a definition of NYHA/WHO functional class provided. Clinicians were told that "there are no intended correct answers" and that the goal was to "simply see if caregivers come to the same conclusions." Respondents were asked to refrain from discussing the questions with others. Participants were assured that all responses would be kept anonymous, with no means by which study personnel could link answers to respondents' identities. Participants were offered entry into a raffle for a free registration at the 2008 international meeting of the Pulmonary Hypertension Association (Silver Spring, MD). Raffle entries were separated from responses to the patient descriptions.

An initial evaluation was performed using paper forms distributed to participants at an investigators' meeting for the Registry to Evaluate Early and Long-Term PAH Disease Management (REVEAL) on March 16, 2007. REVEAL is a multicenter patient cohort study sponsored by Actelion Pharmaceuticals.[9] No direct industry sponsorship was obtained for the current study, which was not part of the REVEAL study protocol. To validate the findings, a second evaluation was performed in which all physician and nurse members of the Pulmonary Hypertension Clinicians and Researchers section of the Pulmonary Hypertension Association were contacted by e-mail and invited to complete the study online. Participants accessed the patient descriptions and entered their judgments of NYHA/WHO class via an internet link at a site maintained by SurveyMonkey, Inc (www.SurveyMonkey.com). Access to the evaluation was available for 1 month (August 2007).

After patient descriptions, information was collected on respondents' occupation, practice, types of patients and medications managed, and factors considered when evaluating NYHA/WHO functional class. Only respondents who indicated active involvement in the clinical care of patients with PAH were included in the analysis. For the online investigation, respondents were asked if they had participated in the initial (paper) evaluation, and those who had were excluded from this second analysis.

To specifically evaluate whether the 6-minute walk distance influenced clinicians' assessments of functional class, a pair of patient case scenarios was presented with different walk distances (300 vs 390 m) but otherwise equivalent variables (age 43 vs 44 years; both with scleroderma-associated PAH treated with "a single oral agent for PAH," both using "supplemental oxygen" and reporting "dyspnea carrying groceries" but denying "light-headedness or peripheral edema") (patients No. 2 and 7, eAppendix online linked to this article). Despite imperfections in this test's performance as a predictor of patient outcomes, the 6-minute walk distance was assessed specifically because of its primacy as an end point in PAH clinical trials.[10,11]

STATISTICAL ANALYSES

Data were entered into Access 5.1 (Microsoft, Redmond, WA), and analysis was performed using SAS 9.1 (SAS, Cary, NC). Agreement between clinician raters regarding the patients' NYHA/WHO functional classes was expressed as intraclass correlation coefficients (ICCs) based on a 2-way random-effects model analysis of variance which included both patients and clinician raters. Variance between patients (reflecting differences in disease impact) was partitioned from the variance within patients (ie, variation in the ratings assigned to any one patient by multiple clinician raters). Despite significant differences in disability between patients, strong "agreement" among the nurse and physician raters would be reflected in a low variance within patients. Perfect agreement between the clinician raters regarding the patients' functional classes would correspond to 0 variance within patients, resulting in an

TABLE 1. **Clinicians Participating in Study of Assessment of NYHA/WHO Functional Class in Patients With PAH**

	Paper	Online
No. of participants	44	69
Physicians	28 (64)	40 (58)
Nurses/nurse practitioners	16 (36)	29 (42)
Specialty		
Pulmonary	29 (66)	43 (62)
Cardiology	15 (34)	23 (33)
Not listed	0	3 (4)
No. of years treating patients with PAH		
0 to <1	0	0
1 to < 3	4 (9)	13 (19)
3 to <5	7 (16)	13 (19)
5 to <10	5 (11)	20 (29)
>10	28 (64)	20 (29)
Not listed	0	3 (4)
Estimated No. of new patients with PAH evaluated annually		
<10	3 (7)	10 (15)
10 to 20	6 (14)	19 (28)
21 to 50	12 (27)	23 (33)
≥50	23 (52)	15 (22)
Not listed	0	2 (3)
PAH treatment prescribed or managed		
Calcium channel blockers	39 (89)	59 (86)
Bosentan	41 (93)	65 (94)
Ambrisentan	17 (39)	41 (59)
Sitaxentan	19 (43)	0 (0)
Sildenafil	40 (91)	66 (96)
Tedalafil	18 (41)	0 (0)
Inhaled iloprost	37 (84)	58 (84)
IV treprostinil	34 (77)	44 (64)
Subcutaneous treprostinil	30 (68)	48 (70)
IV epoprostenol	39 (89)	57 (83)

Data are number (percentage). IV = intravenous; NYHA = New York Heart Association; PAH = pulmonary arterial hypertension; WHO = World Health Organization.

