Comparative Reproducibility and Validity of Systems for Assessing Cardiovascular Functional Class: Advantages of a New Specific Activity Scale

LEE GOLDMAN, M.D., M.P.H., BEVERLY HASHIMOTO, M.D., E. FRANCIS COOK, Sc.D., AND ANITA LOSCALZO, M.S.

SUMMARY Reproducibility and validity are prerequisites for a useful clinical scale. We therefore prospectively tested the reproducibility and validity of the New York Heart Association criteria and the Canadian Cardiovascular Society criteria for the assessment of cardiac functional class and compared these criteria with a new Specific Activity Scale based on the metabolic costs of specific activities. The New York Heart Association estimates made by two physicians had a reproducibility of only 56%, and only 51% of the estimates agreed with treadmill exercise performance. Functional estimates based on the Canadian Cardiovascular Society criteria were significantly more reproducible (73%), but not significantly more valid. The Specific Activity Scale was as reproducible as the Canadian Cardiovascular Society criteria, and its 68% validity was significantly higher than the validities of the other systems. The easily administered Specific Activity Scale was equally reproducible and valid when used by a nonphysician. It was especially better than the other systems for the evaluation of true class II patients and was significantly less likely to underestimate treadmill performance. Although no set of questions can perfectly predict exercise tolerance, the Specific Activity Scale deserves wider prospective testing.

SEVERAL SYSTEMS have been proposed for categorizing the degree of cardiovascular disability so that one patient can be compared with another and so that one patient's status can be monitored over time. The most often used classification system, formulated by the Criteria Committee of the New York Heart Association (table 1), classifies patients according to the degree of symptoms resulting from ordinary or less-than-ordinary activity. Although this widely adopted system has been used to categorize the functional status of patients in numerous clinical studies, no information is available concerning how often physicians would disagree in their definitions of "ordinary" or in their assessments of a patient's functional class, and few data are available regarding the system's correlation with objective measures of exercise tolerance. In 1972, the Canadian Cardiovascular Society proposed more detailed criteria (table 2), which have been adopted by several studies, including the National Heart, Lung and Blood Institute's Coronary Artery Surgery Study. However, the Canadian Cardiovascular Society system has not been evaluated for its reproducibility or for its correlation with objective measures of exercise tolerance.

Because of this uncertainty regarding both the reproducibility and the validity of these two functional classification systems, we developed a new Specific Activity Scale to assess cardiac functional class. In a prospective study in the exercise laboratory of the Peter Bent Brigham Hospital, our simple, easily administered system was significantly more reproducible and significantly more valid than the New York Heart Association criteria and was significantly more valid than the Canadian Cardiovascular Society criteria.

Methods

Development of the Specific Activity Scale

The approximate metabolic costs of a variety of personal care, housework, occupational and recreational activities were estimated based on available data. During a 1-month pretesting phase, we identified questions that referred to activities appropriate for our patient population (table 3). We determined whether specific activities were performed; if the activity was performed, we asked what symptoms were provoked by it; if the activity was not performed, we asked why not. We did not attempt to obtain a detailed cardiovascular history. When performed by a physician, the entire interview normally took less than 2 minutes. A patient was considered able to perform a given number of metabolic equivalents if the appropriate activity could be performed to completion with or without symptoms. Conversely, if the activity was not performed because of symptoms, fear of symptoms, or habit, and if no other activity of approximately equal or higher metabolic cost was performed, the patient was considered unable to attain the given.

From the Cardiovascular Division, and the Division of General Medicine, Brigham and Women's Hospital (Peter Bent Brigham Division) and Harvard Medical School; and the Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts.

Supported in part by contract HRA-230-76-0300 from the National Center for Health Services Research, and by grant 18-P-97517/1-01 from the Health Care Financing Administration.

Dr. Goldman is the recipient of a Teaching and Research Scholarship of the American College of Physicians.

Address for correspondence: Lee Goldman, M.D., Department of Medicine, Brigham and Women's Hospital, 721 Huntington Avenue, Boston, Massachusetts 02115.

Received October 23, 1980; revision accepted March 16, 1981.

Circulation 64, No. 6, 1981.

