Pulmonary Circulation

WILEY

PVRi | Pulmonary Circulation Open Access

RESEARCH ARTICLE OPEN ACCESS

Characteristics of Daily Walking Bouts as Valid and Reliable Indicators of Exercise Capacity in Pulmonary Arterial Hypertension: Insights From the Randomized Controlled Study With Selexipag (TRACE)

Rana Zia Ur Rehman[1] | Federico Parisi[2] | Noman Ashraf[2] | Meenakshi Chatterjee[2] | Robert P. Frantz[3] | Anna R. Hemnes[4] | Nikolay V. Manyakov[5] | Oscar M. Carrasco-Zevallos[2] | Tarik Yardibi[2] | Mona Selej[6] | Tommaso Mansi[7] | Preston Dunnmon[7] | Dzmitry A. Kaliukhovich[5]

[1]Janssen Pharmaceutical Companies of Johnson & Johnson, High Wycombe, Buckinghamshire, UK | [2]Janssen Pharmaceutical Companies of Johnson & Johnson, Cambridge, Massachusetts, USA | [3]Department of Cardiovascular Medicine, Mayo Clinic, Rochester, Minnesota, USA | [4]Division of Allergy, Pulmonary and Critical Care Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA | [5]Janssen Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium | [6]Janssen Pharmaceutical Companies of Johnson & Johnson, South San Francisco, California, USA | [7]Janssen Pharmaceutical Companies of Johnson & Johnson, Spring House, Pennsylvania, USA

Correspondence: Rana Zia Ur Rehman (rrehman5@its.jnj.com)

Received: 18 September 2024 | **Revised:** 23 January 2025 | **Accepted:** 28 April 2025

Funding: The study was funded by Janssen Pharmaceutical Companies of Johnson & Johnson.

Keywords: 6-min walk distance test | actigraphy | bout distance | bout duration | step rate

ABSTRACT

There is a need for objective, continuous and remote assessment of exercise capacity in patients with pulmonary arterial hypertension (PAH). Using data from the TRACE study, in which 108 adult patients with PAH were continuously monitored with a wrist-worn accelerometer, we evaluated whether actigraphy can facilitate continuous monitoring of exercise capacity. Distributions of step rate, distance and duration of patient's walking bouts were estimated at baseline, Week 16 and Week 24 using 2-week periods of actigraphy data. Twenty-one metrics per walking bout characteristic were described (mean, standard deviation, 19 percentiles [5th–95th]). The relationships between these metrics and the 6-min walk distance (6MWD), Borg dyspnea index (BDI), and the PAH Symptoms and Impact questionnaire (PAH-SYMPACT) Physical Impact domain score were assessed at the three timepoints. Test-retest reliability, and discriminant and known-group validity of each metric were also evaluated. All metrics of step rate and bout distance were significantly correlated with 6MWD (Pearson's correlation coefficients: 0.34–0.67; $p < 0.005$) and the PAH-SYMPACT Physical Impact domain score (Pearson's correlation coefficients: −0.51 to −0.32; $p < 0.05$) at all timepoints. Negative correlations were also observed with the BDI and the actigraphy-derived metrics, with the majority reaching significance. Strong test-retest reliability was demonstrated (intra-class correlation coefficient ≥ 0.70). The metrics differentiated well between patients with varying disease severity levels. In conclusion, actigraphy-derived metrics of patients' walking bouts correlate significantly with 6MWD, BDI and physical impacts of PAH, indicating the utility of actigraphy in facilitating continuous monitoring of exercise capacity in adult patients with PAH.

Trial Registration: ClinicalTrials.gov: NCT03078907; URL: clinicaltrials.gov.

1 | Introduction

Pulmonary arterial hypertension (PAH) is a rare, debilitating disease of the pulmonary vasculature, characterized by progressive increases in pulmonary artery pressure and pulmonary vascular resistance, which ultimately leads to right heart failure and death [1, 2]. A decrease in physical activity, together with dyspnea, is a common presenting complaint for patients with PAH that contributes to a reduced overall health-related quality of life [3]. The 6-min walk test (6MWT), which assesses the maximum distance walked in 6 min (the 6-min walk distance [6MWD]), is the most widely used measure of exercise capacity in pulmonary hypertension centers [4]. As an intrinsically important measure of function in PAH [5] and despite not being a surrogate for survival, the 6MWD is considered an intermediate clinical endpoint in PAH, in that some absolute change in distance walked may improve a patient's quality of life [6]. However, it does not fully reflect a patient's overall daily life physical activity, providing only a snapshot within a controlled clinical setting. There is thus a need for objective, continuous and potentially remote assessment of exercise capacity in patients with PAH.

Actigraphy offers a potential alternative to the 6MWD by objectively quantifying physical activity in a continuous, ubiquitous, and remote manner [7, 8]. Actigraphy is a wearable-based, noninvasive method that, through the use of accelerometers, allows the collection of all walking bout data in a patients' everyday life. This provides comprehensive information about the ambulatory behavior and physical activity of patients in their day-to-day naturalistic settings over an extended time [9]. Such devices are practical and passive, requiring minimal engagement and effort from the user. With advancements in consumer wearable technology, accelerated by the COVID-19 pandemic [10], the use of actigraphy and physical activity monitoring devices in clinical research is steadily growing [11].

Recognizing the potential utility of actigraphy-derived measures as an alternative to the conventional measures such as 6MWD, health authorities are beginning to endorse their use as efficacy endpoints in clinical trials. For example, measures of physical activity that were based on activity counts [12] generated continuously by wrist-worn actigraphy devices and deemed a proxy of the activity intensity have been used as the primary efficacy endpoints in some recent clinical trials in PAH [8, 13] and fibrotic interstitial lung disease [14]. The European Medicines Agency also recently qualified actigraphy-derived stride velocity as a primary endpoint in studies of ambulatory Duchenne muscular dystrophy [15], showcasing the utility of measures focusing on specific activity categories, such as walking, rather than on overall physical activity. These examples demonstrate feasibility and utility of actigraphy-derived measures for use as an efficacy endpoint in prospective clinical trials and also provide a roadmap for the development and validation of novel actigraphy-based endpoints.

The phase 4 TRACE study (clinicaltrials.gov, NCT03078907) employed several actigraphy-based measures of physical activity as the primary efficacy endpoints to evaluate the effect of selexipag on daily life physical activity, self-reported symptoms and their impacts in patients with PAH [8]. Changes in daily life physical activity were small and variable, with no statistically significant treatment effect reported. The study nonetheless successfully collected high quality actigraphy data along with several conventional clinical outcome measures, including the 6MWD, Borg dyspnea index, and the PAH—Symptoms and Impact questionnaire (PAH-SYMPACT). The actigraphy and clinical outcomes data were collected throughout the study, enabling testing for relationships between different actigraphy-derived measures of patient's physical activity and well-being. Unlike the primary analysis of the TRACE data [8] and other studies [7, 13, 14, 16, 17] that have explored the relationship between the volume of patient's physical activity (e.g., time spent in non-sedentary or moderate-to-vigorous physical activity, average daily step count or activity time) and patient's well-being, this current analysis of the TRACE data focuses on the characteristics of patient's physical activity. Specifically, the aim of this analysis was to test whether the characteristics of patient's walking bouts, such as step rate (steps per minute), bout distance (distance covered per bout) and duration (duration of a single bout), as assessed via a wrist-worn actigraphy device, convey information on patient exercise capacity. If so, the goal was to test whether these characteristics fulfill regulatory criteria [15] for qualifying as an efficacy endpoint in prospective clinical trials, with the criteria including construct validity (convergent, known-group, and discriminant validity), repetitiveness (test-retest reliability), and responsiveness (sensitivity to change).

