Reliability, validity, and responsiveness of the six-minute walk test in patients with heart failure

Catherine Demers, MD, MSc, FRCP,C, Robert S. McKelvie, MD, PhD, FRCP,C, Abdissa Negassa, PhD,b and Salim Yusuf, MBBS, DPhil, FRCP,C,a for the RESOLVD Pilot Study Investigators Hamilton, Ontario, Canada, and Bronx, NY

Background Our purpose was to evaluate the reliability, validity, and responsiveness of the 6-minute walk test (6MWT) in patients with heart failure (HF) enrolled in the Randomized Evaluation of Strategies for Left Ventricular Dysfunction (RESOLVD) pilot study.

Methods A total of 768 patients was enrolled in a multicenter randomized clinical trial evaluating the effect of candesartan, enalapril, and metoprolol on left ventricular ejection fraction (LVEF), 6MWT distance, neurohormones, and quality of life. The 6MWT was performed once at screening and twice at baseline, 18 weeks, and 43 weeks by a standardized method.

Results Test-retest reliability at baseline (intraclass correlation coefficient [ICC] = 0.90), 18 weeks (ICC = 0.88), and 43 weeks (ICC = 0.91) was very good. Baseline 6MWT distance was weakly inversely correlated to the quality-of-life cumulative score (r = –0.26, P = .0001) and moderately inversely correlated to the New York Heart Association functional classification (NYHA-FC) (r = –0.43, P = .001). In the RESOLVD study, the 6MWT was not responsive to change when effect sizes and standardized response means were used. Disease-specific quality of life was responsive to change in patients treated with candesartan and enalapril and NYHA-FC was responsive to change in the candesartan and enalapril combination and for enalapril alone with small effect sizes. The 6MWT, NYHA-FC, and quality of life were not responsive to change during the metoprolol or placebo phase.

Conclusions The 6MWT is highly reproducible in patients with symptoms of HF. It is somewhat correlated to NYHA-FC and quality of life. Overall, quality of life was most responsive to change, whereas 6MWT and NYHA-FC were comparable but less responsive to change in the RESOLVD study. (Am Heart J 2001;142:698-703.)
markers of functional capacity and the responsiveness to change of the 6MWT, disease-specific quality of life, and New York Heart Association functional class (NYHA-FC) in patients with HF in the setting of the RESOLVD study, a moderate size multicenter randomized controlled clinical trial.

## Methods

### Patient population

Patients with stable HF, NYHA-FC II-IV symptoms, 6MWT distance <500 m, and left ventricular ejection fraction (LVEF) <0.40 were eligible for enrollment into the RESOLVD pilot study. A detailed description of the exclusion criteria and other aspects of the study design has been previously published.\(^{14}\)

### Six-minute walk test

A standardized method was used to administer the 6MWT across all 60 participating centers with a course length ≥20 m. Patients were instructed to walk at their own pace while attempting to cover as much distance as possible during the allotted time. The test was supervised and the time was called out every 2 minutes. Standard encouragement at 30-second intervals was provided in the form of “You are doing well” or “Keep up the good work.” During the test, patients were allowed to rest or stop and then continue as soon as they could resume the walk. At the completion of 6 minutes, the patient was told to “stop” and the distance covered was recorded. The 6MWT was administered once for screening at enrollment into the study. The test was performed in duplicate at baseline (3 weeks after the screening test), 18, and 43 weeks within a 7-day period. The 2 measurements were then averaged for each time point.

### Quality-of-life questionnaire

The Minnesota Living with Heart Failure questionnaire\(^{16}\) was used to assess disease-specific quality of life. This instrument consists of 21 items evaluating the patient’s perceptions concerning the effects of HF on physical, psychologic, and socioeconomic domains. Reproducibility and validity of this questionnaire have been previously assessed.\(^{16,17}\) The questionnaire was self-administered and was scored in a standard fashion across centers.

