

Minimal Clinically Important Difference in the 6-minute-walk Distance for Patients with Pulmonary Arterial Hypertension

Jude Moutchia¹, Robyn L. McClelland², Nadine Al-Naamani³, Dina H. Appleby¹, Kristina Blank², Dan Grinnan⁴, John H. Holmes¹, Stephen C. Mathai⁵, Jasleen Minhas³, Corey E. Ventetuolo⁶, Roham T. Zamanian⁷, and Steven M. Kawut^{1,3}

¹Department of Biostatistics, Epidemiology, and Informatics and ³Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; ²Department of Biostatistics, School of Public Health, University of Washington, Seattle, Washington; ⁴Department of Internal Medicine, Virginia Commonwealth University, Richmond, Virginia; ⁵Department of Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland; ⁶Departments of Medicine and Health Services, Policy and Practice, Brown University, Providence, Rhode Island; and ⁷Department of Medicine, School of Medicine, Stanford University, Palo Alto, California

ORCID ID: 0000-0001-7864-6096 (J.M.).

Abstract

Rationale: The 6-minute-walk distance (6MWD) is an important clinical and research metric in pulmonary arterial hypertension (PAH); however, there is no consensus about what minimal change in 6MWD is clinically significant.

Objectives: We aimed to determine the minimal clinically important difference in the 6MWD.

Methods: We performed a meta-analysis using individual participant data from eight randomized clinical trials of therapy for PAH submitted to the U.S. Food and Drug Administration to derive minimal clinically important differences in the 6MWD. The estimates were externally validated using the Pulmonary Hypertension Association Registry. We anchored the change in 6MWD to the change in the Medical Outcomes Survey Short Form physical component score.

Measurements and Main Results: The derivation (clinical trial) and validation (Pulmonary Hypertension Association

Registry) samples were comprised of 2,404 and 537 adult patients with PAH, respectively. The mean \pm standard deviation age of the derivation sample was 50.5 ± 15.2 years, and 1,849 (77%) were female, similar to the validation sample. The minimal clinically important difference in the derivation sample was 33 meters (95% confidence interval, 27–38), which was almost identical to that in the validation sample (36 m [95% confidence interval, 29–43]). The minimal clinically important difference did not differ by age, sex, race, pulmonary hypertension etiology, body mass index, use of background therapy, or World Health Organization functional class.

Conclusions: We estimated a 6MWD minimal clinically important difference of approximately 33 meters for adults with PAH. Our findings can be applied to the design of clinical trials of therapies for PAH.

Keywords: PAH; walk test; minimum important difference; individual participant data meta-analysis

(Received in original form August 15, 2022; accepted in final form January 10, 2023)

Supported by the Pulmonary Hypertension Association (ATS-PHA Rino Aldrighetti grant [N.A.-N.]); the Cardiovascular Medical Research and Education Fund (S.M.K.); and the National Institutes of Health (K24 HL103844 [S.M.K.]; K23 HL141584 [N.A.-N.], and T32 HL007891 [J.M.]).

Author Contributions: J.M. and S.M.K. designed the study and drafted the manuscript with significant input from R.L.M., N.A.-N., D.G., S.C.M., C.E.V., and R.T.Z. D.H.A., K.B., J.H.H., and J.M. collected and harmonized the individual participant data. J.M., R.L.M., and S.M.K. completed the data analysis. All authors contributed to the interpretation of the results and critical revision of the manuscript and approved the final draft.

Correspondence and requests for reprints should be addressed to Steven M. Kawut, M.D., M.S., 423 Guardian Drive, Philadelphia, PA 19104. E-mail: kawut@upenn.edu.

This article has a related editorial.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Am J Respir Crit Care Med Vol 207, Iss 8, pp 1070–1079, Apr 15, 2023

Copyright © 2023 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.202208-1547OC on January 11, 2023

Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the

Subject: The 6-minute-walk distance (6MWD) is considered an important clinical endpoint for clinical trials of therapy for pulmonary arterial hypertension. The minimal clinically important difference in 6MWD is not established, making planning and interpretation of trials of new therapies challenging.

What This Study Adds to the

Field: This study establishes and validates the minimal clinically important difference in 6MWD using randomized clinical trials and a nationwide prospective registry in the United States.

Pulmonary arterial hypertension (PAH) is characterized by the remodeling of the small muscular pulmonary arteries resulting in elevated pulmonary vascular resistance, increased right ventricular afterload, and right ventricular failure. PAH affects about 15–60 people per million, and the 5-year survival rate in newly diagnosed patients is approximately 61% (1). Randomized clinical trials (RCTs) have led to the regulatory approval of 15 drugs for PAH. The change in 6-minute-walk distance (6MWD) was the primary outcome for most of these RCTs and the basis for regulatory approval (2). Regulators consider the 6MWD an intermediate clinical endpoint in PAH (i.e., intrinsically important to the function of patients, even if not a surrogate for survival) (3). Although studies in PAH often include a decrease in the 6MWD of more than 15% from baseline as a component of a composite time to clinical worsening endpoint, there is no consensus on the change in 6MWD that is clinically relevant in an individual patient or between groups of patients.

The minimal clinically important difference (MCID) is the smallest change in an outcome measure that a patient would identify as meaningful and would mandate a change in the patient's management in the absence of significant side effects or excessive cost (4). The MCID often refers to a mean group difference between patients allocated

to a new intervention compared with those allocated to a control arm (placebo) in a clinical trial. An MCID may also be used to identify responders to an intervention (an individual meeting or exceeding a meaningful threshold), allowing the comparison of response rates between study arms. The values of the MCID for each of these purposes differ, with the responder threshold being larger because it requires consideration of measurement precision and reliability. The MCID is commonly estimated using an anchor, a measure that reflects a true change in the status of an individual. The Medical Outcomes Study 36-item Short Form (SF-36) and 12-item Short Form (SF-12) are generic measures that produce a physical component score (PCS) that has been used previously as a health-related quality of life anchor (5–7).

Investigators usually determine a single MCID value for a population of patients with a disease. However, what constitutes a meaningful change could vary by the clinical profile of the patient. For example, the change in 6MWD that may be important to a 25-year-old man may be different from the change in 6MWD that may be important to a 72-year-old woman. Therefore, a strategy to estimate the MCIDs for specific profiles of patients with PAH could improve the patient-centeredness of this endpoint. Validated 6MWD MCIDs could guide treatment decisions in clinical practice, help with study design and choice of sample size in PAH RCTs, and highlight the importance of the individual patient's perspective and response in their management.