Metabolic costs for some activities were available only from the Department of Occupational Therapy, Massachusetts Rehabilitation Hospital, 125 Nashua Street, Boston, Massachusetts 02114, "Energy Cost of Activities."
ment of functional class because of an acute hospitalization that resulted in major changes in the therapeutic regimen and a concurrent period of enforced restricted activity.

**Interview Protocol**

To determine the Specific Activity Scale class, each patient was independently interviewed by two investigators who had no knowledge of the patient's functional state. If an interview could not be conducted before the exercise test, the interview was still conducted without knowledge of the results of the other interview or of the exercise test itself. One of the two interviews was always conducted by a physician who had not begun internship. To determine whether a non-physician could use the Specific Activity Scale as well as a physician, the identity of the second interviewer varied. About 55% of the patients had their sec-

---

**Table 1. New York Heart Association Functional Classification**

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Patients with cardiac disease but without resulting limitations of physical activity. Ordinary physical activity does not cause undue fatigue, palpitation, dyspnea, or anginal pain.</td>
</tr>
<tr>
<td>II</td>
<td>Patients with cardiac disease resulting in slight limitation of physical activity. They are comfortable at rest. Ordinary physical activity results in fatigue, palpitation, dyspnea, or anginal pain.</td>
</tr>
<tr>
<td>III</td>
<td>Patients with cardiac disease resulting in marked limitation of physical activity. They are comfortable at rest. Less than ordinary physical activity causes fatigue, palpitation, dyspnea, or anginal pain.</td>
</tr>
<tr>
<td>IV</td>
<td>Patients with cardiac disease resulting in inability to carry on any physical activity without discomfort. Symptoms of cardiac insufficiency or of the anginal syndrome may be present even at rest. If any physical activity is undertaken, discomfort is increased.</td>
</tr>
</tbody>
</table>

---

**Table 2. Canadian Cardiovascular Society Functional Classification**

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Ordinary physical activity does not cause angina, such as walking and climbing stairs. Angina with strenuous or rapid or prolonged exertion at work or recreation.</td>
</tr>
<tr>
<td>II</td>
<td>Slight limitation of ordinary activity. Walking or climbing stairs rapidly, walking uphill, walking or stair climbing after meals, or in cold, or in wind, or under emotional stress, or only during the few hours after awakening. Walking more than 2 blocks on the level and climbing more than one flight of ordinary stairs at a normal pace and in normal conditions.</td>
</tr>
<tr>
<td>III</td>
<td>Marked limitation of ordinary physical activity. Walking one to two blocks on the level and climbing one flight of stairs in normal conditions and at normal pace.</td>
</tr>
<tr>
<td>IV</td>
<td>Inability to carry on any physical activity without discomfort — anginal syndrome may be present at rest.</td>
</tr>
</tbody>
</table>

---

**Table 3. Criteria for Determination of the Specific Activity Scale Functional Class**

<table>
<thead>
<tr>
<th>Any yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Can you walk down a flight of steps without stopping? (4.5-5.2 mets)</td>
<td>Go to #2</td>
</tr>
<tr>
<td>2. Can you carry anything up a flight of 8 steps without stopping (5-5.5 mets) or can you:</td>
<td>Go to</td>
</tr>
<tr>
<td>(a) have sexual intercourse without stopping (5-5.5 mets)</td>
<td>#3</td>
</tr>
<tr>
<td>(b) garden, rake, weed (5.6 mets)</td>
<td></td>
</tr>
<tr>
<td>(c) roller skate, dance foxtrot (5-6 mets)</td>
<td></td>
</tr>
<tr>
<td>(d) walk at a 4 miles-per-hour rate on level ground (5-6 mets)</td>
<td></td>
</tr>
<tr>
<td>3. Can you carry at least 24 pounds up 8 steps (10 mets) or can you:</td>
<td>Class</td>
</tr>
<tr>
<td>(a) carry objects that are at least 80 pounds (8 mets)</td>
<td>I</td>
</tr>
<tr>
<td>(b) do outdoor work — shovel snow, spade soil (7 mets)</td>
<td></td>
</tr>
<tr>
<td>(c) do recreational activities such as skiing, basketball, touch football, squash, handball (7-10 mets)</td>
<td></td>
</tr>
<tr>
<td>(d) jog/walk 5 miles per hour (9 mets)</td>
<td></td>
</tr>
<tr>
<td>4. Can you shower without stopping (3.6-4.2 mets) or can you:</td>
<td>Class</td>
</tr>
<tr>
<td>(a) strip and make bed (3.9-5 mets)</td>
<td>III</td>
</tr>
<tr>
<td>(b) mop floors (4.2 mets)</td>
<td></td>
</tr>
<tr>
<td>(c) hang washed clothes (4.4 mets)</td>
<td></td>
</tr>
<tr>
<td>(d) clean windows (3.7 mets)</td>
<td></td>
</tr>
<tr>
<td>(e) walk 2.5 miles per hour (3-3.5 mets)</td>
<td></td>
</tr>
<tr>
<td>(f) bowl (3-4.4 mets)</td>
<td></td>
</tr>
<tr>
<td>(g) play golf (walk and carry clubs) (4.5 mets)</td>
<td></td>
</tr>
<tr>
<td>(h) push power lawn mower (4 mets)</td>
<td></td>
</tr>
<tr>
<td>5. Can you dress without stopping because of symptoms? (2-2.3 mets)</td>
<td>Class</td>
</tr>
<tr>
<td>III</td>
<td>IV</td>
</tr>
</tbody>
</table>