2 | METHODS

2.1 | Study Design and Patient Population

TRACE (Effect of Selexipag on Daily Life Physical Activity of Patients with PAH) was a prospective, multi-center, randomized, placebo-controlled, double-blind exploratory phase 4 study (clinicaltrials.gov, NCT03078907) and has been previously described in detail [8]. Ethical approval was received from independent ethics committees/institutional review boards, and the study was conducted in compliance with the Declaration of Helsinki. All patients provided written informed consent.

Patients were randomized (1:1) to receive selexipag or placebo for 24 weeks (Figure 1A). Selexipag titration was performed as previously described [18]. Eligible patients were 18–75 years of age with PAH (idiopathic, heritable, drug and toxin induced, or associated with connective tissue disease, HIV infection, or corrected congenital heart disease [simple systemic-to-pulmonary shunts ≥ 1 year after repair]) diagnosed by right heart catheterization [19, 20]. Patients were required to be in World Health Organization (WHO) functional class II or III, without hospitalization or worsening WHO functional class in the 30 days before screening and to have a 6MWD ≥ 100 m [8]. Patients had to be receiving an endothelin receptor antagonist alone or in combination with a phosphodiesterase type 5 inhibitor or soluble guanylate cyclase stimulator for ≥ 90 days and at a stable dose for ≥ 30 days before randomization [8]. Participation in an exercise-based rehabilitation program was not permitted in the 8 weeks before study or at any point during the study.

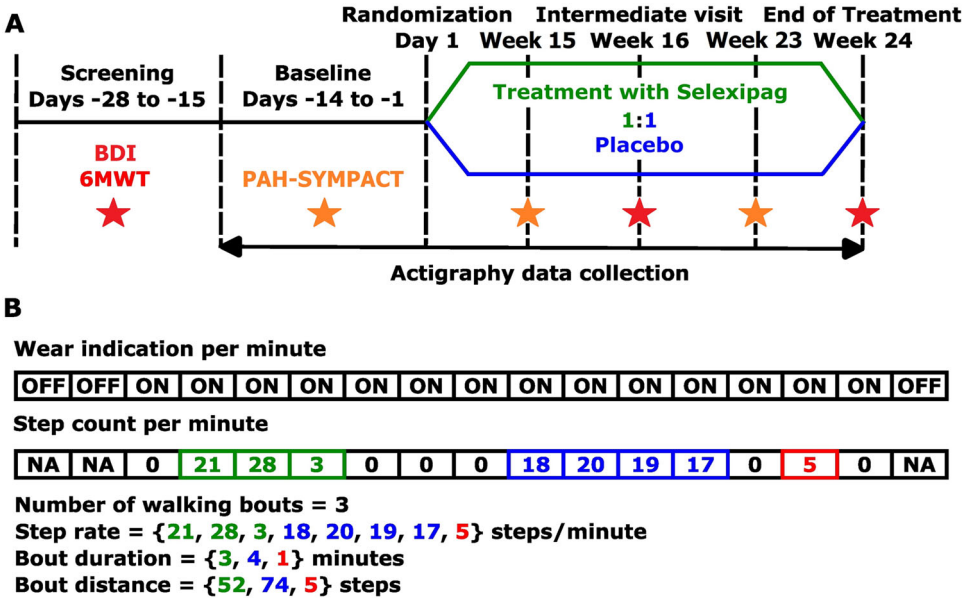

FIGURE 1 | Study design (A), and actigraphy-derived metrics (B). The red stars in panel A denote measurements of the 6MWD and Borg dyspnea index during screening and at the two follow-up clinical visits at Week 16 and 24. The Week 16 and 24 are further referred to as "Intermediate" and "End of Treatment" visits throughout the manuscript, respectively. The orange stars in panel *A* denote administrations of PAH-SYMPACT at baseline, Week 15 and 23. The actigraphy data were collected at baseline and throughout the study till the end of study. Panel *B* shows the two example data streams that include wear indication ("ON" and "OFF") and step counts, as generated by the ActiGraph's proprietary algorithms using epochs of 60 s. Step counts were generated only when the accelerometer was worn by a patient. NA denotes "Not Applicable." The three continuous sequences of positive step counts during wear time, highlighted in green, blue, and red for the sake of illustration, correspond to the three walking bouts. The characteristics of those bouts, including step rate, duration, and distance, were pooled together for the computation of actigraphy-derived metrics. 6MWT, 6 min walk test; BDI, Borg dyspnea index.

2.2 | Clinical Assessment of 6MWD, Borg Dyspnea Index and PAH-SYMPACT Physical Impact and Cognitive/Emotional Impact Domain Scores

6MWD and Borg dyspnea index were assessed at three site visits—at screening and at Weeks 16 and 24 [8] (Figure 1A). The PAH-SYMPACT electronic patient-reported outcome instrument, a PAH-specific questionnaire that assesses symptoms and impacts [21, 22], was completed at baseline, and at Weeks 15 and 23 using the study-provisioned smartphone. The seven physical impact items and the four cognitive/emotional impact domain questions (Supporting Information Table 1), with a 7-day recall period were completed at the end of the 7 days. Each item was scored using a 5-point Likert scale, ranging from 0 to 4, with higher scores indicating worse impact. The overall score of each domain was calculated as the average of the total items within the given domain.

2.3 | Collection of Actigraphy Data

Patients were provided with a lightweight (14 grams), wrist-worn, validated triaxial accelerometer (GT9X Link, ActiGraph LLC) [8, 23–25] at baseline (up to 14 days before randomization) (Figure 1A). Patients were asked to wear the accelerometer on the wrist of their nondominant hand for the duration of the study, 24 h per day. Wear and wake time were algorithmically defined [26–28]. The ActiGraph's proprietary algorithms processed raw acceleration data in epochs of 60 s and generated step counts per minute. Epoch data were uploaded daily via the study-provisioned smartphone. Patients were blinded to their activity data and were not actively encouraged to increase physical activity.

2.4 | Actigraphy-Derived Metrics

For each patient, walking bouts were identified within a 2-week analysis time window at baseline and before the 6MWT at Weeks 16 and 24. A single walking bout was defined as a continuous sequence of 1-min step counts (>0 steps) during wear time (Figure 1B). Each walking bout was characterized by its total duration (bout duration, minutes) and distance covered (bout distance, steps) as well as by a sequence of 1-min step counts (step count, steps/minute). For each patient, frequency distributions of each of these three walking bout characteristics were constructed by combining data from all walking bouts within the 2-week time window adjacent to a given 6MWT. For each characteristic, the frequency distribution was described by its mean, standard deviation, and 19 percentiles (from 5th to 95th percentile with an increment of 5%), resulting in 21 actigraphy-derived metrics per walking bout characteristic.

2.5 | Statistical Analyses

The relationship between the actigraphy-derived metrics and three anchors (6MWD, PAH-SYMPACT Physical Impact domain score and Borg dyspnea index) was assessed with

Pearson and Spearman correlation coefficients. To account for deviations from the normal distribution (Supporting Information Figure 1) and a nonlinear scale (Supporting Information Table 2), the PAH-SYMPACT Physical Impact domain score and Borg dyspnea index, as well as actigraphy-derived metrics, first underwent the Yeo-Johnson transformation followed by the z-score normalization before a correlation coefficient between a given anchor and a metric was computed. No transformation was performed when testing for the relationship between the 6MWD and actigraphy-derived metrics. The correlation coefficients were computed for each clinical visit separately.

