The Minimal Important Difference in the 6-Minute Walk Test for Patients with Pulmonary Arterial Hypertension

Stephen C. Mathai, Milo A. Puhan, Diana Lam, and Robert A. Wise

Objective: To estimate the MID in the 6MWT in patients with PAH.

Methods: Study subjects from the clinical trial of tadalafil in PAH, a 16-week, parallel-group, randomized clinical trial of patients who were treatment naive or on background therapy with an endothelin receptor antagonist, were eligible. 6MWT was performed using a standardized protocol. Distributional and anchor-based methods were used to estimate the MID; the latter method used the Physical Component Summary Score (PCS) of the Medical Outcomes Study 36-item short form (SF-36).

Measurements and Main Results: Four hundred five subjects were analyzed. Domains of the SF-36 were weakly to modestly associated with 6MWT. Change in the PCS of the SF-36 was most strongly associated with change in 6MWT (r = 0.40, P < 0.001) and thus was selected as the anchor for subsequent anchor-based analyses. Distributional analyses yielded estimates of the MID ranging from 25.1 to 38.5 m, whereas anchor-based analyses yielded an estimate of 38.6 m.

Conclusions: Using both distributional and anchor-based methods, the estimated consensus MID in the 6MWT for PAH is approximately 33 m. These results have important implications for (1) assessing treatment responses from clinical trials and metaanalyses of specific PAH therapy, and (2) sample size calculations for future study design.

Keywords: pulmonary hypertension; outcome measures; 6-minute walk test; minimal important difference

Pulmonary arterial hypertension (PAH) comprises a group of clinical disorders characterized by progressive increase in pulmonary artery pressure that often leads to right ventricular failure and death (1). Despite recent advances in therapies for PAH, mortality remains high (2). Although improvements in survival remain the ultimate goal of therapy for PAH, alternate outcome measures have been used to assess efficacy of medications. The 6-minute walk test (6MWT) has been used as a primary outcome measure in many studies of various forms of PAH (3–9). Based on the cumulative data regarding 6MWT in PAH, regulatory agencies currently accept exercise capacity as a primary outcome measure for clinical trials in PAH. Furthermore, these regulatory bodies have approved pharmacologic agents for PAH therapy based on small but statistically significant differences in 6-minute walk distances (6MWD) between treatment and placebo arms. A recent metaanalysis of randomized controlled trials of PAH therapy in patients with pulmonary hypertension reported a weighted mean improvement in 6MWT in subjects in the treatment arm of 42.8 m (range, 10–93 m; 95% confidence interval [CI], 27.8–57.8 m) (10). However, despite its widespread use in clinical trials of PAH, the minimal important difference (MID) of the 6MWT in patients with PAH has not been thoroughly evaluated (11, 12).

The MID is the smallest change or difference in an outcome measure, perceived as beneficial, that would justify a change in the patient’s medical management (13). Although the MID has been determined for the 6MWT in the chronic obstructive pulmonary disease (COPD) population (14, 15), the congestive heart failure population (16), and idiopathic pulmonary fibrosis population (17), this parameter has been reported in only one group of patients with PAH (18). In that study, only distribution-based methods were used to determine the MID; these methods rely solely on statistical methodology and do not include the relationship between patient-important outcome measures, such as health-related quality of life (19). Furthermore, MID estimates derived from distributional-based methods differ sometimes substantially across different methods (20).

With the many new therapeutic agents available for potential treatment of pulmonary hypertension, it is important to establish an MID in the most widely used outcome measure, the 6MWT. The MID for the 6MWT in PAH could be used to assess response to therapy in clinical trials (21). In addition to reporting the mean difference in 6MWT, investigators could report the proportion of
Estimation of the Minimally Important Difference

Both anchor-based and distributional methods for determining the MID were used. The anchor-based methods for determination of the MID use measures for which an MID has been established (the anchor) to estimate the MID for another metric. Furthermore, the anchor must have a relatively strong linear relationship with the metric of interest; therefore, PCS was chosen as the anchor (26). Prior studies of the PCS in chronic diseases have defined the MID as 5 units (22). The numerical value of the MID for the 6MWD was then determined for PCS MID of 5 units using the linear regression of change in PCS against change in 6MWD.