We aimed to determine the MCID for the change in 6MWD using individual participant data from clinical trials in PAH. We sought to develop a scoring system to compute personalized MCIDs and to validate the MCID in a prospective cohort of patients with PAH from throughout the United States.

Methods

Search Strategy and Selection Criteria

This study was registered with the Research Registry (reviewregistry1419). The derivation sample was drawn from individual participants in 21 RCTs of pulmonary hypertension therapies that had been submitted to the FDA (U.S. Food and Drug Administration) since 2000 (Table E1 in the online supplement). RCTs of pulmonary

hypertension therapies that were not submitted to the FDA were not included (8). The derivation sample for this study included adult PAH patients in phase III RCTs that reported 6MWD and SF-36 or SF-12 at two or more time points (Table E2). These trials generally excluded patients with very mild or very severe disease and mostly included patients with a baseline 6MWD of between 150 m and 450 m.

The validation sample was drawn from the study population of the PHAR (Pulmonary Hypertension Association Registry), a prospective registry of newly referred (within 6 months of the first visit) adult and pediatric patients with PAH or chronic thromboembolic pulmonary hypertension (CTEPH) at 67 pulmonary hypertension care centers (PHCCs) across the United States (9). The goal of the PHAR is to assess the quality of care and outcomes at PHCCs. The validation sample included adult patients with PAH in the PHAR with 6MWD and SF-12 PCS values at more than one time point.

Data Collection and Harmonization

The FDA provided raw datasets with individual participant data from each trial as Statistical Analysis System transport files, data dictionaries, and blank case report forms, with the goal of improving clinical trial conduct in PAH. The data harmonization process has been described elsewhere (10–12). Briefly, we used the Study Data Tabulation Model (Version 1.4) to organize the individual participant data into relevant domains. Demographic, body mass index (BMI), PAH diagnosis, WHO (World Health Organization) functional class, and right heart catheterization data were harmonized across the various trials. The time points at which data were captured in the individual trials were maintained, and there was no interpolation or extrapolation of data across time points.

The 6-minute-walk test is a self-paced test of exercise capacity in which patients are asked to walk as far as possible in 6 minutes. It is considered a submaximal exercise test (13). This test was performed several times over the course of the clinical trials using research methodology (e.g., American Thoracic Society guidelines) (14). The 6MWD was recorded in the trials in meters or feet; values in feet were converted to meters by multiplying by 0.3048 for this analysis.

The SF-36 was based on a multidimensional model of health. The 36 items were constructed to capture eight important domains of health: physical functioning, role limitations because of physical health problems, bodily pain, general health, vitality (energy/fatigue), social functioning, role limitations because of emotional problems, and mental health (15). These eight domains are summarized by the PCS and mental component score (MCS). Norm-based scoring of the domain and summary scores standardize results across populations. Twelve of the SF-36 items (i.e., the SF-12) account for more than 90% of the variance in SF-36 PCS and MCS in both general and patient populations (16). ARIES 1 and 2 (17) and SUPER (18) used the SF-36 version 1 (1993 US general population standard). AMBITION (19), SERAPHIN (20), STRIDE-1 (21), and PHIRST (22) used the SF-36 version 2 (1998 U.S. general population standard). AIR (23) used the SF-12 version 1 (1993 U.S. general population standard). SF-36 data from ARIES and SUPER were restandardized on the basis of the 1998 U.S. general population before harmonization. We retained the norm-based domain and component scores of AIR (23) because we were not able to standardize the SF-12 version 1 to the 1998 U.S. general population.

We reproduced the samples from the individual studies by computing descriptive statistics for the baseline data in each trial and compared them to those in the published trial manuscripts. The University of Pennsylvania Institutional Review Board considered the harmonization and secondary use of these data as exempt from approval.

The PHAR collected sociodemographic, anthropometric, right heart catheterization, medical and social history, symptom burden, medication, health-related quality of life, and longitudinal outcome data. The PHAR dataset (locked 03/29/22) included 6MWD in meters recorded from clinically performed 6-minute-walk tests and SF-12 version 2 (1998 U.S. general population standard) data, captured at entry into the registry and approximately after every 6 months during follow-up clinical visits. Informed consent was obtained for patients in the PHAR, which was approved by the University of Pennsylvania single Institutional Review Board.

Data Analysis

See the online supplement for details.

We anchored the change in 6MWD to the change in PCS to estimate the MCID

after calculating Pearson's correlation coefficients between initial, follow-up, and the change over time in 6MWD and PCS. *A priori*, we determined that correlation coefficients of at least 0.3 suggested that PCS was a suitable anchor as previously recommended (24, 25). The predicted value of the change in 6MWD associated with a five-unit change in the PCS from baseline to follow-up (previously reported MCID for PCS in pulmonary disease defining a responder with a significant change in physical health status) was considered to be the MCID (5, 7, 26). Secondary analyses used a 3.4 change in the PCS on the basis of the assumption of a baseline to follow-up correlation of 0.10 and an 80% confidence interval (CI) (7). We repeated the anchor-based analysis using a PCS cutoff of two, as recommended (7), to estimate the 6MWD MCID for mean group differences.

To determine whether an MCID in 6MWD tailored to a patient's characteristics better discriminated patients with a meaningful change in PCS, we computed the accuracy of a personal MCID in the PHAR validation set (i.e., the proportion of patients with an increase in PCS of 5 [and 3.4] or more who were correctly classified on the basis of the increase of 6MWD above their personal MCID). Missing data were not imputed. We performed all analyses initially in the clinical trials dataset and then repeated them in the PHAR dataset. All data analyses were done using R version 4.1.0 (2021–05–18).

Results

Study Participants

Of the 21 available RCTs, 8 phase III RCTs included the SF-36 or the SF-12 (Table E2). Of the 2,810 patients in these studies, we excluded 23 (0.8%) patients younger than 18 years of age, 57 (2.0%) patients with CTEPH, and 326 (11.6%) patients with 6MWD or SF-36 or SF-12 measured less than twice, leaving 2,404 patients with PAH in the derivation set (clinical trials) (Figure 1A). The characteristics of the study sample for this analysis were similar to those of adult patients with PAH in the other phase III trials that did not collect SF-12 or SF-36 data (Table E3). The PHAR included 1,995 patients by March 29, 2,022 (Figure 1B). We excluded 36 (1.8%) patients younger than 18 years of age and 274 (13.7%) patients

with CTEPH. We excluded 1,148 (57.5%) patients who had 6MWD or SF-12 measured less than twice, leaving 537 patients with PAH in the validation set (PHAR). The median (interquartile range) duration from initial to follow-up assessment in the clinical trials was 16 weeks (12, 24), whereas it was 26 weeks (21, 34) in the PHAR.