Test-retest reliability analysis assessed the stability of each actigraphy-derived metric (estimated over the 2-week time window) between baseline (Days −14 to −1) and post-baseline (Days 1 to 14) (Figure 1A). Spearman correlation, Pearson correlation and intra-class correlation coefficients (ICC) were calculated between the estimates at the two time points for each metric separately.

Known-group validity of the actigraphy-derived metrics was evaluated by comparing the metrics between clinically distinct groups of patients. Specifically, patients were assigned to one of the two groups based on (1) WHO functional class (II *vs.* III), (2) Borg dyspnea index (< 3 [no to slight dyspnea] vs. ≥ 3 [moderate to severe dyspnea]; Supporting Information Tables 2), and (3) PAH-SYMPACT Physical Impact domain score (< 1 [no difficulty in physical domain due to a disease] vs. ≥ 1 [at least some difficulty in physical domain due to a disease]; Supporting Information Table 1). As the PAH-SYMPACT Cognitive/Emotional Impact domain score was expected to have only a weak (if any) and nonspecific relationship with exercise capacity in patients with PAH (Supporting Information Table 1), discriminant validity of the actigraphy-derived metrics was evaluated by comparing the metrics between patients with the PAH-SYMPACT Cognitive/Emotional Impact domain score of < 1 (no difficulty in cognitive/emotional domain due to a disease) versus ≥ 1 (at least some difficulty in cognitive/emotional domain due to a disease). The cut-offs were selected to balance out the number of patients between the two groups while preserving their clinical meaningfulness. Separation between the two groups of patients for each actigraphy-derived metric was evaluated using Cohen's d (difference between groups divided by pooled standard deviation; Supporting Material) whereby effect sizes > 0.8 are considered a large effect (0.2 small, 0.5 medium) [29]. Both known-group and discriminant validity were evaluated at baseline only. This prevented potential modulation of the findings by the treatment and/or placebo effect as well as by a decline in volume of the actigraphy data collected over time due to a decrease in patient compliance (Supporting Information Figure 2).

To test for sensitivity to change of the actigraphy-derived metrics, patients were classified as either responders or non-responders. Patients who improved their 6MWD from baseline to the end of treatment by ≥ 24.4 meters were deemed responders, while all other patients were classified as non-responders [30]. The cut-off was selected in an attempt to balance the number of patients classified as responders versus non-responders. Patient classification was performed regardless of the study arm. For each actigraphy-derived metric, changes from baseline to end of treatment were compared between the two groups of patients.

Unless otherwise stated, comparisons between any two continuous distributions were performed with the Mann-Whitney U test. All statistical analyses were carried out using custom written software in Python (version 3.8.8). All reported percentiles were estimated using linear interpolation. All reported p-values are two-sided without adjustment for multiple testing, with significance set at $p < 0.05$.

3 | RESULTS

3.1 | Patient Baseline Characteristics

In TRACE, 53 patients were randomized to receive selexipag and 55 were randomized to placebo [8]. Baseline characteristics were generally well balanced across the two groups [8] (Supporting Information Table 3). Mean 6MWD, Borg dyspnea index and PAH-SYMPACT Physical Impact domain score were similar for both groups at baseline and at the end of treatment [8] (Table 1). Total accelerometer wear time at baseline (measured during the 2-week time window) was also similar between the two study groups (approx. 336 vs 364 h, selexipag arm vs placebo arm; $p = 0.2$; Table 1). Regarding walking bouts during the 2-weeks at baseline, there was no significant difference in the total duration (mean ± standard deviation, selexipag arm: approx. 28 ± 22 h; placebo arm: 26 ± 18 h; $p = 0.65$) or total number (mean ± standard deviation, selexipag arm: 822.5 ± 530.9 bouts; placebo arm: 841.0 ± 533.8 bouts; $p = 0.99$) between the two arms. Step rate, bout duration, and bout distance were also similar between the two groups at each visit (Table 1, Supporting Information Figure 2). As patient compliance decreased over time, total accelerometer wear time, duration and number of walking bouts also decreased (Table 1). Yet the total duration and number of walking bouts normalized by the total wear time remained stable across visits. Specifically, patients on average walked 7.4%–7.8% of the total accelerometer wear time and made a walking bout every 25.3–26.1 min throughout the study.

3.2 | Relationship Between the Actigraphy-Derived Metrics and the Anchors of 6MWD, PAH-SYMPACT Physical Impact Total Score and Borg Dyspnea Index

The 6MWD positively and significantly correlated with all 21 examined actigraphy-derived metrics of step rate and bout distance at baseline, Week 16 and Week 24 (Table 2; Pearson's correlation coefficients range: 0.34–0.67, all $p < 0.005$). Significant and positive correlations were also observed between 6MWD and the majority of metrics of bout durations that could be computed at the three timepoints; correlations below the 45th percentile could not be computed as there was no variability in bout duration for the lower percentiles (Supporting Information Figure 2). Pearson's correlation coefficients and the direction of the correlations were consistent for all metrics.

TABLE 1 | Descriptive statistics on the anchors (6-min walk distance, Borg dyspnea index and PAH-SYMPACT physical impact domain score) and the characteristics of walking bouts at the three timepoints.