The distributional methods used are: (1) effect size (ES), (2) standardized response mean (SRM), (3) standard error of the measurement (SEMeas), and (4) 0.5 times the SD of the baseline measure (0.5 SD). ES is defined as the average of the difference between end-of-treatment and baseline scores divided by the SD of the baseline scores (29). The SRM uses the SD of the change of the measure and therefore accounts for the covariance between the baseline and end-of-study measures (30). The SEMeas was calculated by multiplying the SD of the baseline measurement by 0.5 (32).
TABLE 2. ANCHOR- AND DISTRIBUTIONAL-BASED ESTIMATES OF THE MINIMAL IMPORTANT DIFFERENCE FOR THE 6-MINUTE WALK TEST

<table>
<thead>
<tr>
<th>Method</th>
<th>MID Estimate (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchor method (MID PCS = 5)</td>
<td>38.6</td>
</tr>
<tr>
<td>ES [Eos − baseline/SD(Eos)]</td>
<td>38.4</td>
</tr>
<tr>
<td>SRM [Eos − baseline/SD(Eos)]</td>
<td>28.8</td>
</tr>
<tr>
<td>SEMeas [SD(baseline) × √1−ICC]</td>
<td>25.1</td>
</tr>
<tr>
<td>0.5 SD [0.5 × SD of baseline 6MWT]</td>
<td>38.5</td>
</tr>
</tbody>
</table>

Definition of abbreviations: 6MWT = 6-minute walk test; ICC = intraclass correlation coefficient; Eos = end of study; ES = effect size; MID = minimal important difference; PCS = physical component summary score; SEMeas = standard error of the measurement; SRM = standardized response mean.

RESULTS

As shown in Table 1, 405 subjects who completed the PHIRST trial were included in this analysis. The majority of subjects were white women who were, on average, 53 years of age. Most had idiopathic PAH (IPAH, 60.8%), but approximately one-quarter had PAH related to connective tissue disease (CTD-PAH). Nearly two-thirds of the subjects were World Health Organization (WHO) functional class III, and one-third were WHO functional class II; a minority were in functional classes I and IV. Hemodynamics revealed moderate to severe PAH, with a mean pulmonary artery pressure of 54 ± 13 mm Hg, a mean cardiac index of 2.6 ± 0.7 L/min/m², and a mean pulmonary vascular resistance (PVR) of 11.1 ± 5.2 Wood units. The mean 6MWD was 343.2 ± 76.8 m, suggesting moderate functional impairment. The change in 6MWD between baseline and end-of-study was 33 m (95% CI, 15–50 m).

SF-36 data for each domain and the PCS and mental component summary scores were available on 348 subjects who showed in each of the eight domains of SF-36 were available on 348 subjects who showed the strongest association and therefore was selected as the anchor for subsequent analyses (34).

Data, including demographic, functional class, 6MWD, and SF-36 at baseline, were available on 405 subjects who were included in the distributional analyses. Baseline and end-of-study 6MWT and SF-36 were available on 348 subjects who were ultimately included in the anchor-based analyses. There were no significant differences in the demographic, functional, or hemodynamic characteristics between these groups.

MID Calculations

Table 2 shows the point estimates for the MID for the 6MWT using the anchor-based, ES, SRM, SEMeas, and 0.5 SD methods. The anchor-based analysis using a change in PCS of 5 units yielded an estimate of 38.6 m. ES analyses corresponding to moderate (0.5) effect sizes yielded an estimate of 38.4 m; with SRM-based methods, the estimate was 28.8 m. SEMeas estimate was 25.1 m, and the estimate calculated using 0.5 SD resulted in a value of 38.5 m. The consensus of these values led to an estimated MID of the 6MWT in PAH of 33 m.