*These questions were selected from a much larger group of questions because of their relevance to our patient population.

Abbreviation: mets = metabolic equivalents of activity.
ond interview performed by a medical librarian who had never interviewed a patient and had no background in cardiovascular diseases. The remaining 45% of patients had their second interview performed by a board-certified cardiologist. The second interviewer could not be randomly selected; the research assistant was a half-time employee who interviewed all patients who had exercise tests during her working hours.

In addition to asking the questions directly related to the Specific Activity Scale, the interviewers also estimated the patient's Canadian Cardiovascular Society functional class. For the purposes of this study, symptoms other than angina were considered limitations that would place a patient into the Canadian Cardiovascular Society functional class that best estimated the degree of physical disability.

The patient's New York Heart Association functional class was estimated independently by two physicians. One estimate was made before the exercise test by the board-certified or board-eligible cardiologist who was to conduct the test. This physician did not know the estimates made by the coauthors, but was free to interview the patient to his own satisfaction; the full criteria of the New York Heart Association were posted in the exercise test laboratory for the physician to refer to. Five cardiologists conducted exercise tests during this study. After the exercise test was completed, the physician who had referred the patient to the exercise laboratory was contacted. This physician was asked to estimate the patient's New York Heart Association functional classification before being told the results of the exercise test. We offered to read the specific criteria of the New York Heart Association to the referring physician so there could be no confusion as to what the criteria stated.

Exercise Protocol

All but three patients had standard Bruce protocol treadmill exercise tests as routinely performed at Peter Bent Brigham Hospital. The three other patients had modified protocols using the same equipment. All patients were encouraged to hold onto the handrails during the exercise test. The cardiologist performing the exercise test encouraged patients to reach 85% of their maximal predicted heart rate, but he would stop the test if the patient could go no further or refused to go further because of symptoms or developed asymptomatic electrocardiographic changes, such as worrisome ventricular arrhythmias or severe ST-segment depression. The duration of exercise was timed electronically. True functional class was defined as class I if the patient performed more than 6 minutes of the Bruce protocol or more than 9.5 minutes of the modified protocol; class II if the patient performed more than 3 minutes but not more than 6 minutes of the Bruce protocol, or more than 6.5 minutes but not more than 9.5 minutes of the modified protocol; class III if the patient performed more than 1 minute but not more than 3 minutes of the Bruce protocol or more than 2 minutes but not more than 6.5 minutes of the modified protocol; and class IV if the patient did not perform more than 1 minute of the Bruce protocol or more than 2 minutes of the modified protocol.

Data Analysis

For both the New York Heart Association and the Canadian Cardiovascular Society Systems, patients who were estimated to be asymptomatic and who could perform the activities consistent with class I status were considered as class I rather than as the unofficial, but occasionally used, class 0. We defined reproducibility as interobserver agreement or disagreement on the assessment of functional class and validity as the agreement or disagreement of the functional classification estimates with the patient's true functional classification as determined by the objective exercise test. Based on our pretesting, we hoped to show a 20% increase in reproducibility and a 20% increase in validity using the Specific Activity Scale. To detect such changes with a power of 80% (i.e., a β error of 20%) and an α error of 5%, reproducibility testing was estimated to require 75 patients and validity testing was estimated to require 40 patients (80 estimates). We therefore decided to enroll 75 patients and then terminate the study and analyze the data. Of the first 78 eligible patients, one patient was eliminated because logistic problems prohibited a second Specific Activity Scale interview and two patients were eliminated because the referring physician could not be contacted before he had already obtained the results of the exercise test; the remaining 75 patients constituted the study population.