		Clinical visit					
		Baseline		Week 16		Week 24	
Outcome/characteristic	Study arm	*N*	Mean ± Std. dev.	*N*	Mean ± Std. dev.	> *N*	Mean ± Std. dev.
6MWD, meters	Treatment	53	453.1 ± 129.7	43	450.0 ± 144.0	> 48	464.3 ± 137.4
	Placebo	55	> 449.5 ± 98.9	53	454.4 ± 108.8	> 51	468.6 ± 109.6
	Both arms	108	> 451.2 ± 114.5	96	452.5 ± 125.1	> 99	466.5 ± 123.3
Borg dyspnea index[a]	Treatment	53	3.8 ± 2.4	43	4.0 ± 2.6	> 48	3.6 ± 2.3
	Placebo	55	> 2.8 ± 1.9	52	2.8 ± 1.7	> 51	3.0 ± 2.1
	Both arms	108	> 3.3 ± 2.2	95	3.3 ± 2.2	> 99	3.3 ± 2.2
PAH-SYMPACT physical impact total score[b]	Treatment	52	0.86 ± 0.78	37	1.07 ± 0.88	> 37	0.89 ± 0.74
	Placebo	53	> 0.94 ± 0.75	46	0.91 ± 0.78	> 45	0.83 ± 0.77
	Both arms	105	> 0.90 ± 0.76	83	0.98 ± 0.83	> 82	0.86 ± 0.76
Total accelerometer wear time, minutes	Treatment	53	20,181.5 ± 7471.1	43	18,409.4 ± 1536.5	> 48	17,093.2 ± 2540.9
	Placebo	55	> 21,830.8 ± 9002.1	53	17,467.9 ± 3752.4	> 51	16,367.6 ± 3698.7
	Both arms	108	> 21,021.4 ± 8289.2	96	17,889.6 ± 2995.4	> 99	16,719.4 ± 3195.2
Total duration of walking bouts, minutes	Treatment	53	1692.2 ± 1,342.2	43	1386.6 ± 632.7	> 48	1343.8 ± 647.1
	Placebo	55	> 1586.5 ± 1,075.1	53	1272.0 ± 638.1	> 51	1128.2 ± 603.8
	Both arms	108	> 1638.4 ± 1,209.0	96	1323.3 ± 634.9	> 99	1232.7 ± 631.3
Total number of walking bouts	Treatment	53	822.5 ± 530.9	43	685.2 ± 249.4	> 48	670.9 ± 263.1
	Placebo	55	> 841.0 ± 533.8	53	687.8 ± 315.8	> 51	612.5 ± 289.6
	Both arms	108	> 831.9 ± 530.0	96	686.6 ± 286.5	> 99	640.8 ± 277.3
Mean step rate, steps/minute	Treatment	53	32.82 ± 9.61	43	33.59 ± 11.04	> 48	32.65 ± 10.13
	Placebo	55	> 33.01 ± 8.53	53	32.91 ± 8.92	> 51	32.68 ± 9.17
	Both arms	108	> 32.91 ± 9.03	96	33.22 ± 9.88	> 99	32.66 ± 9.60
Mean bout duration, minutes	Treatment	53	1.94 ± 0.44	43	1.94 ± 0.49	> 48	1.93 ± 0.45
	Placebo	55	> 1.84 ± 0.37	53	1.84 ± 0.36	> 51	1.80 ± 0.34
	Both arms	108	> 1.89 ± 0.41	96	1.89 ± 0.43	> 99	1.86 ± 0.40
Mean bout distance, steps	Treatment	53	67.36 ± 33.42	> 43	> 70.21 ± 40.84	48	66.99 ± 34.82
	Placebo	55	63.45 ± 29.22	53	63.14 ± 29.20	51	61.61 ± 28.06
	Both arms	108	65.37 ± 31.27	96	66.31 ± 34.88	99	64.22 ± 31.47

Note: The characteristics related to walking bouts were first computed for each patient and clinical visit separately, and then aggregated across patients. *N* indicates the number of patients.
Abbreviations: 6MWD, 6 min walk distance; Std. Dev, standard deviation.
[a]Borg dyspnea index is rated 0–10, with 0 indicating no difficulty in breathing and 10 indicating maximal difficulty in breathing.
[b]PAH-SYMPACT scores were assessed by using a five-point Likert scale, ranging from 0 to 4, with higher scores indicating worse impact.

Figure 2 shows the correlations between 6MWD and three example actigraphy-derived metrics of step rate (standard deviation, median and 95th percentile) at baseline, Week 16 and Week 24 (see also Supporting Information Figures 3 and 4 for the correlations with similar example metrics of bout distance and duration).

The PAH-SYMPACT physical impact domain score was significantly and negatively associated with all 21 metrics of step rate and bout distance at each time point (Supporting Information Table 4; Pearson's correlation coefficients range: −0.32 to −0.51; all $p < 0.05$). Most negative correlations with metrics of bout durations that could be computed were also significant (mean, median and ≥ 55th percentile). Negative correlations were also observed between the Borg dyspnea index and all actigraphy-derived metrics, with the majority reaching significance (Supporting Information Table 5). The consistency of these negative correlations between both the PAH-SYMPACT Physical Impact domain score and Borg dyspnea index and example metrics of step rate are demonstrated in Supporting Information Figures 5 and 6 (see also Supporting Information Figures 7–10 for similar example metrics of bout distance and duration).

Similar results for all three anchors were obtained when using Spearman correlation coefficients (Supporting Information Table 4–6). In addition, similar correlations were reported between the PAH-SYMPACT Physical Impact domain score and Borg dyspnea index and actigraphy-derived metrics when

TABLE 2 | Relationship between the 6-min walk distance and *non-transformed* actigraphy-derived metrics, as assessed with *Pearson* correlation coefficients.

Actigraphy-derived metric	Step rate, steps/minute			Bout distance, steps			Bout duration, minutes		
	Baseline	Week 16	Week 24	Baseline	Week 16	Week 24	Baseline	Week 16	Week 24
5th percentile	**0.55*** **	**0.60*** **	**0.55*** **	**0.48*** **	**0.58*** **	**0.52*** **	NA	NA	NA
10th percentile	**0.60*** **	**0.63*** **	**0.60*** **	**0.51*** **	**0.42*** **	**0.56*** **	NA	NA	NA
15th percentile	**0.65*** **	**0.61*** **	**0.60*** **	**0.58*** **	**0.44*** **	**0.60*** **	NA	NA	NA
20th percentile	**0.65*** **	**0.52*** **	**0.60*** **	**0.59*** **	**0.46*** **	**0.59*** **	NA	NA	NA
25th percentile	**0.67*** **	**0.53*** **	**0.58*** **	**0.59*** **	**0.40*** **	**0.57*** **	NA	NA	NA
30th percentile	**0.64*** **	**0.49*** **	**0.59*** **	**0.55*** **	**0.40*** **	**0.55*** **	NA	0.03	NA
35th percentile	**0.64*** **	**0.50*** **	**0.59*** **	**0.54*** **	**0.41*** **	**0.56*** **	NA	0.03	NA
40th percentile	**0.63*** **	**0.51*** **	**0.58*** **	**0.52*** **	**0.39*** **	**0.57*** **	NA	0.03	NA
45th percentile	**0.64*** **	**0.53*** **	**0.56*** **	**0.52*** **	**0.39*** **	**0.57*** **	0.13	0.01	NA
50th percentile	**0.63*** **	**0.53*** **	**0.55*** **	**0.53*** **	**0.34*** **	**0.57*** **	**0.21***	0.07	0.12
55th percentile	**0.62*** **	**0.55*** **	**0.56*** **	**0.53*** **	**0.35*** **	**0.56*** **	**0.23***	**0.21***	0.16
60th percentile	**0.62*** **	**0.56*** **	**0.55*** **	**0.53*** **	**0.34*** **	**0.56*** **	**0.35*** **	**0.45*** **	**0.40*** **
65th percentile	**0.60*** **	**0.56*** **	**0.56*** **	**0.53*** **	**0.35*** **	**0.56*** **	**0.38*** **	**0.56*** **	**0.56*** **
70th percentile	**0.59*** **	**0.57*** **	**0.54*** **	**0.52*** **	**0.36*** **	**0.57*** **	**0.53*** **	**0.49*** **	**0.53*** **
75th percentile	**0.59*** **	**0.60*** **	**0.54*** **	**0.53*** **	**0.39*** **	**0.57*** **	**0.50*** **	**0.44*** **	**0.49*** **
80th percentile	**0.58*** **	**0.61*** **	**0.55*** **	**0.55*** **	**0.44*** **	**0.54*** **	**0.49*** **	**0.49*** **	**0.56*** **
85th percentile	**0.58*** **	**0.61*** **	**0.55*** **	**0.54*** **	**0.48*** **	**0.54*** **	**0.49*** **	**0.45*** **	**0.53*** **
90th percentile	**0.57*** **	**0.61*** **	**0.60*** **	**0.55*** **	**0.47*** **	**0.52*** **	**0.59*** **	**0.56*** **	**0.61*** **
95th percentile	**0.57*** **	**0.65*** **	**0.65*** **	**0.52*** **	**0.53*** **	**0.49*** **	**0.52*** **	**0.56*** **	**0.54*** **
Std. dev	**0.58*** **	**0.64*** **	**0.64*** **	**0.56*** **	**0.56*** **	**0.54*** **	**0.55*** **	**0.56*** **	**0.54*** **
Mean	**0.63*** **	**0.64*** **	**0.61*** **	**0.62*** **	**0.59*** **	**0.60*** **	**0.59*** **	**0.60*** **	**0.61*** **

Note: Cells include correlation coefficients between the 6MWD and actigraphy-derived metrics computed at each clinical visit separately. The data were pooled across the two study arms. *p-value < 0.05, **p-value < 0.01, ***p-value < 0.005, where all p-values are two-sided. Significant correlations (p-value < 0.05) are highlighted in bold. Abbreviations: NA, not available; Std. dev, standard deviation.

no transformation was performed (Supporting Information Tables 7 and 8).