Table 3 shows the point estimates for the MID for the 6MWT for certain subgroups in the study. Overall, the estimates for individual subgroups were similar to the estimates for the overall cohort (range, from 24.4–40.6 m); for instance, no significant differences were noted between estimates for the treatment-naïve group and the background-therapy group. Although the anchor-based estimate for the CTD group was smaller than other subgroup estimates (24.4 m), this estimate remained in the range of estimates from the overall cohort (Table 1).

DISCUSSION

In this study, we report several estimates of the MID for the 6MWT for PAH. To our knowledge, this is the first report of the MID for the 6MWT in PAH using both anchor-based and distributional methods. Our data show that estimates of MID for the 6MWT are remarkably similar between these methods. The consistency of the estimates derived from different methods supports the validity and robustness of this estimate of the MID. Based on these findings, we estimate the MID of the 6MWT in PAH to be approximately 33 m.

Determination of an MID for an outcome measure is an integral component in the validation of that measure in a particular disease state and in the elucidation of the distinction between statistical and clinical significance. Because sample size heavily influences the power to detect a statistically significant change in the outcome of interest in any study (i.e., the larger the sample size, the smaller the statistically significant change that can be detected), assessing the clinical relevance of the minimum detectable difference is of paramount importance when designing and reporting the results of clinical trials. However, despite its importance, few parallel-group randomized clinical trials across the spectrum of human disease report the MID for the respective outcome measures used (35). To our knowledge, only two prior studies in PAH have provided estimates of MID for clinically relevant outcome measures: stroke volume by cardiac magnetic resonance imaging (36), SF-36 (18), and 6MWT (18).

TABLE 3. ANCHOR- AND DISTRIBUTIONAL-BASED ESTIMATES OF THE MINIMAL IMPORTANT DIFFERENCE FOR PULMONARY ARTERIAL HYPERTENSION SUBGROUPS

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Baseline 6MWD</th>
<th>Change 6MWD</th>
<th>R value*</th>
<th>Distributional (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPAH (n = 246)</td>
<td>346 (76.6)</td>
<td>30.2 (59.3)</td>
<td>0.32</td>
<td>39.0</td>
</tr>
<tr>
<td>CVD-PAH (n = 96)</td>
<td>326 (81.2)</td>
<td>12.4 (51.5)</td>
<td>0.51</td>
<td>24.4</td>
</tr>
<tr>
<td>Treatment naïve (n = 177)</td>
<td>339 (78.1)</td>
<td>25.5 (60.8)</td>
<td>0.47</td>
<td>38.7</td>
</tr>
<tr>
<td>Background therapy (n = 197)</td>
<td>354 (74.9)</td>
<td>31.2 (54.4)</td>
<td>0.33</td>
<td>36.4</td>
</tr>
</tbody>
</table>

Definition of abbreviations: 6MWT = 6-minute walk test; ES = effect size; MID = minimal important difference; PCS = physical component summary score; SRM = standardized response mean.

All results presented as mean (SD) unless otherwise specified.

*All P values < 0.0001.
Mathai, Puhan, Lam, et al.: MID in 6MWT for PAH

The estimate of the MID for the 6MWT in PAH reported in the current study differs slightly from the prior study of patients with PAH (18), which reported an MID of 41 m (range, 18.7–74.15 m). Similar to the current study, Gilbert and colleagues (18) used a large, randomized controlled study (the SUPER trial [4]) of sildenafil therapy in PAH. The patient populations included in the SUPER and PHIRST trials had similar proportions of IPAH and APAH subjects, similar distribution of WHO functional class, and similar disease severity as assessed by baseline 6MWD and hemodynamics. Thus, perhaps it is to be expected that the estimates were similar between studies. For instance, when compared with other randomized, double-masked, parallel group studies in PAH, distributional estimates of the MID, calculated by 0.5 times SD of the baseline 6MWD, fall within a range of 35.7 to 45 m (data not shown) (4–9, 37).