ICC of 1.[12] The agreement in ratings for the 2 patients differing only in 6-minute walk distance was assessed with a κ coefficient, quantifying the agreement between 2 ratings over the raters, beyond chance.[13] The use of 10 patients and at least 99 total clinician raters provided 80% power to distinguish between moderate and substantial agreement.[14]

To evaluate whether groups of raters who judge functional class in similar fashions could be identified, and if those who differed did so in a systematic manner, nonhierarchical clustering was performed using the FASTCLUS procedure in SAS.[15] A "distance" representing the dissimilarity between each pair of raters was determined on the basis of their NYHA/WHO rankings of 10 patients. The 2 raters with the smallest distance or dissimilarity were initially joined to form a group (or "cluster"), and subsequently groups of raters were systematically combined.[16] Cluster size was determined by the rule that each must contain at least 10% of the sample to maximize the differentiation of groups while retaining sufficient cluster size to permit a statistically meaningful analysis.[17] A potential asso-

ciation between occupation (physician or nurse) and group/cluster membership was assessed by χ^2. Differences in the paper and online groups in the variables they reported using when they considered functional class were assessed with the Fisher exact test for contingency tables.

This study was approved by the University of Pennsylvania Institutional Review Board.

RESULTS

CLINICIAN PARTICIPANTS

A total of 113 clinicians participated in the 2 evaluations (Table 1). The initial paper administration involved 28 physicians and 16 nurses; 40 physicians and 29 nurses completed the online version. An additional 6 physicians who completed the online evaluation indicated that they had also participated in the initial paper version and thus were excluded from the second analysis. All participants indicated that they were actively involved in the clinical care of patients with PAH. In the initial evaluation, all participants practiced in the United States, as did 63 (91%) in the online sample; 3 of the online respondents practiced in Canada and 1 each in Argentina, China, Mexico, and the Netherlands. Of the participants in the 2 evaluations, 111 (98%) practiced at academic medical centers, and 88 (78%) indicated they had enrolled patients into clinical trials or studies of PAH. Moreover, 101 (89%) of the physicians and nurses believed that the patients described in the study resembled those they might see in their own practices.

Of the 113 participants, 108 (96%) indicated that they use the NYHA/WHO functional classification as part of their evaluation when selecting therapy for their patients. Factors they considered when formulating functional class assessments are listed in Table 2.

AGREEMENT IN NYHA/WHO FUNCTIONAL CLASS ASSIGNMENTS

The NYHA/WHO functional class assessments varied, with a wide range observed in the rankings assigned to

TABLE 2. **Factors Clinicians Considered When Formulating Functional Class Assessments**