### Other outcome measurements

In all centers, measurements were performed at baseline and at 18 and 43 weeks. The NYHA-FC and LVEF, measured by radionuclide angiography, were evaluated as described previously.\(^{14}\)

### Statistical analysis

#### Reliability

The test-retest reliability or the ability to measure the same results on repeated tests was measured with use of the intraclass correlation coefficient (ICC), calculated with the variance components of a random effects 2-way analysis of variance model. The overall test reliability was assessed with the variance components derived from analysis of variance considering the 2 measurements on a given visit as random replicates. A mixed-effects analysis of variance was also used considering the intervention/treatment or learning effect within visits, visits considered as a fixed effect. An ICC >0.75 was considered adequate and >0.9 was considered excellent.\(^{18}\)

#### Construct validity

Construct validity describes the relationship between an attribute under evaluation and other attributes.\(^{19,20}\) The relationship between these attributes is then examined and if both are measuring something similar then the theoretical construct, for example, measurement of functional capacity in patients with heart failure, is correct. Construct validity was assessed by correlating the baseline 6MWT distance with baseline NYHA-FC and quality of life. The Pearson correlation coefficient was used to examine the relationship between the 6MWT distance and quality of life. The Spearman correlation coefficient was used to evaluate the relationship between the 6MWT and NYHA-FC.

#### Responsiveness

Responsiveness or sensitivity to change assesses the ability of a diagnostic test to measure change over time or the ability to measure the main effect of treatment.\(^{20}\) Traditionally, responsiveness has been assessed with the t test or analysis of variance. Other statistical methods have been proposed to evaluate responsiveness.\(^{21-24}\) These alternative methods may help us better define clinically important changes when data from clinical trials are used. In this article, effect sizes and standardized response means were used to assess responsiveness. (1) Effect sizes were calculated by dividing the observed change by the SD of the baseline score.\(^{24}\) An effect size of >0.8 was considered large, 0.5 to 0.8 was moderate, and 0.2 to 0.5 was small.\(^{25}\) (2) The standardized response mean is defined as the observed change divided by the SD of the difference scores.\(^{24}\) This statistic is very similar to the paired t test but avoids the use of the SEM and thus is less influenced by the sample size. It reflects the signal-to-noise ratio probably better than the effect size. A higher standardized response mean indicates greater sensitivity to change and a standardized response mean <0.5 is considered to be insensitive to change.\(^{26}\) CIs were calculated by the method described by Liang et al.\(^{27}\) Effect sizes and standardized response means provide us with the ability to compare responsiveness with change of different measurements of functional capacity and between different drug treatments. 6MWT responsiveness was evaluated between baseline and 18 weeks and 18 to 43 weeks, and responsiveness was also assessed for the NYHA-FC and quality of life. The screening 6MWT was excluded from analysis to avoid regression to the mean. These results are derived from post hoc analyses of the RESOLVD database. Results were considered statistically significant for a P value ≤0.05. Statistical analysis was performed with SAS software (SAS Institute, Cary, NC).

### Results

#### Baseline characteristics

Seven hundred sixty-eight patients were randomized to candesartan, candesartan plus enalapril, or enalapril alone and followed up for 43 weeks. At 18 weeks, 426 of the 768 patients were further randomized to metoprolol or placebo while maintaining first-stage randomization medication. Eighty-three percent of patients were male, mean age 63 ± 11 years, and 1% were in NYHA-FC I, 65% were in NYHA-FC II, 32% were in NYHA-FC III, and 2% were in NYHA-FC IV. Mean LVEF
was 0.27 ± 0.10 and ischemic heart disease was the etiology of HF in 71% of patients. Ninety-five percent of patients were receiving angiotensin-converting enzyme inhibitors at baseline, 84% were on diuretics, 15% were on β-receptor antagonists, and 69% were on digoxin.

**Six-minute walk test**

The mean distance walked was 381 ± 84 (mean ± SD) m at baseline, 387 ± 94 m at 18 weeks, and 387 ± 103 m at 43 weeks. The mean difference between the first and second walk distance at each measurement time was 57 m. The average number of days between the 2 measurements was 2.2 ± 1.5 at baseline, 3.7 ± 3.0 at 18 weeks, and 3.1 ± 3.5 at 43 weeks.

**Reliability.** Test-retest reliabilities were very good to excellent at all measurement times with ICCs of 0.85 (baseline 0.90, 18 weeks 0.88, 43 weeks 0.91). Overall reliability, when all 6MWTs performed during the study by a random effects model were considered, was good at 0.81. With a mixed-effects model, taking into account the learning or treatment effect, overall reliability was adequate at 0.80.

**Construct validity.** The baseline 6MWT distance was weakly inversely correlated with disease-specific quality of life (r = -0.26, P = .0001) and moderately inversely correlated with NYHA-FC (r = -0.43, P = .001).