The mean \pm standard deviation age of the derivation (clinical trial) study sample was 50.5 ± 15.2 years, and 1,849 (77%) were female, similar to those in the validation (PHAR) study sample (Table 1). About three-quarters of the patients were White, and there were 11% Hispanic or Latino patients in both samples. Idiopathic PAH was more common in the derivation (clinical trial) study sample than in the validation (PHAR) study sample, and drug and toxin-related PAH and portopulmonary hypertension were slightly more common in the validation (PHAR) study sample. The baseline 6MWD and PCS scores were similar between study samples.

The correlations between initial, follow-up and the change over time in 6MWD and PCS in the clinical trials sample and the PHAR are shown in Table E4 and Figure E1. There were clinically (and statistically) significant correlations (most r of at least 0.30) for all studies except the correlation in change from baseline for AIR (23), supporting the use of PCS as an anchor for 6MWD.

Population 6MWD MCID

The estimates of the population 6MWD MCID for responders from primary and sensitivity analyses in the derivation (clinical trial) and validation (PHAR) samples are shown in Figure 2. The responder 6MWD MCID using PCS of at least five from the one-step generalized estimating equation linear regression model in the derivation (clinical trial) sample was 33 m (95% CI, 27–38) and almost identical in the validation (PHAR) data set (36 m [95% CI, 29–43]). The two-step analysis of the derivation (clinical trial) data set provided a similar MCID (33 m [95% CI, 26–38]) (Figures 2 and E2).

The primary estimates were robust to the use of other anchor-based methods in both the derivation (clinical trials) dataset and validation (PHAR) dataset (e.g., receiver operating characteristic curve and change difference) (Figures 2 and E3). Empirical cumulative distribution function curves and probability density function curves for a PCS

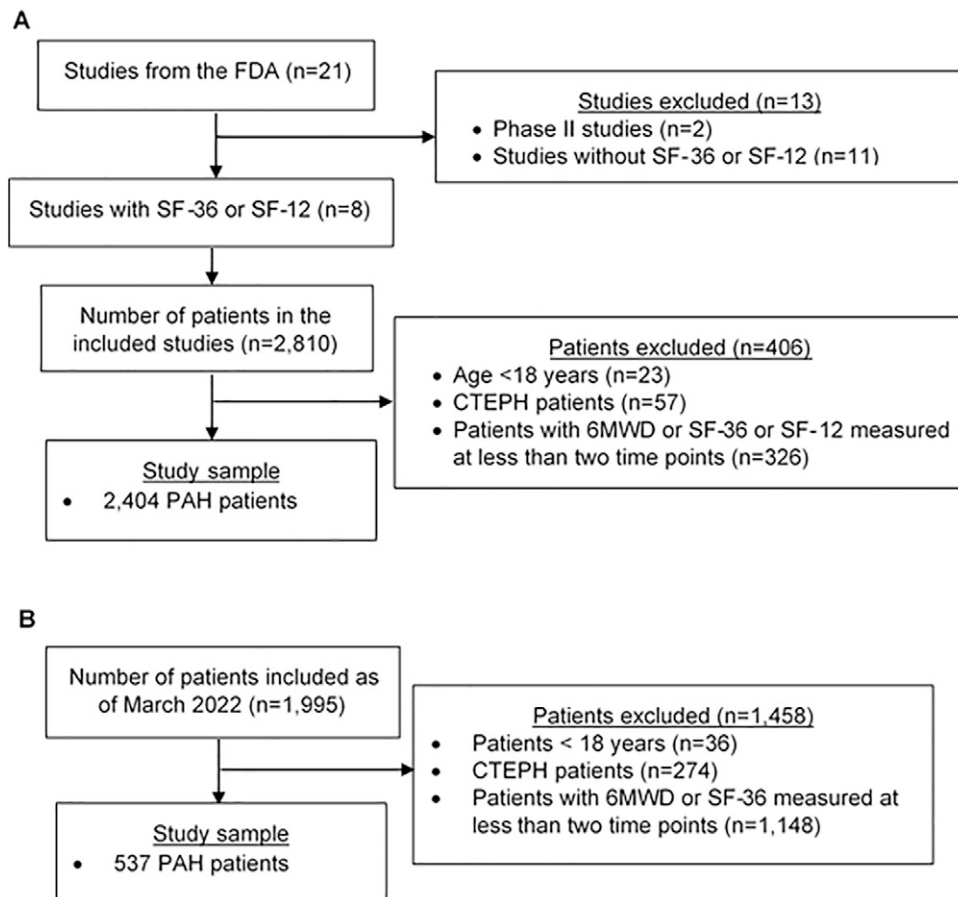


Figure 1. Participant flowchart. (A) Study sample from FDA clinical trials (derivation sample). (B) Study sample from the Pulmonary Hypertension Association Registry (validation sample). 6MWD = 6-minute-walk distance; CTEPH = chronic thromboembolic pulmonary hypertension; FDA = U.S. Food and Drug Administration; PAH = pulmonary arterial hypertension; SF = Short Form.

cut-off of five are shown in Figure E4. The anchor-based methods produced similar point estimates and overlapping 95% CIs both in the derivation (clinical trial) and validation (PHAR) data sets.

The secondary distributional methods showed that the standardized response mean and standard error of measurement MCIDs were similar to the anchor-based approach, whereas the standardized mean difference produced an MCID much greater than the other anchor-based and distributional approaches (Figure 2).

The 6MWD MCID did not significantly differ by age, sex, race, PAH etiology, BMI, WHO functional class, and background PAH therapy in the derivation (clinical trials) or validation (PHAR) data sets (Figures 3 and 4). However, patients with lower baseline 6MWD had larger 6MWD MCIDs in the derivation sample (P for interaction less than 0.001) and possibly in the validation sample (P for interaction equal to 0.09).

Analyses using a cutoff of 3.4 for the PCS anchor in the one-step analysis showed a similar MCID estimate of 28 m (95% CI, 23–34) in the derivation data set and 33 m (95% CI, 26–40) in the validation data set (Figure E5). Other analyses using the PCS 3.4 unit as an anchor are shown in Table E7 and Figures E6–E10.