3.3 | Test-Retest Reliability of the Actigraphy-Derived Metrics

There was a strong and significant agreement between the actigraphy-derived metrics of step rate obtained over a 2-week time window at baseline versus post-baseline (Table 3 and Figure 3). Specifically, the strength of agreement, as assessed by ICC, varied between 0.78 and 0.88 for all examined actigraphy-derived metrics of step rate, demonstrating good test-retest reliability. Similarly, all examined actigraphy-derived metrics of bout distance agreed well between the two periods, with ICC varying between 0.76 and 0.90 (Table 3, Supporting Information Figure 11). As bout duration showed no variability across patients (mean ± standard deviation = 1 ± 0 min) below the 45th percentile, comparison of bout duration between the two periods was infeasible for the lower percentiles. All other metrics of bout duration demonstrated a high reproducibility between the two periods, with the ICC ranging from 0.68 to 1.0 (Table 3, Supporting Information Figure 12). Pearson and Spearman correlation coefficients were in good agreement with the ICC (Supporting Information Table 10).

3.4 | Known-Group and Discriminant Validity of the Actigraphy-Derived Metrics

All 21 actigraphy-derived metrics of step rate (Figure 4A) and bout distance (Figure 4B) differed significantly between patients in WHO functional class II ($N = 74$ patients) and WHO functional class III (N = 34 patients; all $p < 0.005$). Cohen's d evaluating separation between the two groups of patients ranged from 0.9 to 1.3 and from 1.0 to 1.3 across the metrics of step rate and bout distance, respectively (Table 4). Similar findings were obtained for mean and standard deviation (both $p < 0.005$) as well as for the 55th ($p < 0.01$) and higher percentiles (all $p < 0.005$) of bout duration, with Cohen's d varying between 0.6 and 1.2 (Table 4).

Patients with no to slight dyspnea, as judged from the Borg dyspnea index, walked significantly faster (N = 40 patients; step rate: all metrics $p < 0.01$; Figure 4D) and covered significantly longer distances in a single walking bout (bout distance: all

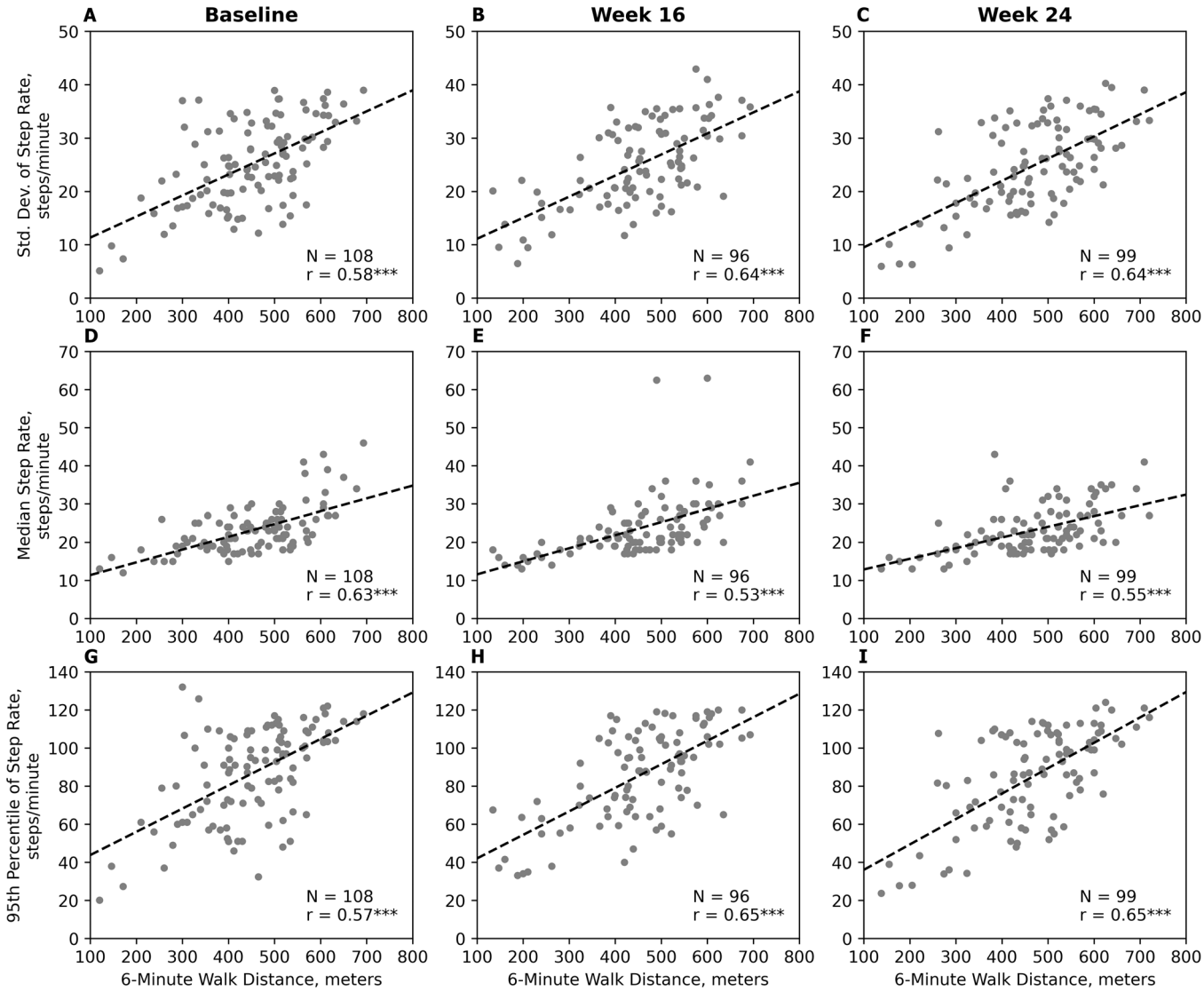

FIGURE 2 | Comparison between the 6-min walk distance and example actigraphy-derived metrics of step rate. (A, B, C) Standard deviation of step rate. (D, E, F) 50th percentile of step rate. (G, H, I) 95th percentile of step rate. The metrics were computed at baseline in A, D, G, at Week 16 in B, E, H, and at Week 24 visit in C, F, I. The dots in all scatter plots represent individual patients, with *N* denoting the total number. The dashed lines in each plot represent the best linear regression fits to the data presented. r corresponds to a Pearson correlation coefficient computed between the 6-min walk distance and a given actigraphy-derived metric, with the asterisks denoting the corresponding two-sided *p*-value: *< 0.05, **< 0.01, ***< 0.005. Std. Dev, standard deviation.

metrics $p < 0.005$; Figure 4E) than those experiencing moderate to severe dyspnea ($N = 68$ patients). Cohen's d varied between 0.5 and 0.8, and between 0.4 and 0.8 across the metrics of step rate and bout distance, respectively (Supporting Information Table 11). Similar trends were observed for the actigraphy-derived metrics of bout duration (Figure 4F), with Cohen's d varying 0.5 and 0.8 (Supporting Information Table 11).