The estimate of MID for the 6MWT in PAH is remarkably similar to estimates of the MID for the 6MWT in other disease states (Table 4). Using the National Emphysema Treatment Trial data, Puhan and colleagues described an MID of around 26 m, derived by both anchor-based and distributional methods (14). Holland and colleagues found an MID estimate of 25 m using similar methodology in a smaller cohort of patients with COPD who participated in pulmonary rehabilitation (38). Recently, du Bois and colleagues reported an MID ranging from 24 to 45 m in the 6MWT for patients with idiopathic pulmonary fibrosis, again comparable to the current estimates for MID in patients with PAH (17). Thus, even the anchor-based estimate of MID for the CTD group in the current study (24.4 m), despite differing quantitatively from the other subgroup estimates in the stratified analyses, falls within a range of expected variability. Whether the consistency of these MID estimates across pulmonary disease states reflects an intrinsic characteristic of response to any disease or similarities in the functional limitations independent of disease pathogenesis is not known.

The current study used several distributional-based methods and an anchor-based method to estimate the MID in an effort to generate a robust and reliable estimate; however, there is no consensus about which method or quantity of change in the construct of a particular method truly represents the MID. There remain strengths and weaknesses of each of the methods used (19, 28). In the current study, the estimates derived from the distributional methods varied based on the SD of the parameter used in the calculations. The ES method uses the SD of the baseline 6MWT, whereas the SRM method uses the SD of the change in 6MWT (end-of-study minus baseline). Because the variation of the baseline 6MWT is larger than the variation of the change in 6MWT over time, the ES method will yield a larger MID estimate, as it does in the current study. However, because an intervention study using an exercise measure such as 6MWT is intended to interpret changes over time, the SRM method is likely more appropriate (34).

Anchor-based estimates of MID use an external clinical or patient-based measure to group patients by magnitude of response: no change, small, moderate, or large positive (or negative) changes (19). The anchor chosen should be relevant to the disease and have clinical usefulness in the disease state. In the current study, we investigated the relationship between change in each of the eight domains of the SF-36 along with the summary scores (PCS and MCS). The strongest correlation between parameters of the SF-36 and 6MWT was found with PCS; this parameter has particular relevance in PAH as the primary symptom in the disease is dyspnea on exertion. Furthermore, preliminary data suggest an independent relationship between PCS and survival in PAH (39). Therefore, PCS appears to have both face validity and clinical usefulness in PAH and, thus, was an appropriate anchor for this analysis.

**Limitations**

Despite the use of multiple methods to estimate the MID for the 6MWT in this study, several limitations exist. First, although PCS has face validity and clinical usefulness in PAH, as a generic measure of health-related quality of life, it is less responsive to change than disease-specific measures (27). However, among the parameters of the SF-36 and other metrics, such as preference-based instruments, the PCS demonstrates the most

---

**Table 4. Minimal Important Difference Estimates for the 6-Minute Walk Test in Other Chronic Cardiopulmonary Diseases**

<table>
<thead>
<tr>
<th>Study Population</th>
<th>Reference Number</th>
<th>Distributional</th>
<th>Anchor</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>COPD</td>
<td>14*</td>
<td>25.7 to 30.6</td>
<td>18.9 to 26.4</td>
<td></td>
</tr>
<tr>
<td>COPD</td>
<td>38†</td>
<td>25.5 to 26.5</td>
<td>24.5</td>
<td></td>
</tr>
<tr>
<td>COPD</td>
<td>34‡</td>
<td>29 to 42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COPD</td>
<td>54‡‡</td>
<td>45 (42 to 47)</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>IPF</td>
<td>42††</td>
<td>33</td>
<td>30.5 (19 to 45)</td>
<td></td>
</tr>
<tr>
<td>DPLD</td>
<td>16‡‡</td>
<td>19 to 22</td>
<td>–43 (–48.6 to –47)</td>
<td></td>
</tr>
<tr>
<td>CHF</td>
<td></td>
<td>47 to 49</td>
<td>Unable to estimate</td>
<td></td>
</tr>
</tbody>
</table>