The MCID for mean group differences (with a PCS cutoff of two, as recommended [7]) using one-step linear regression in the derivation (clinical trial) population was 24 m (95% CI, 18–31) and in the validation (PHAR) population 31 m (95% CI, 24–37), as shown in Figures E11–E14. The two-step estimate from the clinical trial population was 24 m (95% CI, 17–31).

Personalized 6MWD MCID

See the online supplement for details.

The MCID did not differ significantly on the basis of other covariates (Figure 3), so the equation to estimate a personalized 6MWD

MCID from the derivation set (clinical trials) only included the baseline 6MWD: $6MWD\ MCID = 63.70 + 2.69 \times PCS(5) - 0.12 \times \text{baseline } 6MWD$, simplified to $6MWD\ MCID = 77.15 - 0.12 \times \text{baseline } 6MWD$.

The accuracy of various 6MWD MCID cutoffs in discriminating patients with PAH with a meaningful clinical difference is shown in Table 2. The accuracies of the various MCID cutoffs were similar, and their CIs overlapped.

Discussion

We found that the 6MWD MCID to identify responders was approximately 33 m using anchor-based approaches in (to our knowledge) the first study using meta-analyzed individual clinical trial participant data and registry-based real-world data from almost 3,000 patients with PAH. Sensitivity analyses showed that

Table 1. Characteristics of Study Participants

Characteristic	Derivation Set (Clinical Trials)		Validation Set (PHAR)		Standardized Difference*
	n	Statistic	n	Statistic	
Age (yr), mean ± SD	2,404	50.5 ± 15.2	537	53.7 ± 16.0	0.20
Sex (female), n (%)	2,404	1,849 (76.9)	537	414 (77.1)	0.01
Race, n (%)	2,404		537		0.34
White		1,845 (76.7)		410 (76.4)	
Asian		239 (9.9)		27 (5.0)	
Black		107 (4.5)		50 (9.3)	
Other		28 (1.2)		23 (4.3)	
Unknown		185 (7.7)		27 (5.0)	
Ethnicity, n (%)	2,404		537		0.29
Hispanic or Latino		272 (11.3)		60 (11.2)	
Not Hispanic or Latino		2,131 (88.6)		455 (84.7)	
Unknown		1 (0.0)		22 (4.1)	
Body mass index (kg/m ²), mean ± SD	2,128	27.1 ± 6.3	529	29.9 ± 7.2	0.42
PAH etiology, n (%)	2,401		537		0.56
Idiopathic		1,407 (58.6)		224 (41.7)	
Associated with connective tissue disease		718 (29.9)		167 (31.1)	
Associated with congenital heart disease		154 (6.4)		33 (6.1)	
Drug and toxin-induced		64 (2.7)		58 (10.8)	
Heritable/familial		32 (1.3)		12 (2.2)	
Associated with HIV infection		26 (1.1)		7 (1.3)	
Portopulmonary hypertension		0 (0.0)		32 (6.0)	
Persistent pulmonary hypertension of the newborn		0 (0.0)		2 (0.4)	
Pulmonary veno occlusive disease or pulmonary capillary hemangiomatosis		0 (0.0)		2 (0.4)	
Mean right atrial pressure (mm Hg), median (IQR)	1,992	8.0 (5.0, 11.0)	527	9.0 (5.0, 13.0)	0.26
Mean pulmonary artery pressure (mm Hg), median (IQR)	2,131	50 (40, 60)	529	50 (40, 60)	0.07
Cardiac index (L/min/m ²), median (IQR)	1,980	2.32 (1.90, 2.83)	497	2.13 (1.73, 2.70)	0.24
Pulmonary vascular resistance (Wood units), median (IQR)	2,083	10.0 (6.4, 14.4)	483	9.5 (6.0, 13.3)	0.11
Pulmonary artery wedge pressure (mm Hg), median (IQR)	2,062	9.0 (7.0, 12.0)	506	11.0 (8.0, 14.0)	0.38
WHO functional class, n (%)	2,403		523		0.40
I		13 (0.5)		42 (8.0)	
II		927 (38.6)		210 (40.2)	
III		1,378 (57.3)		247 (47.2)	
IV		85 (3.5)		24 (4.6)	
6-minute-walk distance (m), mean ± SD	2,404	356 ± 90	537	347 ± 126	0.08
Physical component score, mean ± SD	2,404	35.3 ± 8.9	537	35.8 ± 10.7	0.05
Mental component score, mean ± SD	2,404	44.2 ± 12.1	537	48.5 ± 11.6	0.36

Definition of abbreviations: HIV = human immunodeficiency virus; IQR = interquartile range; PAH = pulmonary arterial hypertension; PHAR = Pulmonary Hypertension Association Registry; SD = standard deviation.

*Standardized mean differences less than 0.5 are considered small.

the estimate and confidence intervals were robust to differences in the modeling approach. MCIDs that identify responders should always be considered an approximation because of several sources of variability. However, our results were consistent across age, sex, race, BMI, type of PAH, background treatment status, and functional class, even though we have shown that sex and BMI may impact 6MWD (27). Contrary to our expectations, the MCID was larger in patients with lower 6MWD at initial assessment in both the clinical trial and registry study samples, although this baseline dependence of the MCID is likely spurious, as shown in prior studies (discussed below [28]). Personalized adjusted MCIDs were similar in accuracy to

the population MCID, and adjustment for baseline 6MWD may not be appropriate, so the evidence does not support a role for personalized MCIDs in clinical care or research. Finally, we have provided estimates of the MCIDs for between-group mean differences.