In agreement with the findings above, all actigraphy-derived metrics of step rate and bout distance were significantly greater in patients experiencing no difficulty in physical domain due to PAH ($N = 64$ patients), as judged from the PAH-SYMPACT Physical Impact domain score, as compared to those experiencing at least some difficulty ($N = 41$ patients; all Cohen's d: 0.6–0.9; all $p < 0.005$; Supporting Information Figure 13A, B; Supporting Information Table 12). Similar trends were apparent across the actigraphy-derived metrics of bout duration (Supporting Information Figure 13C), with Cohen's d ranging from 0.4 to 1.0.

Patients experiencing no difficulty in cognitive/emotional domain due to PAH ($N = 62$ patients), as judged from the PAH-SYMPACT Cognitive/Emotional Impact domain score, revealed greater step rates than their peers reporting at least some difficulty in the same domain ($N = 43$ patients). This difference reached statistical significance only for several actigraphy-derived metrics of step rate (mean, standard deviation, the 55th-75th, 90th, and 95th percentiles: all $p < 0.05$; the 80th and 85th percentiles: both $p < 0.01$; Supporting Figure 13D), with Cohen's d for those metrics varying between 0.3 and 0.5 (Supporting Information Table 13). Only a handful of metrics of bout distance (mean, standard deviation, the 90th and 95th percentiles: $p < 0.05$; Cohen's d: 0.3–0.4; Supporting Information Figure 13E) and bout duration (standard deviation:

TABLE 3 | Test-retest reliability of the actigraphy-derived metrics, as assessed by *intra-class correlation* coefficients.

		Step rate, steps/minute		Bout duration, minutes		Bout distance, steps	
Statistic	**Period of time**	**Mean ± Std. dev.**	**ICC**	**Mean ± Std. dev.**	**ICC**	**Mean ± Std. dev.**	**ICC**
5th percentile	Baseline	9.8 ± 1.0	0.78	1.0 ± 0.0	NA	11.1 ± 1.1	0.80
	Post-baseline	9.9 ± 1.0		1.0 ± 0.0		11.1 ± 1.1	
10th percentile	Baseline	11.7 ± 1.2	0.86	1.0 ± 0.0	NA	12.5 ± 1.3	0.87
	Post-baseline	11.7 ± 1.2		1.0 ± 0.0		12.6 ± 1.2	
15th percentile	Baseline	13.0 ± 1.5	0.88	1.0 ± 0.0	NA	13.7 ± 1.6	0.88
	Post-baseline	13.0 ± 1.4		1.0 ± 0.0		13.8 ± 1.5	
20th percentile	Baseline	14.1 ± 1.7	0.88	1.0 ± 0.0	NA	14.9 ± 1.8	0.88
	Post-baseline	14.1 ± 1.6		1.0 ± 0.0		14.9 ± 1.7	
25th percentile	Baseline	15.1 ± 2.0	0.86	1.0 ± 0.0	NA	16.1 ± 2.2	0.90
	Post-baseline	15.2 ± 1.8		1.0 ± 0.0		16.2 ± 2.1	
30th percentile	Baseline	16.3 ± 2.5	0.84	1.0 ± 0.0	NA	17.5 ± 2.8	0.90
	Post-baseline	16.3 ± 2.2		1.0 ± 0.0		17.5 ± 2.7	
35th percentile	Baseline	17.6 ± 3.1	0.82	1.0 ± 0.0	NA	19.3 ± 3.6	0.89
	Post-baseline	17.5 ± 2.7		1.0 ± 0.0		19.2 ± 3.4	
40th percentile	Baseline	19.0 ± 3.8	0.81	1.0 ± 0.0	NA	21.4 ± 4.8	0.88
	Post-baseline	19.0 ± 3.4		1.0 ± 0.0		21.2 ± 4.4	
45th percentile	Baseline	20.9 ± 4.8	0.81	1.0 ± 0.1	1.00	24.0 ± 6.1	0.88
	Post-baseline	20.7 ± 4.3		1.0 ± 0.1		23.8 ± 5.6	
50th percentile	Baseline	23.1 ± 6.1	0.81	1.1 ± 0.2	0.71	27.2 ± 7.3	0.89
	Post-baseline	22.9 ± 5.6		1.0 ± 0.2		26.9 ± 6.9	
55th percentile	Baseline	25.8 ± 7.7	0.81	1.2 ± 0.4	0.68	30.8 ± 8.8	0.88
	Post-baseline	25.6 ± 7.2		1.2 ± 0.4		30.3 ± 8.1	
60th percentile	Baseline	29.1 ± 9.5	0.81	1.4 ± 0.5	0.74	35.2 ± 10.7	0.88
	Post-baseline	28.8 ± 9.1		1.4 ± 0.5		34.5 ± 9.7	
65th percentile	Baseline	33.1 ± 12.0	0.82	1.6 ± 0.5	0.79	40.4 ± 13.3	0.85
	Post-baseline	33.0 ± 11.5		1.6 ± 0.5		39.5 ± 11.9	
70th percentile	Baseline	38.3 ± 15.0	0.85	1.9 ± 0.4	0.74	47.1 ± 16.8	0.84
	Post-baseline	37.9 ± 14.3		1.8 ± 0.4		45.9 ± 14.8	
75th percentile	Baseline	44.3 ± 18.2	0.87	2.0 ± 0.4	0.70	56.1 ± 21.7	0.80
	Post-baseline	44.1 ± 17.8		2.0 ± 0.5		54.8 ± 19.1	
80th percentile	Baseline	51.6 ± 21.2	0.88	2.3 ± 0.6	0.75	68.3 ± 28.7	0.79
	Post-baseline	51.3 ± 21.3		2.2 ± 0.5		67.4 ± 25.8	
85th percentile	Baseline	60.2 ± 23.4	0.88	2.7 ± 0.7	0.68	88.3 ± 41.2	0.76
	Post-baseline	60.2 ± 23.5		2.6 ± 0.7		86.2 ± 36.9	
90th percentile	Baseline	71.6 ± 25.0	0.86	3.4 ± 0.9	0.76	124.5 ± 69.1	0.78
	Post-baseline	71.1 ± 24.9		3.4 ± 0.9		123.3 ± 65.0	
95th percentile	Baseline	86.6 ± 24.6	0.84	4.7 ± 1.7	0.73	218.0 ± 144.0	0.80
	Post-baseline	86.0 ± 25.3		4.7 ± 1.7		225.1 ± 154.3	
Std. dev	Baseline	25.2 ± 7.8	0.87	2.0 ± 1.2	0.86	148.0 ± 117.0	0.87
	Post-baseline	25.0 ± 8.1		2.0 ± 1.1		145.3 ± 111.5	
Mean	Baseline	32.9 ± 9.0	0.88	1.9 ± 0.4	0.84	65.4 ± 31.3	0.84

(Continues)

TABLE 3 | (Continued)

Statistic	Period of time	Step rate, steps/minute		Bout duration, minutes		Bout distance, steps	
		Mean ± Std. dev.	ICC	Mean ± Std. dev.	ICC	Mean ± Std. dev.	ICC
	Post-baseline	32.7 ± 9.0		1.9 ± 0.4		64.2 ± 29.2	

Note: Cells include intraclass correlation coefficients computed between the estimates of a given actigraphy-derived metric obtained at baseline versus post-baseline. The intra-class correlation coefficients were obtained with a two-way random effect model (absolute agreement) and computed using function "intraclass_corr" from package "pingouin" (version 0.5.3) in Python. ICC, intra-class correlation coefficient; NA, not available; Std. dev, standard deviation.