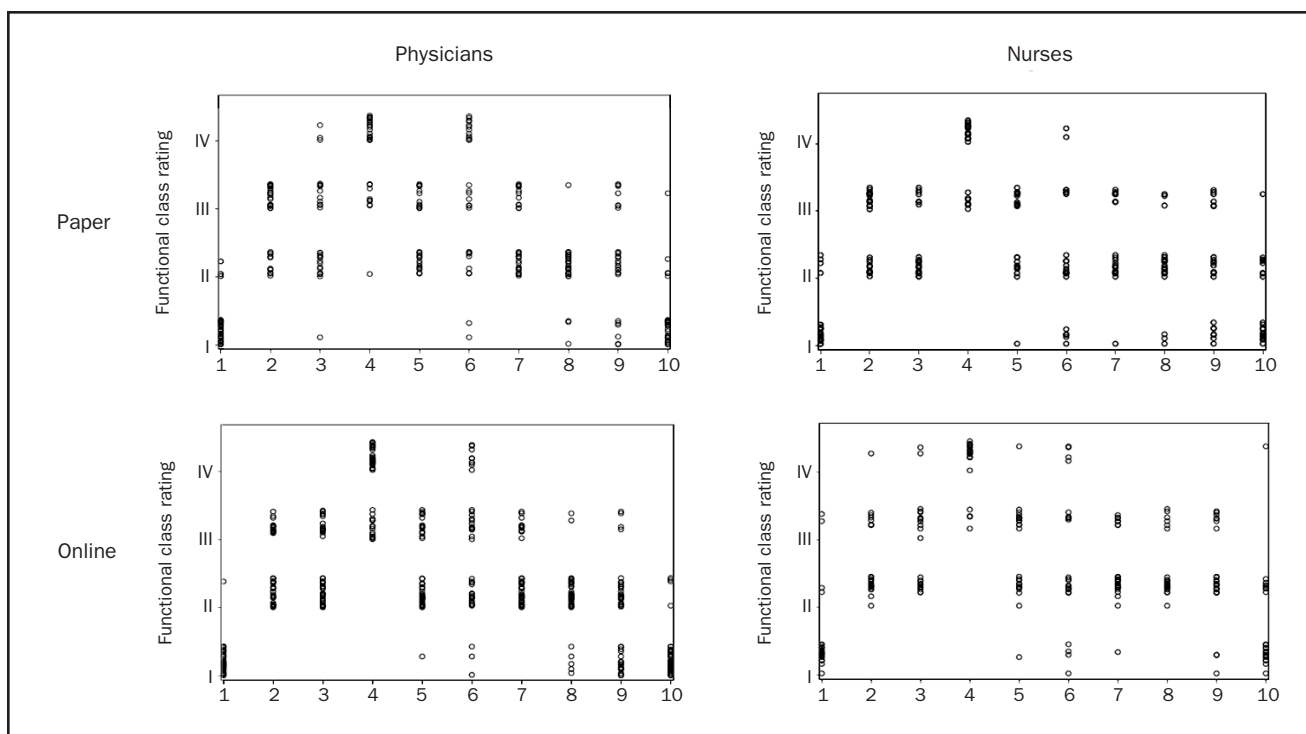|	Paper	Online	P value
Symptoms	41 (93)	68 (99)	.13
Medication used	10 (23)	17 (25)	>.99
Hemodynamic values	11 (25)	0	<.001
Physical examination	24 (55)	31 (45)	.34
6-minute walk distance	18 (41)	41 (59)	.08
Echocardiographic findings	8 (18)	0	<.001
Patient's lifestyle	33 (75)	46 (67)	.40
Patient's occupation	23 (52)	31 (45)	.56
Insurance requirements	9 (20)	0	.01

Data are number (percentage).

FIGURE 1. Individual physician and nurse's ranking of New York Heart Association/World Health Organization functional class for each of 10 patient case scenarios. Physician and nurses' ranking on the initial paper evaluation are shown at the top and from the subsequent online groups of clinicians on the bottom. No significant difference was seen in the distribution of functional class assignments made by each of the physician and nursing groups participating in the paper and online assessments (χ^2 [9] = 10.83; P=.29 for physicians; χ^2 [9] = 8.39; P=.50 for nurses as assessed through a generalized estimating equations approach accounting for the multiple ratings per clinician).

individual patients by both physicians and nurses. Results of the initial paper and subsequent online evaluations were similar, with consistent disagreement between the clinician raters occurring at each administration. Nearly identical distributions were seen in the functional class assignments made by each of the physician and nursing groups participating in the paper and online assessments (Figure 1). Each of the 10 patient case scenarios was assigned to at least 2 functional classes. At best, an individual patient was judged as being in either of 2 functional classes by the initial paper group of raters (eg, functional class II or class III by 15 [34%] and 29 [66%], respectively; patient No. 2, eAppendix online linked to this article). In the online evaluation group, this same patient was ranked as either class II or class III by 44 (64%) and 25 (36%), respectively. For the clinicians as a group, this patient was ranked as either class II (by 60 clinicians [53%]) or class III (by 53 [47%]). More commonly, discrepancies between clinician evaluations resulted in each patient being ranked in any of 3 functional classes. At worst, a single patient was judged as class I, II, III, and IV by 4 (9%), 16 (36%), 10 (23%) and 14 (32%), respectively, of clinicians in the initial group and by 7 (10%), 28 (40%), 19 (28%) and 15 (22%) of the online group (patient No. 6,

eAppendix online linked to this article). The dominant functional class assignment of the raters for each patient and the percentage of exact matches and those within 1 functional class are shown in Table 3.

Intraclass correlation coefficients for agreement among the physicians and nurse raters in each evaluation group are shown in Table 4. Similarly poor to modest agreement was seen in the original paper survey and subsequent online physician and nursing groups. Overall, ICCs were slightly higher among physicians than nurses.