Previous studies of the MCID in patients with PAH used distributional methods (29) rather than multiple anchor-based methodologic approaches as recommended (25), included only one trial (which was also included in this analysis) (30), and anchored on clinical events (3); none included validation analyses in real-life data from the current era. Although we used a larger multinational study sample over a longer time span, our population 6MWD

MCID estimates from both clinical trial and registry study samples using one- and two-step models were consistent with those from the study using the RCT of tadalafil that triangulated anchor-based and distribution-based estimates (30), supporting the generalizability of this estimate across studies, time, and geography. However, our MCID was lower than the estimate obtained from a study that used only distributional methods in PAH (41 m) (29). One prior study from an earlier population of patients from the PHAR derived the MCID for the emPHasis-10 (a pulmonary hypertension-specific quality of life instrument) using distributional methods linked with a change in 6MWD of 35 m (31), providing further validation of our findings anchored to the

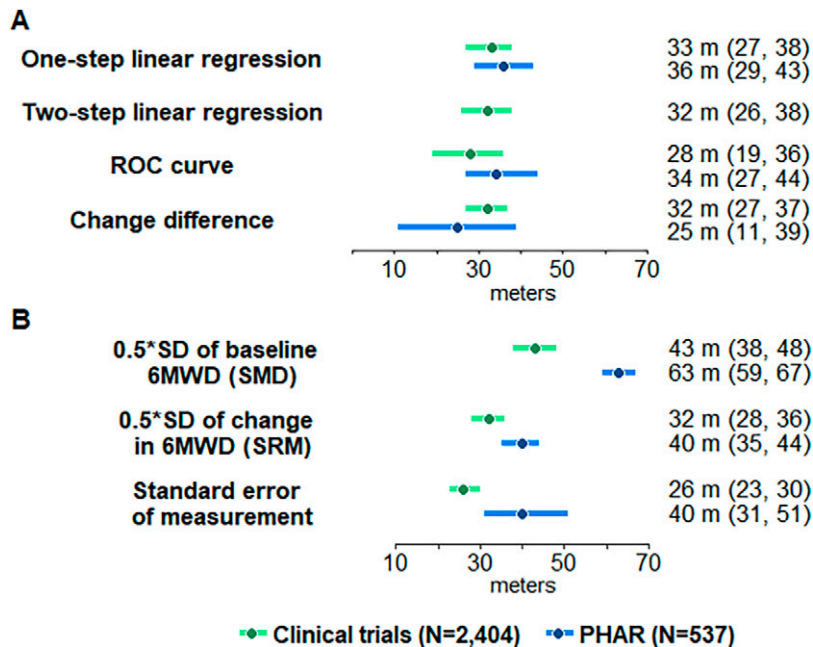


Figure 2. Minimal clinically important difference (95% confidence interval) in the 6-minute-walk distance using (A) anchor-based (physical component score change of at least five) and (B) distributional methods. 6MWD = 6-minute-walk distance; PHAR = Pulmonary Hypertension Association Registry; ROC = receiver operating characteristic; SMD = standardized mean difference; SRM = standardized response mean.

PCS. Finally, we distinguished the MCID for responders from the MCID for studying group differences; other studies have not provided specific estimates.

We previously published a study including some of the clinical trials in the current sample, which showed that 6MWD was not a good surrogate endpoint for short-term outcomes with a threshold effect of greater than 40 m required to possibly infer a reduction in clinical worsening (3). This discrepancy could be because of a focus on patient perception of a change in their physical health (captured by the PCS) in this study rather than a purely statistical relationship between some change in 6MWD and the risk of clinical worsening.

Although younger patients, males, and patients with lower BMI tended to have larger MCIDs, these (and other) differences were not statistically significant in either derivation or validation samples. Also, the type of PAH and use of background therapy did not significantly affect the MCID. Patients with lower baseline 6MWD had significantly higher 6MWD MCID, which is counterintuitive at first glance. We assumed that more severely affected patients would require a smaller

change in 6MWD to perceive this as clinical improvement. Consistent between the clinical trial and registry real-world populations, this finding could be explained in several ways. First, a severely impaired patient may require a greater change in 6MWD to detect the improvement because of factors that dull the perception of physical wellness. In this scenario, patients perceive the same change in PCS differently. Second, this observation could be a result of the floor and ceiling effects of the change value. Relationships between lower baseline measurements and higher MCIDs in various metrics and data simulations have been well-documented (32), suggesting that differences in responsiveness of the sample and nonrandom error in the change score across subgroups may account for this finding (28, 33, 34). Finally, type I error is possible (as the *P* value in the validation group was borderline); however, the consistently strong inverse relationship between baseline 6MWD and the MCID in our two independent study samples and the similar baseline dependency of the MCID in other disease states (32, 34) makes this less likely.

Some have recommended using the percent change in 6MWD as an endpoint

(often used as a component of the time to clinical worsening composite endpoint in PAH). However, we found significant heteroscedasticity when the percent change in 6MWD was plotted against the baseline 6MWD and an inverse relationship between the MCID (modeled as percent change in 6MWD) and the baseline 6MWD (data not shown), suggesting that modeling as percent change does not fully address the baseline dependency of the 6MWD. The use of percent change in 6MWD in a composite endpoint (being considered equivalent to hospitalization, lung transplantation, or death) likely has very different implications than those of an MCID, which identifies the smallest threshold in change that may be clinically important.

The establishment of 6MWD MCIDs in PAH may significantly advance clinical trial design. Regulatory bodies, clinicians, and investigators often want to know the probability of a meaningful response to an intervention (especially how a patient feels or functions) to decide if the cost, side effects, and burden are justified by the chance that a patient will benefit. This requires a validated responder threshold linked to a clinically important measure. The downsides to this approach have been recognized (35), especially the loss of power entailed in dichotomizing a continuous variable. The other traditional role of MCIDs is in comparing group means between arms of a clinical trial and could be used for minimum sample size calculation. We found an MCID of approximately 24 m linked to an increment of PCS of two, considered a reasonable increment for group mean differences (7).

Although the 6MWD MCID can be used by clinicians in defining treatment goals and monitoring the evolution of the disease, the accuracy in discriminating patients who reported a meaningful change at their follow-up visit was only adequate (with *c*-statistics around 0.60). The 6MWD in PHAR was not always measured on the same day as PCS, and clinical 6-minute-walk tests may not be as rigorous as research-based testing from clinical trials. The clinical utility of the MCIDs in this study needs further investigation before widespread use.