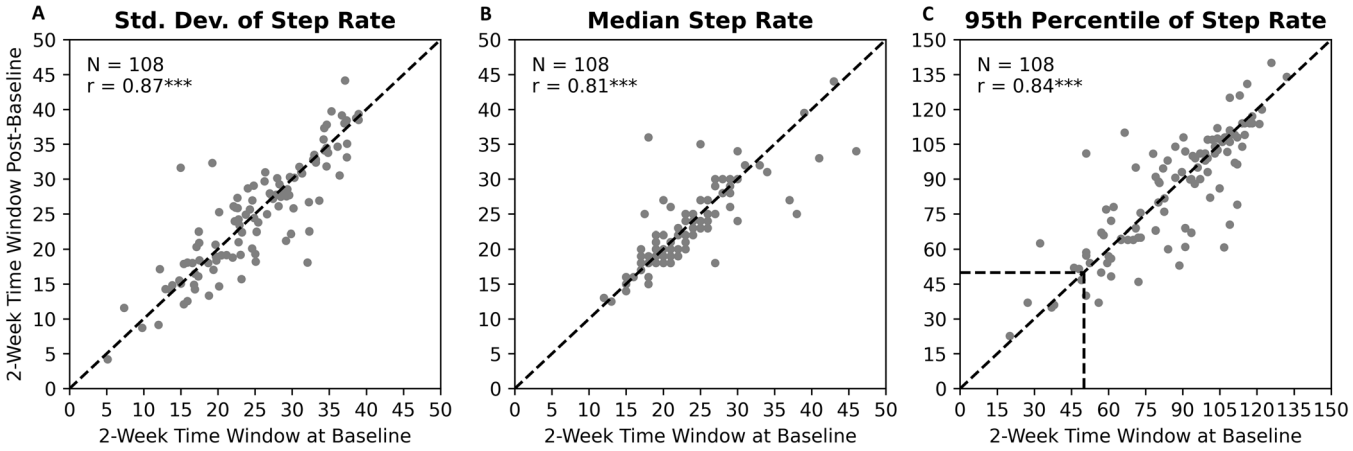

FIGURE 3 | Test-retest reliability of example actigraphy-derived metrics of step rate. (A) Standard deviation of step rate. (B) 50th percentile of step rate. (C) 95th percentile of step rate. The dashed black line in each panel corresponds to a diagonal. Note differences in the range of values between panels A, B (between 0 and 50) versus C (between 0 and 150). For the sake of comparison, the dashed lines at the bottom-left corner of panel C indicate the same range of values as in panels A and B. The dots in all scatter plots represent individual patients, with N denoting the total number. r corresponds to a Pearson correlation coefficient with the asterisks denoting the corresponding two-sided *p*-value: ***< 0.005. Std. Dev, standard deviation.

$p < 0.05$; Cohen's d = 0.3; Supporting Information Figure 13F) showed similar statistically significant differences between the two groups of patients.

3.5 | Sensitivity to Change

Change in the 6MWD from baseline to the end of treatment was significantly greater in responders ($N = 38$ patients [19 from the selexipag arm and 19 from the placebo arm]; mean ± standard deviation: 69.6 ± 47.5 meters) than in non-responders ($N = 61$ patients [29 from the selexipag arm and 32 from the placebo arm]; −16.6 ± 27.1 meters; $P < 10^{-16}$). Supporting Table 14 reports mean changes in the actigraphy-derived metrics from baseline to the end of treatment for responders and non-responders separately. Separation between the two groups tended to be greater along the metrics reflecting higher percentiles of the walking bout characteristics than along those reflecting the lower percentiles. In total, 16 of the 21 actigraphy-derived metrics of step rate demonstrated an increase from baseline to the end of treatment in responders, whereas no change or decline in the same metrics was observed in non-responders. Similar observations were made for several actigraphy-derived metrics of bout distance (85th -95th percentiles, mean, and standard deviation) and bout duration (65th, 85th −95th percentiles). Supporting Information Figure 14 shows cumulative distributions functions of changes in the 95th percentile of the three walking bout characteristics that showed the best separation between responders and non-responders. Remarkably, qualitatively similar results were obtained when applying a more conservative cut-off of 33 meters [5] (results not shown here).

4 | DISCUSSION

Our findings demonstrate that continuous and remote monitoring of physical activity by actigraphy can provide a thorough and unbiased evaluation of exercise capacity in PAH patients, potentially overcoming the limitations of momentary in-clinic measures such as the 6MWD. Here, we explored the construct validity, test-retest reliability and sensitivity to change of actigraphy-derived metrics as measures of exercise capacity in patients with PAH. Specifically, whether patient's walking bouts in everyday life, captured through wrist-worn accelerometers, reflect the physical exercise capacity of patients with PAH was examined. The data demonstrate that these metrics—step rate, walking bout distance and duration—correlate significantly with established and commonly used clinical measures such as the 6MWD, Borg dyspnea index, and the PAH-SYMPACT Physical Impact domain score. These correlations were observed consistently at multiple time events highlighting the robustness of the tested actigraphy-derived metrics. The metrics also exhibited high test-retest reliability, with ICCs greater than 0.70, indicating strong consistency over time. The ability of these metrics to distinguish between distinct levels of disease