The overall proportion of times that the functional class estimates were reproducible (percentage of patients assigned to the same functional class by both observers using the same system) and valid (percentage of functional class estimates that agreed with the treadmill performance class) for each of the three functional classification systems was analyzed. Simultaneous comparisons of the performances of all three classification systems were analyzed using a chi-square statistic with 2 degrees of freedom. If the overall chi-square statistic was significant, two-by-two comparisons of one system with another were made by the Mantel-Haenszel matched analysis because the

<table>
<thead>
<tr>
<th>Table 4. Summary of Criteria for Specific Activity Scale Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class I:</strong> Patient can perform to completion any activity requiring ≥ 7 metabolic equivalents</td>
</tr>
<tr>
<td><strong>Class II:</strong> Patient can perform to completion any activity requiring ≥ 5 metabolic equivalents but cannot or does not perform to completion activities requiring ≥ 7 metabolic equivalents</td>
</tr>
<tr>
<td><strong>Class III:</strong> Patient can perform to completion any activity requiring ≥ 2 metabolic equivalents but cannot or does not perform to completion any activities requiring ≥ 5 metabolic equivalents</td>
</tr>
<tr>
<td><strong>Class IV:</strong> Patient cannot or does not perform to completion activities requiring ≥ 2 metabolic equivalents.</td>
</tr>
</tbody>
</table>
performances of all three classification systems were
measured on a common set of patients and, in some
instances, by the same investigator.14 Statistical
significance was defined as \( p \leq 0.05 \).

To take into account the degree of reproducibility
and validity that could be expected to occur by chance,
kappa statistic statistics were also generated.16–18 The kappa
statistic, which is designed to compare an observed
reproducibility or validity to what could be expected
by chance, is scaled to range from \(-1\) to \(+1\). A value
of \(+1\) refers to perfect agreement, a value of 0 refers
to chance agreement, and a value of \(-1\) refers to com-
plete disagreement. To adjust for the degree of ob-
served disagreements, we used weighted kappa statis-
tics in which disagreement by two functional classes
was considered twice as bad, and disagreement by
three functional classes was considered four times as
bad, as disagreement by one functional class.18 The
values of the weighted kappa statistics were compared
with the degree of agreement that could be expected
by chance.18

To correlate functional class estimates with the ac-
tual duration of treadmill exercise in seconds (to be
distinguished from true functional classification based
on the exercise duration), we computed Spearman
correlation coefficients. The Spearman correlation co-
efficient \( \rho \) is a nonparametric coefficient based on
the relative rankings of the variables, not on their absolute
values. For the three patients who had modified tread-
mill protocols, the duration of treadmill exercise in
seconds was converted to the number of seconds of
the standard Bruce protocol that was estimated to require
the same energy expenditure.12, 18, 18

In four of the 150 New York Heart Association es-
timates and three of the 150 Canadian Cardiovas-
cular Society estimates, an interviewer could not de-
cide between two functional classes (e.g., class I vs
class II). These patients were not excluded from the
study, because to do so might bias our results. Thus, in
our reproducibility and validity calculations, the choice
among the functional class estimates was deter-
mined by a coin flip (class I or class II). In the calcula-
tion of Spearman correlation coefficients, however,
the functional class estimate was considered inter-
mediate (class 1.5).

Results

Patient Population

For 56 patients (75%) the primary indication for the
exercise test was the evaluation of chest pain. Twelve
patients (16%) performed exercise tests for evaluation
of other symptoms, and seven asymptomatic patients
(9%) performed exercise tests for assessment of func-
tional status after their hospital discharge after val-
vular surgery, coronary bypass surgery or myocardial
infarction.

Exercise was terminated because of fatigue in 26
patients, chest pain in 21, leg complaints in nine, other
symptoms in 14 and asymptomatic electrocardio-
graphic findings in five. Forty-nine patients (65%) had positive exercise tests or achieved at least 85% of
their maximal predicted heart rate and had non-
diagnostic or negative exercise ECGs. Based on the
duration of exercise associated with our definition
of true functional class, 38 patients had exercise tests
that placed them in class I, 23 in class II, 13 in class
III and one patient in class IV.