Strengths of our study include a large multinational clinical trial population over a long time span, the use of real-world

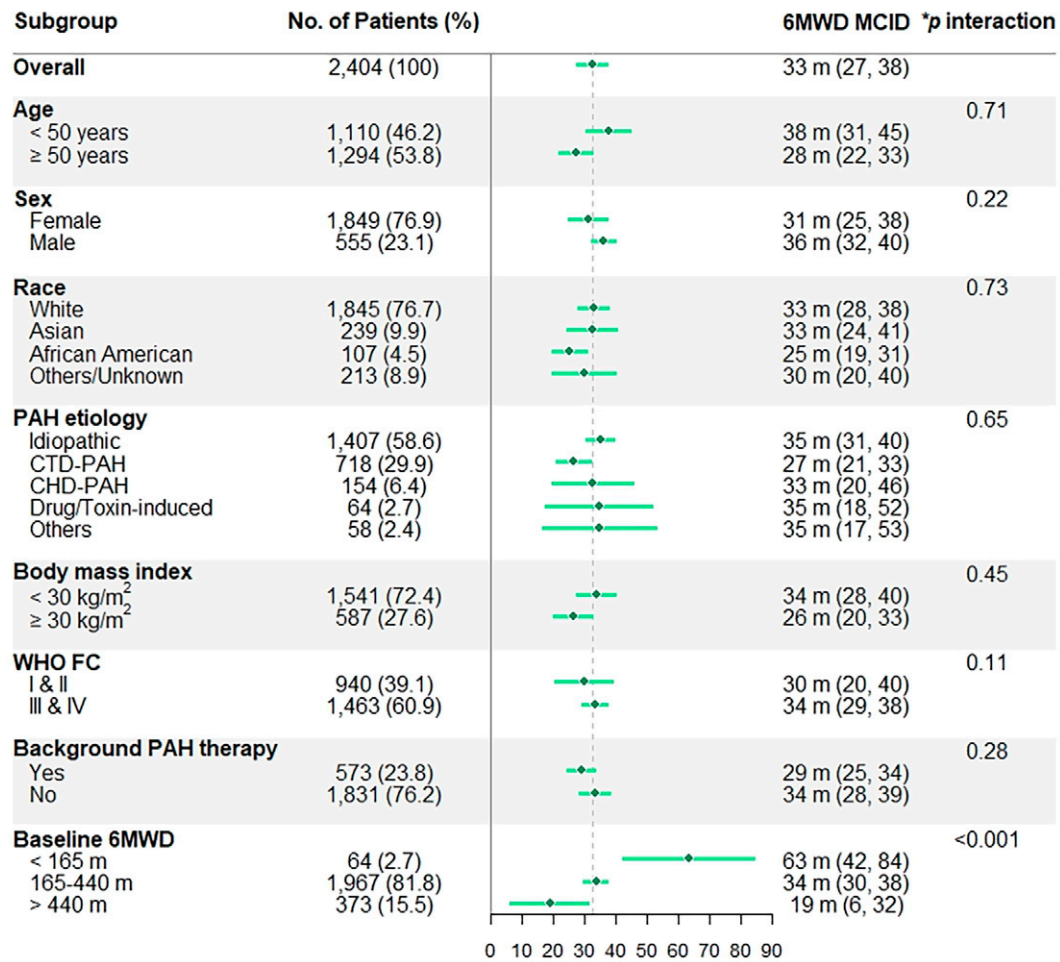


Figure 3. Subgroup 6-minute-walk distance (6MWD) MCID (95% confidence interval) estimates from the derivation (clinical trial) data on the basis of a change in physical component score of five. *Age and baseline 6MWD are analyzed as continuous variables. CHD-PAH = PAH associated with congenital heart disease; CTD-PAH = PAH associated with connective tissue disease; MCID = minimal clinically important difference; PAH = pulmonary arterial hypertension; WHO FC = World Health Organization functional class.

current registry data for validation of findings, multiple sensitivity analyses, and distinguishing MCID for responders versus MCID for use in studying group differences. Our study also has some limitations. It is possible that individual patients were in more than one of the clinical trials; however, we were unable to determine this. This is likely a relatively small number considering the time span of the trials. The dates of the trial recruitment and the PHAR did not overlap, so it is very unlikely that patients would have been in both datasets. Also, response shift could lead to PCS measurement bias (36). In response shift, patient-reported outcomes improve over time because the patient adapts to match their new

life circumstances to better cope with them (37).

The anchoring approach may present some issues in terms of the selection of the anchor, the correlation with the 6MWD, and the estimated MCID may be anchor-dependent (25). The SF-36 (or SF-12) is a generic health status questionnaire that may not be as sensitive as disease-specific questionnaires. However, 6MWD MCIDs in the PHAR obtained using PCS were similar to 6MWD MCIDs obtained using emPHASIS-10, which is a pulmonary hypertension-specific instrument (Figure E17). The PCS may not always reflect the same physical health status across patients, regions, countries, and eras. We normalized these

scores to the 1998 U.S. norms, which may not be valid for the countries from which our derivation (clinical trial) study population was drawn or the timeframes (10). The high degree of agreement in findings between the derivation and validation cohorts (the latter solely U.S.-based) makes such differences less likely to impact our findings. The exclusion of patients with fewer than two assessments of 6MWD or SF-36 (or SF-12) could lead to selection bias (excluding sicker patients who died before reassessment, leaving a healthier sample for analysis). Finally, we only had a small number of patients with low 6MWD at baseline (by design of the trials), so the generalizability of these findings to such patients is unknown.