**Responsiveness.** The effect sizes and standardized response means for the 6MWT were small, but the standardized response means were statistically significant for patients receiving candesartan 16 mg and enalapril (Table I). There was no significant change of the 6MWT compared with baseline in patients randomized to metoprolol or placebo with all measures of responsiveness (Table II).

### Table I. Responsiveness of the 6MWT, quality of life, and NYHA-FC from baseline to 18 weeks

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Test</th>
<th>4 mg (n = 105)</th>
<th>8 mg (n = 103)</th>
<th>16 mg (n = 103)</th>
<th>4 mg + 20 mg (n = 159)</th>
<th>8 mg + 20 mg (n = 158)</th>
<th>20 mg (n = 106)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect size*</td>
<td>6MWT</td>
<td>0.07</td>
<td>0.07</td>
<td>0.11</td>
<td>0.02</td>
<td>-0.005</td>
<td>0.16</td>
</tr>
<tr>
<td>QOL</td>
<td>-0.19</td>
<td>-0.21</td>
<td>-0.22</td>
<td>-0.12</td>
<td>-0.12</td>
<td>-0.14</td>
<td>-0.14</td>
</tr>
<tr>
<td>NYHA-FC</td>
<td>-0.08</td>
<td>-0.07</td>
<td>-0.13</td>
<td>-0.14</td>
<td>-0.18</td>
<td>-0.26</td>
<td></td>
</tr>
<tr>
<td>Standardized response mean†</td>
<td>6MWT</td>
<td>0.092 ([-0.139-0.324])</td>
<td>0.108 ([-0.094-0.310])</td>
<td>0.198 ([0.018-0.378])</td>
<td>0.042 ([0.114-0.198])</td>
<td>-0.009 ([0.176-0.158])</td>
<td>0.269 ([0.103-0.435])</td>
</tr>
<tr>
<td>QOL</td>
<td>-0.232 ([-0.418-0.046])</td>
<td>-0.266 ([-0.486-0.046])</td>
<td>-0.253 ([-0.459-0.047])</td>
<td>-0.157 ([-0.300-0.014])</td>
<td>-0.157 ([-0.316-0.002])</td>
<td>-0.022 ([-0.395-0.045])</td>
<td></td>
</tr>
<tr>
<td>NYHA-FC</td>
<td>-0.13 ([-0.282-0.022])</td>
<td>-0.089 ([-0.276-0.098])</td>
<td>-0.126 ([-0.344-0.092])</td>
<td>-0.157 ([-0.304-0.010])</td>
<td>-0.194 ([-0.349-0.039])</td>
<td>-0.342 ([-0.493-0.191])</td>
<td></td>
</tr>
</tbody>
</table>

QOL, Quality of life.

*Effect size = Observed change/SD of the baseline test.
†Standardized response mean = Observed change/SD of the difference scores (95% CI).

### Table II. Responsiveness of the 6MWT, quality of life, and NYHA-FC from 18 to 43 weeks

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Test</th>
<th>Metoprolol (n = 199)</th>
<th>Placebo (n = 193)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect size*</td>
<td>6MWT</td>
<td>-0.06</td>
<td>-0.05</td>
</tr>
<tr>
<td>QOL</td>
<td>0.099</td>
<td>0.125</td>
<td></td>
</tr>
<tr>
<td>NYHA-FC</td>
<td>-0.096</td>
<td>-0.039</td>
<td></td>
</tr>
<tr>
<td>Standardized response mean†</td>
<td>6MWT</td>
<td>-0.077 ([-0.220-0.066])</td>
<td>-0.071 ([-0.203-0.061])</td>
</tr>
<tr>
<td>QOL</td>
<td>0.147 (0.020-0.268)</td>
<td>0.161 (0.020-0.300)</td>
<td></td>
</tr>
<tr>
<td>NYHA-FC</td>
<td>-0.1 ([-0.24-0.041])</td>
<td>-0.04 (-0.182-0.100)</td>
<td></td>
</tr>
</tbody>
</table>

QOL, Quality of life.

*Effect size = Observed change/SD of the test at 18 weeks.
†Standardized response mean = Observed change/SD of the difference scores (95% CI).
Quality of life was improved in both the metoprolol and placebo groups (Table II) with small effect sizes and standardized response means. The effect sizes and standardized response means were small for the NYHA-FC compared with baseline (Table II).