Reproducibility Testing

The Specific Activity Scale and the Canadian Car-
diovascular Society system both had a reproducibility
of 73%, which was significantly higher than the 56%
reproducibility of the New York Heart Association
classification system (table 5). The reproducibilities
of the Canadian Cardiovascular Society system and of
the Specific Activity Scale did not depend on which of
the coauthors performed the patient interview. The
relative reproducibilities of all three classification
systems were independent of the patient's presenting
complaint, the patient’s ability to achieve 85% of the
maximal predicted heart rate and the results or dura-
tion of the exercise test.

The reproducibility of the New York Heart As-
sociation estimates was similar to that found for unreli-
able radiographic criteria (table 6). Conversely, the
better reproducibilities of the Specific Activity Scale
and of the Canadian Cardiovascular Society system
were comparable to those for common physical ex-
amination data.

Validity Testing

The New York Heart Association estimates agreed
with exercise treadmill performance only 51% of the
time; this validity rate was not quite significantly low-
er than the 59% validity rate for the Canadian Car-
diovascular Society system, but was significantly low-

<table>
<thead>
<tr>
<th>Table 5. Reproducibility Testing (n = 75)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. and % of times that two observers:</td>
</tr>
<tr>
<td>A. agreed on the assessment of an individual patient</td>
</tr>
<tr>
<td>B. differed by one functional class</td>
</tr>
<tr>
<td>C. differed by two functional classes</td>
</tr>
<tr>
<td>D. differed by three functional classes</td>
</tr>
</tbody>
</table>

* \( \chi^2 \) (two degrees of freedom) = 6.85, \( p < 0.05 \). By matched analysis, results with both the CCS and the SAS criteria were significantly more reproducible than results with the NYHA criteria (\( p < 0.05 \)).

† All three systems were significantly more reproducible than could be expected by chance. The CCS system and the SAS had slightly different kappa statistics because the CCS system had one estimate that differed by three functional classes.

Abbreviations: NYHA = New York Heart Association; CCS = Canadian Cardiovascular Society; SAS = Specific Activity Scale.
er than the 68% validity rate for the Specific Activity Scale (table 7). Also, in the matched-pair analysis of the 37 instances in which the true functional class agreed with a coauthor's Canadian Cardiovascular Society estimate or with his or her own Specific Activity Scale estimate, but not with both estimates, the Specific Activity Scale estimate was significantly more likely to predict treadmill performance (25 matched pairs vs 12 matched pairs, \( \chi^2 = 4.57, p = 0.033 \)). The Specific Activity Scale had the highest weighted kappa statistic (0.54), followed by the Canadian Cardiovascular Society system (0.47) and the New York Heart Association system (0.33). Similarly, the Specific Activity Scale had the strongest correlation \( (r = -0.66) \) (table 8) with the duration of treadmill exercise in seconds, followed by the Canadian Cardiovascular Society system \( (r = -0.64) \) and the New York Heart Association system \( (r = -0.54) \).

All three systems were equally likely to overestimate treadmill performance, but both the New York Heart Association (41 estimates) and the Canadian Cardiovascular Society (35 estimates) systems were significantly more likely to underestimate treadmill performance than was the Specific Activity Scale (20 estimates; both \( p < 0.01 \)). The mean Specific Activity Scale estimate for the entire series (1.65) was very similar to the patients' true mean functional class (1.69). The mean Canadian Cardiovascular Society system estimate (1.81) was actually farther from the true mean functional class than was the mean New York Heart Association estimate (1.77). The tendency of both the Canadian Cardiovascular Society system and the New York Heart Association system to underestimate performance is consistent with the systems' criteria that classify patients based on the activities that are associated with symptoms. Conversely, the Specific Activity Scale classified patients

<table>
<thead>
<tr>
<th>Table 6. Comparative Interobserver Reproducibilities of Various Radiologic Signs, Physical Examination Findings, and Functional Classification Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kappa</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

The kappa statistic values for the three functional classification systems are weighted (see Methods) to adjust for the degree of disagreement, whereas the other kappa statistic values reported here are unweighted because they measured on a yes-no scale. The use of unweighted kappa statistics for the three functional classification systems would make small changes in the reported values and would not alter the clustering shown above.