TABLE 3. **Observed Agreement Among Clinicians**

Patient No.	Dominant functional class response (%)	Exact functional class match, %	Within 1 functional class, %
1	I (90.2)	80.5	96.6
2	II (52.2)	49.7	100
3	II (57.5)	47.2	95.1
4	IV (71.7)	58.5	98.7
5	II (51.3)	45.5	95.7
6	II (38.4)	28.0	68.4
7	II (63.7)	52.7	99.4
8	II (83.9)	70.2	97.0
9	II (49.4)	35.7	84.7
10	I (80.9)	64.4	90.2

TABLE 4. **Intraclass Correlation Coefficients for Each Group of Participants**

	Correlation coefficients (95% CIs)		
	Paper	Online	Combined
Physicians	0.611 (0.407-0.841)	0.681 (0.495-0.879)	0.649 (0.459-0.860)
Nurses	0.550 (0.354-0.806)	0.602 (0.404-0.838)	0.582 (0.389-0.825)
Both physicians and nurses	0.577 (0.377-0.829)	0.645 (0.457-0.859)	0.616 (0.427-0.843)

CI = confidence interval.

CLUSTER ANALYSIS

To assess whether clinicians tend to differ from each other in a consistent manner regarding functional class assessments, a nonhierarchical cluster analysis was performed. Two groups of clinicians were identified on the basis of their overall differences in patient rankings (Figure 2). Members of one group routinely assigned higher functional class rankings to each patient than did members of the other. This difference was not explained by occupation because the distribution of nurses and physicians was nearly identical in the 2 groups ($\chi^2[1] = .02$; $P=.09$ for an association between group membership and occupation). Physicians accounted for 60% (46) and 61% (22) of the clinicians in groups 1 and 2, respectively; nurses, 40% (31) and 39% (14). Approximately two-thirds of the clinicians were in the group that consistently ranked patients into numerically higher functional classes.

EFFECT OF 6-MINUTE WALK DISTANCE ON FUNCTIONAL CLASS ASSESSMENTS

To assess the potential influence of the 6-minute walk distance on clinicians' assessments of NYHA/WHO func-
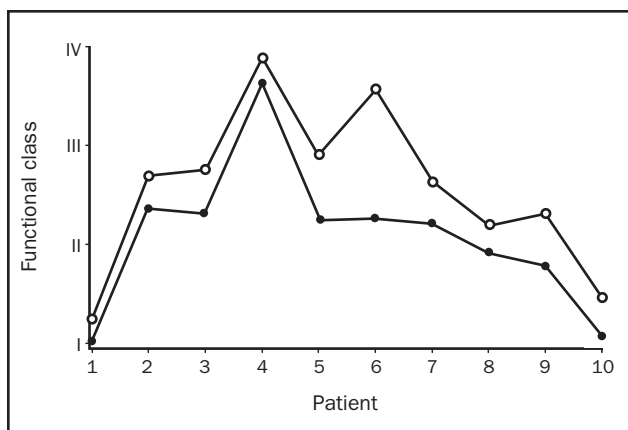

FIGURE 2. Cluster analysis identified 2 groups of clinicians according to their tendency to rank patients with higher or lower New York Heart Association/World Health Organization functional class compared with other clinicians. Of the 113 clinicians, 77 (68%) consistently ranked patients in higher functional classes (open circles), and 36 (32%) consistently assigned lower functional classes (closed circles). Independent analyses of the physicians and nurses participating in each of the paper and online groups resulted in an equivalent group pairing (data not shown).

tional class, a pair of patient case scenarios differing only in the 6-minute walk distance (300 and 390 m) with otherwise equivalent variables was presented. Of the 113 clinicians, 83 (73%) assigned the same functional class to the 2 patients (κ for concordance of evaluations, 0.49; 95% confidence interval, 0.35-0.73; $P=.005$); 20 (18%) of the clinicians assigned a higher (worse) functional class to the patient with the lower 6-minute walk distance, each involving a difference of a single class (from class I to II or from class II to III); and 10 (9%) of the clinicians assigned a lower (better) functional class to the patient case scenarios with the lower 6-minute walk distance.

DISCUSSION

We found a wide variation in the NYHA/WHO functional class assessments made by clinicians who care for and conduct clinical research in patients with PAH. The findings were consistent in 2 independent groups of clinicians assessed with paper and online presentations of patients with PAH. Agreement on a single functional class was not seen for any patient, and in most cases judgments spanned 2 or 3 classes. Consistent with the discrepant conclusions reached was the reported wide variation in factors considered when assessing functional class. Although some clinicians use only patient symptoms to judge functional class, others use varying combinations of symptoms, demographics, exercise, and hemodynamic variables. Indeed, many report using multiple factors not mentioned in the originally defined NYHA or WHO functional class definitions.[1,6] Other potential reasons for the variation in approach might include differences in physician training; inappropriate assumptions regarding the relationships among hemodynamics, exercise capacity and functional class; and a perceived need to "adjust" assignments to meet restrictions in the availability of treatments.