From baseline to 18 weeks, effect sizes and standardized response means were the largest for quality of life followed by NYHA-FC and 6MWT. When patients were treated with metoprolol versus placebo, effect sizes and standardized response means were similar for the quality of life and 6MWT and were smaller for NYHA-FC.

Discussion

In the RESOLVD pilot study, the 6MWT was highly reliable when repeated twice at each of the 3 measurement times, was moderately inversely correlated with the NYHA-FC, and was weakly inversely correlated with quality of life measured with the Minnesota Living with Heart Failure questionnaire. For the responsiveness analysis, we used statistical analyses that are not commonly used in health sciences. The 6MWT was only responsive to change for patients allocated to the highest dose of candesartan or to enalapril alone. It was not responsive during the candesartan plus enalapril combination or with metoprolol in the RESOLVD pilot study. Quality of life was more responsive to change compared with NYHA-FC or 6MWT during the candesartan, enalapril, or combination phase of the trial, although this change was small when effect sizes and standardized response means were compared. All measurements of functional capacity and quality of life were poorly responsive during the metoprolol versus placebo phase.

Reliability of the 6MWT

Previous reports support the findings from the RESOLVD pilot study on the high reliability of the 6MWT. Smaller studies have demonstrated the reproducibility of the 6MWT in patients with HF. In a group of elderly patients, two 6MWTs were performed within 3 to 8 weeks. In the 24 patients who had reported no change in clinical status, the ICC or test-retest reliability was 0.91, which was similar to that of our analysis. Guyatt et al evaluated 25 patients with chronic lung disease and 18 patients with HF, who performed the 6MWT 6 times with 2-week intervals separating each test. Reliability was evaluated with use of the 4 last walking tests, and the within-patient SD was less than 6% of the mean score 65% of the time and within 12% of the mean score 95% of the time.

Construct validity of the 6MWT

In the current study, quality of life measured with the Minnesota Living with Heart Failure questionnaire was weakly inversely correlated with baseline 6MWT distance. The SOLVD quality-of-life substudy found similar results, that baseline quality of life measured with the Minnesota Living with Heart Failure was weakly inversely correlated with 6MWT distance ($r = -0.39$). A weak correlation was also reported in the SOLVD Registry substudy and in patients with advanced HF. The correlation between the 6MWT distance and the physical domain of the Minnesota Living with Heart Failure questionnaire was not assessed in this analysis.

The 6MWT distance was moderately inversely correlated to NYHA-FC in the RESOLVD pilot study. These results have been corroborated by other smaller studies. Guyatt et al reported a correlation coefficient of $-0.45$ between the walking test score and the NYHA-FC in 18 patients with HF ($P = 0.058$). In the SOLVD Registry Substudy 6MWT performance was consistent with functional status in the extremes, with longer walking distances for patients without symptoms (NYHA-FC I) and shorter walking distances for patients with NYHA-FC III/IV symptoms. Because the NYHA-FC better reflects activities of daily living, a closer correlation with the 6MWT was expected.

Responsiveness

In the RESOLVD pilot study, the 6MWT was only responsive in patients treated with enalapril when effect sizes and standardized response means were used. The original RESOLVD pilot study publication showed no significant improvement with use of repeated-measures analysis of variance (ANOVA), which is consistent with the effect sizes and standardized response means analysis in the current article. In a study of 60 elderly patients with HF receiving no specific treatment intervention, measuring change of clinical status with a transitional scale from “much worse” to “much better,” the 6MWT was significantly responsive when measured with the responsiveness coefficient, defined as the variance of change divided by the sum of the variance of change and the error variance derived from ANOVA. Significant effect sizes for patients who had reported improvement (effect size = 0.85) and deterioration (effect size = 2.13) in their symptoms were also noted.

Our results are similar to those of a study evaluating the effects of bucindol on functional capacity and quality of life. In patients with HF treated with bucindol for 12 weeks, 6MWT distance, NYHA-FC, and quality of life measured with the Minnesota Living with Heart Failure questionnaire were not significantly improved. In the Prospective Randomized Evaluation of Carvedilol in Symptoms and Exercise (PRECISE) trial, although 6MWT distance was improved with use of nonparametric testing with rank transformation of the data ($P = 0.048$), quality of life measured with the Minnesota Living with Heart Failure questionnaire was not significantly improved. These results suggest the possible limitations of submaximal and maximal exercise testing in assessing the treatment effect of β-blockers in HF. Inconsistencies
across studies may reflect the negative chronotropic effect of β-blocker therapy on exercise capacity or that functional improvement may actually be difficult to demonstrate in patients with little disability at baseline.