<table>
<thead>
<tr>
<th>Table 7. Validity Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No. and % of the 150 estimates that:</strong></td>
</tr>
<tr>
<td><strong>A. agreed with treadmill performance</strong></td>
</tr>
<tr>
<td><strong>B. disagreed with performance by one class</strong></td>
</tr>
<tr>
<td><strong>C. disagreed with performance by two classes</strong></td>
</tr>
<tr>
<td><strong>D. disagreed with performance by three classes</strong></td>
</tr>
</tbody>
</table>

| Weighted kappa statistic | 0.33† | 0.47† | 0.54† |

*\( \chi^2 (two \ degrees \ of \ freedom) = 8.65, p = 0.013 \). By matched analysis, the proportion of correct estimates made by the SAS was significantly higher than the proportions made by either the NYHA criteria \( (p < 0.001) \) or the CCS criteria \( (p < 0.05) \). The proportion of correct estimates made by the CCS system was not quite statistically significantly higher than the proportion of correct estimates made by the NYHA criteria \( (\chi^2 = 3.27, p = 0.07) \).

†All three kappa statistics were significantly higher than could be expected by chance.

Abbreviations: NYHA = New York Heart Association; CCS = Canadian Cardiovascular Society; SAS = Specific Activity Scale.

### Table 8: Comparison of Classification Systems with Treadmill Performance: Overestimates and Underestimates

<table>
<thead>
<tr>
<th>Treadmill Performance (n = 75)</th>
<th>New York Heart Association Criteria (n = 150)</th>
<th>Canadian Cardiovascular Society Criteria (n = 150)</th>
<th>Specific Activity Scale (n = 150)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Functional class</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class I</td>
<td>38 (51%)</td>
<td>70 (47%)</td>
<td>68 (45%)</td>
</tr>
<tr>
<td>Class II</td>
<td>23 (31%)</td>
<td>48 (32%)</td>
<td>45 (30%)</td>
</tr>
<tr>
<td>Class III</td>
<td>13 (17%)</td>
<td>29 (19%)</td>
<td>35 (23%)</td>
</tr>
<tr>
<td>Class IV</td>
<td>1 (1%)</td>
<td>3 (2%)</td>
<td>2 (1%)</td>
</tr>
<tr>
<td><strong>B. Mean functional class</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>C. Correlation (r) with duration of treadmill exercise in seconds</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.54</td>
<td>-0.64</td>
<td>-0.66</td>
</tr>
</tbody>
</table>
by the activities that could be performed to completion regardless of symptoms.

When using the New York Heart Association classification system, the validity rates of the cardiologists performing the exercise tests were identical to those of the referring physicians. Similarly, the two physician coauthors and the research assistant were all equally valid in their use of the Canadian Cardiovascular Society criteria and of the Specific Activity Scale. There was no evidence that the examiners' predictions of exercise tolerance became more or less valid during the study.

The relative validities of the three functional classification systems did not depend on the reasons for which the exercise test was performed, the percent of predicted heart rate achieved, why the patient ceased exercising, or the degree of electrocardiographic changes precipitated by the exercise test. The Specific Activity Scale was as valid in patients who did not perform activities because of fear or disinterest as in patients who ceased activities because of cardiac symptoms. No patient subsets could be identified in which either the New York Heart Association classification system or the Canadian Cardiovascular Society system performed significantly better than the Specific Activity Scale, but the increased validity of the Specific Activity Scale was most notable in the 23 patients with true class II exercise tolerance: 32 of 46 (70%) Specific Activity Scale estimates were valid, compared with 19 of 46 (41%) Canadian Cardiovascular Society estimates and only 16 of 46 (35%) New York Heart Association classification estimates (both p < 0.01). Because only one of our 75 patients had class IV treadmill performance, we cannot compare the relative validities of the three classification systems for class IV patients.

Discussion

Transforming clinical data into useful clinical scales is one of the most challenging problems in clinical research. For a measurement scale to be valuable, it must be reproducible from one observer to another, and it should be reasonably well correlated with a relatively objective standard. Thus, the activity of daily living scale and the sickness impact profile represent measurement scales that have successfully transformed clinical data into health status measures.