Because PAH results in both substantial morbidity and mortality, the success of new therapies has been assessed not only by survival but also by measures of cardiovascular impairment (eg, hemodynamics), as well as by measures of both functional capacity and functional status.[18] *Functional capacity* represents a patient's potential for physical activity, frequently measured by maximal aerobic capacity or 6-minute walk distance. However, patients do not neces-

sarily make use of their full capacity in daily life. *Functional performance* embodies the physical and emotional realm of activity in which a patient lives day to day and may be assessed by measures such as the physical and emotional domain scores of quality-of-life questionnaires. The difference between functional performance and capacity represents a patient's reserve or the *comfort zone* in which the patient lives.[19] This interplay between performance, capacity, and reserve is the functional status and can be described by the NYHA/WHO functional class.

Progression in a patient's NYHA/WHO functional class (from class I through IV) characterizes his or her declining comfort zone and the impact of debilitating symptoms on daily life. Therefore, assigning functional class requires an integration of a patient's capacity and current performance, both of which are influenced not only by cardiovascular disease and comorbid conditions but also by personal, cultural, and environmental factors that alter the patient's (and the clinician's) perceptions and expectations.[7] Because functional class is a subjective tool, it is not surprising that what is judged as "undue dyspnea" or even "ordinary activity" may differ widely among clinicians assessing NYHA/WHO functional class.

The variation in functional class assessments that we observed does not negate the value of the NYHA/WHO as a prognostic tool.[20] Although clinicians may differ in the conclusions they draw regarding functional class, all clinicians, regardless of individual approach or numerical rank assigned, likely aim to rank the most ill patients in more advanced functional classes. Therefore, the potential concern raised by the current findings is not whether patients judged to be more ill by NYHA/WHO functional class will indeed have worse outcomes than those judged to be in a less severely impaired functional class. Rather, the inconsistencies in the approach taken to functional class assignment make it more difficult to compare results across studies, to evaluate the meaning of changes in functional class reported in response to therapy, and to translate the results of clinical trials to the care of individual patients.

Our study is limited by the lack of actual clinician-patient interaction and the inability of clinicians to pose questions they might ask in routine evaluations; thus, the information available for determining functional class is limited. The consistent distribution of functional class responses found in each of the evaluations and the clear separation of clinicians into 2 groups that consistently ranked patients in either higher or lower functional classes suggest that, despite the inability to fully mimic the process by which judgments are made in clinical practice, true differences exist in the approaches taken. Indeed, the clustering of "higher" and "lower" ratings suggests that sys-

tematic differences may exist. The participating clinicians (and those involved in designing and choosing the 10 patients) may be viewed as having failed to adhere to the "correct" approach in assessing NYHA/WHO functional class (as defined by its original designers or by current "expert opinion"). Our study was designed to ignore any such "gold standard," and thus none was defined. The goal was not to assess whether clinicians arrived at "correct" answers but rather whether they arrived at the same answers when determining functional class as they would normally in clinical practice. We cannot confirm the self-reported experience/expertise of participants in the evaluation of PAH, although the consistent demographic reports from the 2 participating groups suggest that we have captured a group representative of those with a concerted focus in the care of patients with PAH. Finally, we cannot rule out the possibility that the wide variation in responses observed may have been due in part to careless or haphazard completion of functional class assessments by some respondents.

Development of tools to promote a uniform approach to NYHA/WHO functional classification would be an important step in helping to standardize the clinical care of patients with PAH and in performing and interpreting clinical research.[2,21] An alternative means of assessing a patient's clinical status, such as an overall estimation of "higher" or "lower" risk according to a combination of physiologic and other variables (including a uniformly applied NYHA/WHO functional class), might be similarly valuable.[1] The interrater reliability of any such method of evaluation should be confirmed to achieve the goal of a standardized approach to both clinical evaluation and research.