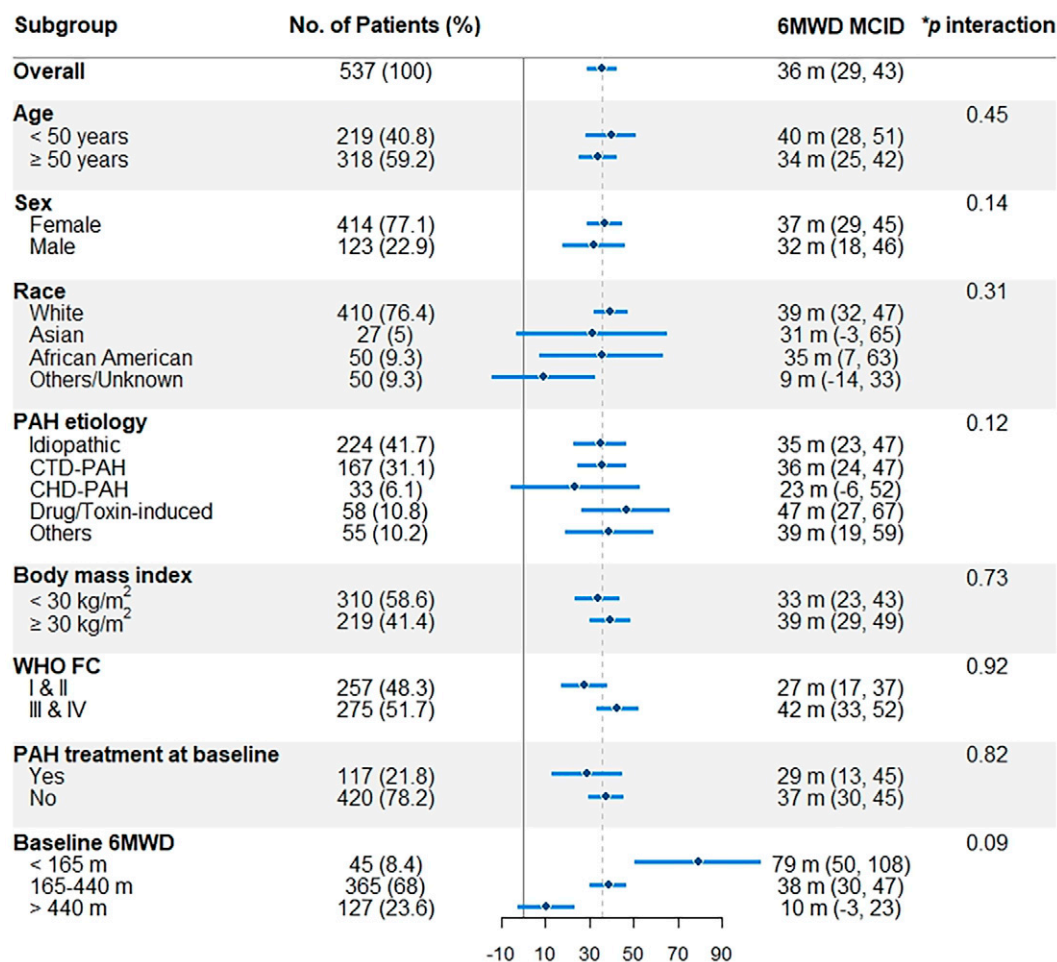


Figure 4. Subgroup 6-minute-walk distance (6MWD) MCID (95% confidence interval) estimates from the validation (Pulmonary Hypertension Association Registry) data on the basis of a change in physical component score of five. *Age and baseline 6MWD are analyzed as continuous variables. For definition of abbreviations, see Figure 3.

Table 2. Accuracy of Various 6-minute-walk Distance Minimal Clinically Important Difference Cutoffs in Detecting Patients with Pulmonary Arterial Hypertension with a Change in Physical Component Score of at least Five

6MWD MCID Cutoff	Validation Data	Accuracy (95% CI)*
Population MCID of 33 m	PHAR G1 (n = 537) PHAR G2 (n = 354) PHAR G3 (n = 225)	60.0 (55.8–64.1) 68.1 (63.2–72.9) 66.7 (60.5–72.8)
Population MCID stratified by baseline 6MWD: 63 m for patients with baseline 6MWD < 165 m; 34 m for patients with baseline 6MWD 165–440 m; and 19 m for patients with baseline 6MWD > 440 m	PHAR G1 (n = 537) PHAR G2 (n = 354) PHAR G3 (n = 225)	59.2 (55.1–63.4) 66.7 (61.8–71.6) 64.9 (58.7–71.1)
Individual MCID obtained from a prediction model: Change in PCS (5) + baseline 6MWD	PHAR G1 (n = 537) PHAR G2 (n = 354) PHAR G3 (n = 225)	60.1 (56.0–64.3) 66.1 (61.2–71.0) 63.6 (57.3–69.8)
Individual MCID obtained from a prediction model: Change in PCS (5) + baseline 6MWD + change in PCS (5) × baseline 6MWD	PHAR G1 (n = 537) PHAR G2 (n = 354) PHAR G3 (n = 225)	59.6 (55.4–63.7) 65.3 (60.3–70.2) 63.1 (56.8–69.4)

Definition of abbreviations: 6MWD = 6-minute-walk distance; CI = confidence interval; G1: Between baseline and 1st follow-up visit; G2: Between 1st and 2nd follow-up visit; G3: Between 2nd and 3rd follow-up visit. MCID = minimal clinically important difference; PCS = physical component score. *Proportion of patients with and without a change in PCS ≥ 5 who are correctly classified.

Conclusions

Using a large number of patients in RCTs of pulmonary arterial hypertension therapies and a registry from centers across the United States, we estimated a 6MWD MCID threshold of approximately 33 m to identify a significant clinical response in a patient with PAH for research purposes. A smaller MCID of approximately 24 m is warranted for

interpreting group mean differences in RCTs. The integration of such patient-focused study endpoints with other indicators of morbidity and mortality in early- and late-phase clinical trials in PAH requires further investigation. ■

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

Acknowledgment: The PHAR (Pulmonary Hypertension Association Registry) is supported by the Pulmonary Hypertension Care Centers, Inc., a supporting organization of the Pulmonary Hypertension Association. The authors thank the other investigators, the staff, and particularly participants of the PHAR for their valuable contributions. A full list of participating PHAR sites and institutions can be found at www.PHAssociation.org/PHAR.