Our standard, the exercise treadmill test, provides an estimate of oxygen consumption but is unlikely to be identical to true oxygen consumption. However, in the study by Franciosa and colleagues, measured oxygen consumption and exercise duration were highly correlated. Thus, treadmill exercise performance appears to be a reasonably valid standard for assessing functional capacity and classification.

As noted by Selzer and Cohn, the widely accepted criteria of the New York Heart Association have never been subjected to the systematic reproducibility and validity testing that would normally be required of any clinical assay or diagnostic test. In our series, the New York Heart Association classification system's reproducibility of only 56% was significantly lower than the reproducibility of the other two systems. We cannot be sure that this low reproducibility of the New York Heart Association system was not biased by the fact that it was administered by the referring physician and the exercise laboratory physician, whereas the authors administered the other two systems. However, because the New York Heart Association criteria were posted in the exercise laboratory and because we offered to read the criteria to all referring physicians, this bias should have been minimized. Further, our reproducibility and validity rates did not increase over the course of the study, suggesting that further experience with the Canadian Cardiovascular Society criteria or with the Specific Activity Scale was not important. Thus, we believe that the higher reproducibility of the Canadian Cardiovascular Society system was related to the greater details included in its definition, but that the further detail in our Specific Activity Scale could not improve upon the reproducibility attained by the Canadian Cardiovascular Society criteria.

Our findings regarding the poor validity of the New York Heart Association system were consistent with the findings of previous studies. In one series, the New York Heart Association functional classification as estimated by the consensus of two physicians predicted exercise tolerance in 32 of 43 patients (74%). In another series, however, the New York Heart Association functional classification as estimated by a single physician predicted exercise tolerance in only 16 of 44 patients (36%) with congestive heart failure. The pooled results of these two studies would yield a validity of 55% (48 of 87 patients), which is similar to our validity rate of 51%. By comparison, the Specific Activity Scale was significantly more valid than either the New York Heart Association system or the Canadian Cardiovascular Society system. This increased validity of the Specific Activity Scale was not surprising because we asked standardized questions related to the individual activities that a patient could perform and then compared the responses with a patient's actual ability to exercise. Using a similar approach, Taylor and colleagues showed that the sum of energy expenditures required for the many leisure activities an adult performs has a multiple correlation coefficient (R) of 0.45 with treadmill exercise capacity. The study's higher correlation coefficient in our study (r = 0.66) may indicate that the standardized questionnaire approach is more accurate in patients with symptoms than it is in asymptomatic adults.

Our study design allowed the New York Heart Association criteria to be applied by experienced physicians, including the patient's own doctor. Thus, we believe that the 56% reproducibility and 51% validity of the New York Heart Association criteria indicate that this system is not adequate to evaluate a patient's response to therapy or to compare one patient with another. The Canadian Cardiovascular Society system and the Specific Activity Scale, as applied by interviewers who did not know the patient and who were often less experienced than the physicians who
estimated New York Heart Association classifications, had significantly higher reproducibilities, but only the Specific Activity Scale had a significantly higher validity.

Other authors have proposed classification systems that were not tested in our investigation. Feinstein proposed a clinical taxonomy that used a detailed questioning of a patient’s occupation, customary activities and sporadic activities. This system is advantageous for the in-depth assessment of an individual patient’s change over time, but it was not designed to compare one patient with another. The anginal severity system of Peduzzi and Hultgren, which was proposed in response to the perceived shortcomings of the New York Heart Association criteria, integrated the amount of therapeutic intervention and the severity of the patient’s angina. However, Peduzzi and Hultgren’s questions regarding level of activity are of unknown reproducibility, and the system as a whole may be too cumbersome to use outside of the clinical research setting.

The popularity of the New York Heart Association classification system is at least partly based on its simplicity. Any system that might replace this standard should be more accurate without being appreciably more complex. We believe that the Specific Activity Scale, which is based on the metabolic requirements of some of the same types of activities that are often included in present routine assessments, represents an acceptable tradeoff between precision and complexity. The Specific Activity Scale was equally valid when administered by a nonphysician, and its questions were no more time consuming than those used by the participating cardiologists to estimate New York Heart Association classifications.