When repeated-measures ANOVA is used as described in the RESOLVD pilot study, NYHA-FC and quality of life were not improved. Although a standardized method was used in an attempt to maintain consistency between participating centers in the RESOLVD pilot study, it is difficult to ensure that these are actually properly implemented in each center in a large multicenter study compared with a single-center study. The lack of responsiveness to change noted in this study is most likely related to the lack of significant effect of candesartan, enalapril, or the combination and of metoprolol on exercise capacity measured with the 6MWT. However, our study represents what is practical in a multicenter study, and individual reproducibility was very good in this setting.

Study limitations

The RESOLVD pilot study enrolled patients with HF with ischemic and dilated cardiomyopathy with LVEF <40 and walking distance of <500 m. The nonresponsiveness of the 6MWT could reflect that patients were receiving optimal maximum tolerated angiotensin-converting enzyme inhibitor before the onset of this study or that duration of treatment was too short to clearly demonstrate a clinical benefit. Candesartan, enalapril, and metoprolol may not significantly alter exercise capacity for this study period. In patients who have little disability at baseline, measurement of exercise capacity may be an inadequate tool to assess responsiveness. This study further demonstrates the lack of agreement on the use of exercise testing as an intermediate outcome in HF trials and on which measurement of exercise response (eg, exercise duration, peak oxygen consumption, submaximal exercise testing) better reflects changes in functional status, particularly with β-blockers.

Conclusion

This is the first report, on a large number of patients with HF, on the reliability, validity, and responsiveness of the 6MWT. This simple tool is highly reliable when performed twice within a 5-day period: it has a weak inverse correlation to disease-specific quality of life and is moderately inversely correlated to NYHA-FC. 6MWT responsiveness was limited in the RESOLVD pilot study with slightly better results for quality of life measured with the Minnesota Living with Heart Failure questionnaire and NYHA-FC. Although some therapeutic agents have been associated with improvement in exercise capacity, these changes have not always been correlated with improved survival. Tools to evaluate functional capacity remain important in evaluating the patient’s prognosis and responsiveness to therapy in clinical practice. The 6MWT is simple and easy to administer and should still be used as an outcome in clinical trials designed to evaluate different treatment modalities in patients with HF because inability to perform daily activities remains an important impediment in this population.

References

Medical and socioenvironmental predictors of hospital readmission in patients with congestive heart failure

Miyuki Tsuchihashi, MS,a Hiroyuki Tsutsui, MD,b Kazunori Kodama, MD,b Fumiyoshi Kasagi, PhD,c Soko Setoguchi, MD,a Masahiro Mohr, MD,a Toru Kubota, MD,a and Akira Takeshita, MDa Fukuoka and Hiroshima, Japan

Background Patients with chronic congestive heart failure (CHF) require frequent rehospitalization because of the exacerbation of CHF. It is of clinical importance to determine predicting factors for readmission to reduce this likelihood. Previous studies have focused primarily on the demographic and medical characteristics in selected subsets of patients. Therefore, within a broad cohort of consecutively hospitalized patients, we sought to identify not only demographic and medical predictors but also socioenvironmental factors associated with readmission.

Methods We assessed demographic (age, sex), medical (etiology of CHF, New York Heart Association functional class, left ventricular ejection fraction, previous admission for CHF, length of hospital stay, comorbidity, and medications), and socioenvironmental variables (occupation, financial resources, living alone, and follow-up visits) in 230 patients discharged with a diagnosis of CHF and recorded hospital readmission.

Results Within 1 year after discharge, 81 patients (35%) were readmitted. Five variables, including poor follow-up visits [odds ratio (OR) 4.9, 95% CI 2.0-11.8], previous admission for CHF [OR 3.3, 95% CI 1.8-6.1], no occupation [OR 2.6, 95% CI 1.2-5.5], longer hospital stay [OR 3.2, 95% CI 1.2-8.5], and hypertension [OR 2.0, 95% CI 1.1-3.7], were identified as significant independent predictors for readmission by multivariate logistic regression analysis.

Conclusions Our independent predictors of readmission support the importance of medical and socioenvironmental factors in the deterioration of CHF. Therefore interventions to decrease readmission should also target social management in all hospitalized patients. (Am Heart J 2001;142:ae7.)

4/1/117964