CONCLUSION

We observed a wide variation in the NYHA/WHO functional class assignments made by both physicians and nurses for descriptions of patients with PAH. Despite the wide use of the NYHA/WHO functional class in clinical care and as a research tool, interrater agreement may be inadequate. Our findings further suggest potential systematic differences in clinicians' approach, such that some routinely rank patients in higher (or lower) functional class than do other practitioners. Regardless of whether a "correct" approach exists, the current study suggests it is not being used in the application of functional class assessment by many of those who provide care to patients with PAH and perform PAH research. Future studies involving direct patient (or simulated patient) interactions are justified to further assess differences in clinicians' approaches. Efforts to promote a uniform approach of evaluating functional class or to identify more

reliable functional parameters might help to standardize PAH clinical care and research.

REFERENCES

1. McLaughlin VV, McGoon MD. Pulmonary arterial hypertension. *Circulation*. 2006;114(13):1417-1431.

2. Hoeper MM, Oudiz RJ, Peacock A, et al. End points and clinical trial designs in pulmonary arterial hypertension: clinical and regulatory perspectives. *J Am Coll Cardiol*. 2004;43(12)(suppl S):48S-55S.

3. Badesch DB, Abman SH, Ahearn GS, et al. Medical therapy for pulmonary arterial hypertension: ACCP evidence-based clinical practice guidelines. *Chest*. 2004;126(1 suppl):35S-62S.

4. Badesch DB, Abman SH, Simonneau G, Rubin LJ, McLaughlin VV. Medical therapy for pulmonary arterial hypertension: updated ACCP evidence-based clinical practice guidelines. *Chest*. 2007;131(6):1917-1928.

5. Barst RJ, McGoon M, Torbicki A, et al. Diagnosis and differential assessment of pulmonary arterial hypertension. *J Am Coll Cardiol*. 2004;43(12, suppl S):40S-47S.

6. Criteria Committee of the New York Heart Association. *Nomenclature and Criteria for Diagnosis of Diseases of the Heart and Great Vessels*. 9th ed. Boston, MA: Little, Brown; 1994.

7. Bennett JA, Riegel B, Bittner V, Nichols J. Validity and reliability of the NYHA classes for measuring research outcomes in patients with cardiac disease. *Heart Lung*. 2002;31(4):262-270.

8. Coyne KS, Allen JK. Assessment of functional status in patients with cardiac disease. *Heart Lung*. 1998;27(4):263-273.

9. McGoon MD, Barst RJ, Doyle RL, Liou TG, Miller D, Feldkircher K. REVEAL registry: treatment history and treatment at baseline. *Chest*. 2007;132(4 suppl):631S.

10. Macchia A, Marchioli R, Marfisi R, et al. A meta-analysis of trials of pulmonary hypertension: a clinical condition looking for drugs and research methodology. *Am Heart J*. 2007;153(6):1037-1047.

11. Rich S. The value of approved therapies for pulmonary arterial hypertension [editorial]. *Am Heart J*. 2007;153(6):889-890.

12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.

13. Kraemer HC, Bloch DA. Kappa coefficients in epidemiology: an appraisal of a reappraisal. *J Clin Epidemiol*. 1988;41(10):959-968.

14. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med*. 1998;17(1):101-110.

15. Statistical Analysis Systems Institute. *SAS/STAT User's Guide*. 4th ed. Cary, NC: SAS Institute; 1990.

16. Johnson RA, Wichern DW. *Applied Multivariate Statistical Analyses*. 3rd ed. Englewood Cliffs, NJ: Prentice Hall; 1992.

17. Morral AR, Iguchi MY, Belding MA, Lamb RJ. Natural classes of treatment response [published correction appears in *J Consult Clin Psychol*. 1997;65(5):732]. *J Consult Clin Psychol*. 1997;65(4):673-685.

18. McLaughlin VV, Presberg KW, Doyle RL, et al. Prognosis of pulmonary arterial hypertension: ACCP evidence-based clinical practice guidelines. *Chest*. 2004;126(1 suppl):78S-92S.

19. Miller-Davis C, Marden S, Leidy NK. The New York Heart Association Classes and functional status: what are we really measuring? *Heart Lung*. 2006;35(4):217-224.

20. Friedman EB, Palevsky HI, Taichman DB. Classification and prognosis of pulmonary arterial hypertension. In: Mandel J, Taichman DB, eds. *Pulmonary Vascular Disease*. Philadelphia, PA: Elsevier Science, 2006:66-82.

21. Murugappan M, Sundy R, Byers D, et al. Application of a standardized questionnaire to assign World Health Organization functional classification in pulmonary arterial hypertension patients [abstract C46]. *Am J Respir Crit Care Med*. 2008;177:A919.