The 73% reproducibility and 68% validity of the Specific Activity Scale probably indicate that no cardiovascular functional classification system based on the medical history can be perfectly reproducible or valid. Even with the Specific Activity Scale format, there was some variation in the exact way that questions were asked; some patients gave patently contradictory answers to the same questions, and some patients’ reproducible responses were not predictive of their objective exercise tolerance.

There are no data regarding the prognostic value of a patient’s Specific Activity Scale class, and any new system may perform better for its originators than it will for others. Nevertheless, the Specific Activity Scale appears to have distinct advantages over the Canadian Cardiovascular Society system, which in turn was better than the New York Heart Association system. If further trials at other institutions substantiate our experience, the Specific Activity Scale or its future modifications might appropriately replace these previous systems in both clinical practice and clinical investigation.

Acknowledgment

The authors thank Drs. Peter Stone, Joshua Wynne, John Somberg, James Marsh, Jonas Galper and Lawrence DeBoer, and Violet Desingco for assistance and Drs. Alvan Feinstein, Herbert Sherman, James Muller, and Peter Cohn for reviewing the manuscript.

References

25. Feinstein AR: Clinical biostatistics. XLI. Hard science, soft data, and the challenges of choosing clinical variables in
SUMMARY  
In a consecutive autopsy series of 17 patients with tetralogy of Fallot (TOF), including five patients with associated pulmonary atresia (TOF + PA), radiopaque material was injected into the pulmonary and collateral arteries and the lungs were examined using quantitative techniques to assess pulmonary vascular and alveolar structure and growth.

Three types of systemic collateral arteries (SCA) were distinguished by their origin, each showing a characteristic type of “anastomosis” with a pulmonary artery: Bronchial artery collaterals arise from bronchial arteries (so judged by origin and distribution) and anastomose with pulmonary arteries inside the lung. Direct aortic collaterals arise from the descending aorta, enter the lung at the hilum and supply a lobe, or at least a segment. Inside the lung, structure and distribution are those of a pulmonary artery; outside the lung, those of a systemic artery. These vessels sometimes anastomose inside the lung with pulmonary arteries from neighboring segments of lobes. Indirect aortic collaterals arise from major branches of the aorta other than bronchial arteries (e.g. internal mammary, subclavian), and usually anastomose with the central pulmonary arteries outside the lung.

All three types of SCAs were found in TOF + PA, but only bronchial artery collaterals were found in TOF. SCAs narrowed at the site of anastomosis with a pulmonary artery; this apparently protected the peripheral intraacinar arteries from high flow and pressure and prevented changes of excessive muscularity and intimal hyperplasia.

In most patients, the intraacinar arteries were smaller than normal; in patients who had large, surgically created shunts of long duration they were fewer in number, of greater muscularity and with severe occlusive changes of intimal hyperplasia. The number of alveoli was reduced in almost all patients, but because alveolar size was increased, only a few patients had a small lung volume.

Early surgical correction of TOF may prevent the impairment in alveolar and vascular growth described here. In TOF + PA patients, selective arteriography of all large aortic collaterals should be performed preoperatively to determine their intrapulmonary distribution and relationship with central pulmonary arteries. This will help in selecting the surgical procedure and identify collaterals that can be safely ligated.

RECENT ADVANCES in cardiac surgery allow correction of tetralogy of Fallot (TOF) in infancy, which may be desirable for normal structural development of the lung. However, no study has been carried out to quantify the extent of impaired lung and vascular growth from infancy to adulthood in patients with TOF, including those with previous palliative procedures. In patients with TOF and associated pulmonary atresia (PA), it is difficult to determine the type and timing of surgical interventions that will optimally increase pulmonary blood flow. This is because of the difficulty in assessing preoperatively the relative importance of the central pulmonary and systemic collateral arteries (SCAs) and the degree of structural abnormality in the peripheral pulmonary vascular bed.

In the present study, we used arteriography and morphometric techniques to quantitatively assess lung growth and vascular development in a consecutive postmortem series of patients with TOF, including some with and some without PA. For each type of SCA distinguished by site of origin, we identified a pattern of anastomosis with pulmonary arteries.

L Goldman, B Hashimoto, E F Cook and A Loscalzo

Circulation. 1981;64:1227-1234
doi: 10.1161/01.CIR.64.6.1227

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 1981 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/64/6/1227

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation is online at:
http://circ.ahajournals.org/subscriptions/