Small meaningful change

Substantial change

Definition of abbreviations: CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; DPLD = diffuse parenchymal lung disease; IPF = idiopathic pulmonary fibrosis; NETT = National Emphysema Treatment Trial.

* From NETT study; severe COPD only.
† Small cohort (n = 44); anchor was patient-reported change.
‡ Poor correlations between 6MWT and patient-reported outcomes precluded anchor-based analyses; study cohort included patients with both moderate and severe COPD.
‡† Patient-reported global rating of change; comparisons to other patients.
‡‡ Large cohort; criterion reference used for anchor-based analysis.
‡‡‡ Small cohort (n = 48); anchor was patient-reported change in symptoms.
‡‡‡‡ Small cohort (n = 45); patient-reported global rating of change; effect sizes larger for deterioration than improvement.
‡‡‡‡‡ Cohort of elderly in strength training trial, subacute stroke survivors, or prospective community-based study; determined both smallest meaningful change and substantial change based on standard error of the measurement, ES = 0.2, and ES = 0.5.
responsiveness to change in chronic lung disease (27). To account for this diminished responsiveness, we chose to examine effect sizes of a moderate magnitude (change in PCS = 5 units). Thus, we expect the MID estimate derived from the anchor-based method adequately represents the MID, despite the limitations of the PCS as an anchor. Use of additional patient-reported outcomes to use as anchors would be ideal, but were not available in this dataset or did not have an established MID to use as an anchor (20). The US Food and Drug Administration has presented results of a pooled analysis of 13 randomized, double-masked, parallel group studies in PAH that correlates change in 6MWD with change in PVR index (PVRI), showing that change in PVRI is moderately associated with change in 6MWD (40). However, the clinical significance of change in PVR (or PVRI), specifically related to a patient-important outcome, has recently been drawn into question. van de Veerdonk and colleagues reported that although baseline PVR predicted outcome in a large cohort of patients with IPAH, change in PVR over time was not related to survival (41). Furthermore, there are no data from randomized trials relating changes in hemodynamic parameters to changes in quality of life. Therefore, the usefulness of determining an MID for a surrogate parameter such as PVR may be limited, in particular if the surrogate correlates poorly with the outcome that it is supposed to represent. Third, the study cohort was predominantly composed of patients with New York Heart Association functional class II or III disease. This limited sample may impact the estimate derived from distributional analyses, as the SD for the 6MWT may be smaller than that of a cohort that includes patients from functional classes I through IV and thus limit the generalizability of this estimate. However, because the SD of the change in 6MWT (for the SRM method) was used, this particular limitation may be mitigated.

Conclusions

Using both anchor-based and distributional-based methods, the estimated MID for the 6MWT in patients with PAH is around 33 m. The anchor-based method and distributional-based estimates were very similar and comparable to estimates of MID 5 units). The anchor-based method and distributional-based estimates to changes in quality of life. Therefore, the usefulness of determining an MID for a surrogate parameter such as PVR may be limited, in particular if the surrogate correlates poorly with the outcome that it is supposed to represent. Third, the study cohort was predominantly composed of patients with New York Heart Association functional class II or III disease. This limited sample may impact the estimate derived from distributional analyses, as the SD for the 6MWT may be smaller than that of a cohort that includes patients from functional classes I through IV and thus limit the generalizability of this estimate. However, because the SD of the change in 6MWT (for the SRM method) was used, this particular limitation may be mitigated.

Author disclosures are available with the text of this article at www.atsjournals.org.

References