References

- Farber HW, Miller DP, Poms AD, Badesch DB, Frost AE, Muros-Le Rouzic E, et al. Five-year outcomes of patients enrolled in the REVEAL registry. *Chest* 2015;148:1043–1054.
- Sitbon O, Gombert-Maitland M, Granton J, Lewis MI, Mathai SC, Rainisio M, et al. Clinical trial design and new therapies for pulmonary arterial hypertension. *Eur Respir J* 2019;53:1801908.
- Gabler NB, French B, Strom BL, Palevsky HI, Taichman DB, Kawut SM, et al. Validation of 6-minute walk distance as a surrogate end point in pulmonary arterial hypertension trials. *Circulation* 2012;126:349–356.
- Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407–415.
- Witt S, Krauss E, Barbero MAN, Müller V, Bonniaud P, Vancheri C, et al. Psychometric properties and minimal important differences of SF-36 in idiopathic pulmonary fibrosis. *Respir Res* 2019;20:47.
- Puhan MA, Guyatt GH, Goldstein R, Mador J, McKim D, Stahl E, et al. Relative responsiveness of the chronic respiratory questionnaire, St. Georges respiratory questionnaire and four other health-related quality of life instruments for patients with chronic lung disease. *Respir Med* 2007;101:308–316.
- Maruish ME. User's manual for the SF-36v2 health survey. Lincoln, RI: Quality Metric Incorporated; 2011.
- Liu HL, Chen XY, Li JR, Su SW, Ding T, Shi CX, et al. Efficacy and safety of pulmonary arterial hypertension-specific therapy in pulmonary arterial hypertension: a meta-analysis of randomized controlled trials. *Chest* 2016;150:353–366.
- Gray MP, Kawut SM. The pulmonary hypertension association registry: rationale, design, and role in quality improvement. *Adv Pulm Hypertens* 2018;16:185–188.
- Min J, Appleby DH, McClelland RL, Minhas J, Holmes JH, Urbanowicz RJ, et al. Secular and regional trends among pulmonary arterial hypertension clinical trial participants. *Ann Am Thorac Soc* 2021;19:952–961.
- Urbanowicz RJ, Holmes JH, Appleby D, Narasimhan V, Durborow S, Al-Naamani N, et al. A semi-automated term harmonization pipeline applied to pulmonary arterial hypertension clinical trials. *Methods Inf Med* 2021;61:3–10.
- McCarthy BE, McClelland RL, Appleby DH, Moutchia JS, Minhas JK, Min J, et al. Body mass index and treatment response in patients with pulmonary arterial hypertension: a meta-analysis. *Chest* 2022;162:436–447.
- Holland AE, Spruit MA, Troosters T, Puhan MA, Pepin V, Saey D, et al. An official European Respiratory Society/American Thoracic Society technical standard: field walking tests in chronic respiratory disease. *Eur Respir J* 2014;44:1428–1446.
- ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the six-minute-walk test. *Am J Respir Crit Care Med* 2002;166:111–117.
- Ware JE, Kosinski M, Dewey JM, Gandek B. SF36 health survey: manual and interpretation guide. Lincoln, RI: Quality Metric Incorporated; 1993.
- Ware JE. How to score version 2 of the SF-12 health survey: (with a supplement documenting version 1). Lincoln, RI: Quality Metric Incorporated; 2002.
- Galiè N, Olschewski H, Oudiz RJ, Torres F, Frost A, Ghofrani HA, et al.; Ambrisentan in Pulmonary Arterial Hypertension, Randomized, Double-Blind, Placebo-Controlled, Multicenter, Efficacy Studies (ARIES) Group. Ambrisentan for the treatment of pulmonary arterial hypertension: results of the ambrisentan in pulmonary arterial hypertension, randomized, double-blind, placebo-controlled, multicenter, efficacy (ARIES) study 1 and 2. *Circulation* 2008;117:3010–3019.
- Galiè N, Ghofrani HA, Torbicki A, Barst RJ, Rubin LJ, Badesch D, et al.; Sildenafil Use in Pulmonary Arterial Hypertension (SUPER) Study Group. Sildenafil citrate therapy for pulmonary arterial hypertension. *N Engl J Med* 2005;353:2148–2157.
- Galiè N, Barberà JA, Frost AE, Ghofrani H-A, Hoeper MM, McLaughlin VV, et al.; AMBITION Investigators. Initial use of ambrisentan plus tadalafil in pulmonary arterial hypertension. *N Engl J Med* 2015;373:834–844.
- Pulido T, Adzerikho I, Channick RN, Delcroix M, Galiè N, Ghofrani H-A, et al.; SERAPHIN Investigators. Macitentan and morbidity and mortality in pulmonary arterial hypertension. *N Engl J Med* 2013;369:809–818.
- Barst RJ, Langleben D, Frost A, Horn EM, Oudiz R, Shapiro S, et al.; STRIDE-1 Study Group. Sitaxsentan therapy for pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2004;169:441–447.
- Galiè N, Brundage BH, Ghofrani HA, Oudiz RJ, Simonneau G, Safdar Z, et al.; Pulmonary Arterial Hypertension and Response to Tadalafil (PHIRST) Study Group. Tadalafil therapy for pulmonary arterial hypertension. *Circulation* 2009;119:2894–2903.
- Olschewski H, Simonneau G, Galiè N, Higenbottam T, Naeije R, Rubin LJ, et al.; Aerosolized Iloprost Randomized Study Group. Inhaled iloprost for severe pulmonary hypertension. *N Engl J Med* 2002;347:322–329.
- Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102–109.
- Swigris J, Foster B, Johnson N. Determining and reporting minimal important change for patient-reported outcome instruments in pulmonary medicine. *Eur Respir J* 2022;60:2200717.
- Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991;59:12–19.
- Ventetuolo CE, Moutchia J, Baird GL, Appleby DH, McClelland RL, Minhas J, et al. Baseline sex differences in pulmonary arterial hypertension randomized clinical trials. *Ann Am Thorac Soc* 2023;20:58–66.
- Terluin B, Roos EM, Terwee CB, Thorlund JB, Ingelsrud LH. Assessing baseline dependency of anchor-based minimal important change (MIC): don't stratify on the baseline score! *Qual Life Res* 2021;30:2773–2782.
- Gilbert C, Brown MCJ, Cappelleri JC, Carlsson M, McKenna SP. Estimating a minimally important difference in pulmonary arterial hypertension following treatment with sildenafil. *Chest* 2009;135:137–142.
- Mathai SC, Puhan MA, Lam D, Wise RA. The minimal important difference in the 6-minute walk test for patients with pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2012;186:428–433.
- Borgese M, Badesch D, Bull T, Chakinala M, DeMarco T, Feldman J, et al.; PHAR Study Group. EmPHasis-10 as a measure of health-related quality of life in pulmonary arterial hypertension: data from PHAR. *Eur Respir J* 2021;57:2000414.

32. Ward MM, Guthrie LC, Alba M. Dependence of the minimal clinically important improvement on the baseline value is a consequence of floor and ceiling effects and not different expectations by patients. *J Clin Epidemiol* 2014;67:689–696.
33. Ward MM, Alba MI. Estimates of minimal clinically important improvements vary with the responsiveness of the sample. *J Clin Epidemiol* 2022;142: 110–118.
34. Alma H, de Jong C, Jelusic D, Wittmann M, Schuler M, Kollen B, *et al.* Baseline health status and setting impacted minimal clinically important differences in COPD: an exploratory study. *J Clin Epidemiol* 2019;116: 49–61.
35. Harrell FE. Responder analysis: loser x 4. 2019 [accessed 2022 Jul 20]. Available from: <https://discourse.datamethods.org/t/responder-analysis-loser-x-4/1262>.
36. Schwartz CE, Sprangers MAG. Response shift. In: Michalos AC, editor. *Encyclopedia of quality of life and well-being research*. Springer Netherlands: Dordrecht, the Netherlands; 2014. pp. 5542–5547.
37. Ilie G, Bradfield J, Moodie L, Lawen T, Ilie A, Lawen Z, *et al.* The role of response-shift in studies assessing quality of life outcomes among cancer patients: a systematic review. *Front Oncol* 2019;9:783.