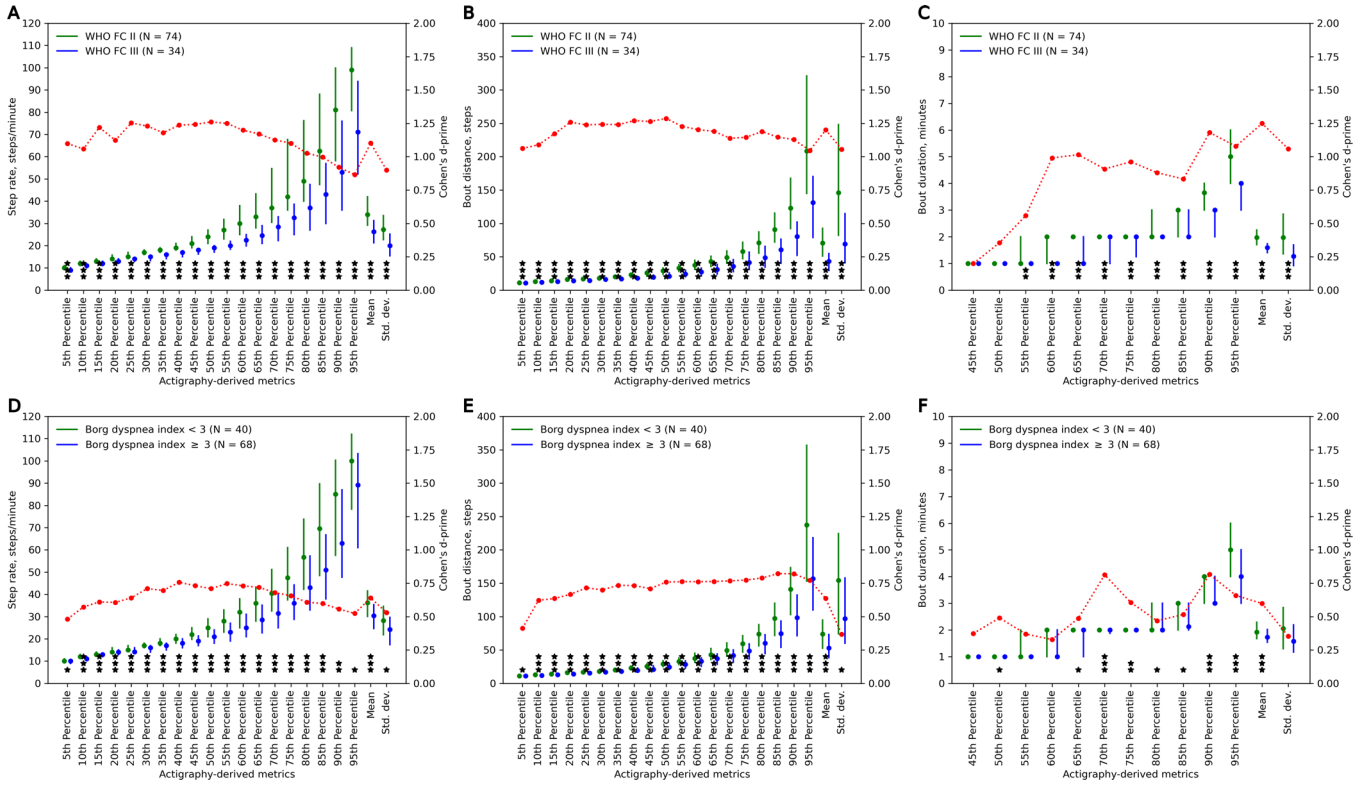

FIGURE 4 | Known-group validity of the actigraphy-derived metrics using the WHO functional class and Borg dyspnea index. (A and D) Step rate. (B and E) Bout distance. (C and F) Bout duration. Patients were assigned to one of the two distinct groups using either the WHO functional class in A, B, C, or the Borg dyspnea index in D, E, F. Data of patients in WHO functional class II and III in A, B, C are highlighted in green and blue, respectively. Similarly, data of patients experiencing no to slight and moderate to severe dyspnea in D, E, F are highlighted in green and blue, respectively. *N* indicates the number of patients in each of the two groups. Data of each actigraphy-derived metric were compared using both the Mann-Whitney U test, with the asterisks along the x-axis denoting the corresponding two-sided *p*-value (*< 0.05, **< 0.01, ***< 0.005), and Cohen's d highlighted in red. Std. Dev, standard deviation.

severity further supports their validity as indicators of exercise capacity in PAH patients.

Higher step rates, longer walking bouts, and greater distances covered during individual walking bouts were associated with greater exercise capacity, as indicated by the positive associations between the actigraphy-derived metrics and the 6MWD. These findings are of high importance as they demonstrate the utility of actigraphy-derived metrics as consistent and unbiased measurements that parallel the well-established and widely used 6MWT. The results are consistent with prior research that has used actigraphy to monitor physical activity in PAH patients, showing significant correlations with clinical outcomes such the 6MWD. For example, Hughes et al. reported a significant positive correlation (Pearson correlation coefficient 0.7) between the average daily step count and the 6MWD in patients with PAH ($N = 45$) [31]. Similarly, strong correlations (Pearson correlations coefficients 0.55–0.65) between the 6MWD and the average daily step count were reported in TRACE by Howard et al in patients with PAH ($N = 108$) [8]. In this latter study, the strength of the relationship between the 6MWD and time spent in non-sedentary and in moderate-to-vigorous physical activity, as judged from Pearson correlation coefficients, ranged between 0.31 and 0.38, and 0.31–0.41, respectively. In a further study by Lachant et al., [16] Spearman correlation coefficients computed between the 6MWD and average activity parameters (daily step count, daily activity time, light activity time, moderate activity time) in patients with PAH ($N = 30$) ranged from 0.44 to 0.59. In our analysis, significant Pearson correlation coefficients computed between the 6MWD and the examined actigraphy-derived metrics across the three different time points varied between 0.21 and 0.64, on par with the strength of associations reported above. Comparisons across studies regarding the strength of associations between the 6MWD and a given actigraphy-derived metric should however be interpreted with some caution. For example, differences across studies in the time periods used to estimate the metric or in the volume of data may result in different levels of variance in the metric which can impact on the strength of the association, as may differences in disease severity. Importantly, many correlation coefficients reported in our analysis were within the range (0.55–0.66) accepted by the European Medicines Agency for the qualification of the 95th percentile of stride velocity as the first digital actigraphy-based efficacy endpoint in studies of ambulatory Duchenne Muscular Dystrophy [15].

While the correlations between the actigraphy-derived metrics and conventional measures such as the 6MWD were significant, they were not perfect, and this is beneficial for several reasons. The 6MWD provides only a momentary assessment of patient's exercise capacity in a highly controlled environment, and as reflected in our data, we see that patients with PAH do not often

TABLE 4 | Comparison of the actigraphy-derived metrics between patients in *WHO functional class* II versus III for each of the three characteristics of walking bouts at randomization.

	Step rate, steps/minute			Bout distance, steps			Bout duration, minutes		
Actigraphy-derived metric	**FC II ($N = 74$)**	**FC III ($N = 34$)**	**Cohen's d**	**FC II ($N = 74$)**	**FC III ($N = 34$)**	**Cohen's d**	**FC II ($N = 74$)**	**FC III ($N = 34$)**	**Cohen's d**
5th percentile	10.0	9.0	**1.1*** **	11.0	10.9	**1.1*** **	1.0	1.0	NA
10th percentile	12.0	11.0	**1.1*** **	13.0	12.0	**1.1*** **	1.0	1.0	NA
15th percentile	13.0	12.0	**1.2*** **	14.0	13.0	**1.2*** **	1.0	1.0	NA
20th percentile	14.0	13.0	**1.1*** **	16.0	14.0	**1.3*** **	1.0	1.0	NA
25th percentile	15.0	14.0	**1.3*** **	17.0	14.5	**1.2*** **	1.0	1.0	NA
30th percentile	17.0	15.0	**1.2*** **	18.0	16.0	**1.2*** **	1.0	1.0	NA
35th percentile	18.0	16.0	**1.2*** **	20.0	17.0	**1.2*** **	1.0	1.0	NA
40th percentile	19.0	17.0	**1.2*** **	23.0	18.0	**1.3*** **	1.0	1.0	NA
45th percentile	21.0	18.0	**1.2*** **	26.0	19.5	**1.3*** **	1.0	1.0	0.2
50th percentile	24.0	19.0	**1.3*** **	29.0	21.0	**1.3*** **	1.0	1.0	0.4
55th percentile	27.0	20.0	**1.2*** **	33.0	24.0	**1.2*** **	1.0	1.0	**0.6** **
60th percentile	30.0	22.5	**1.2*** **	37.5	27.5	**1.2*** **	2.0	1.0	**1.0*** **
65th percentile	33.0	24.5	**1.2*** **	43.0	31.0	**1.2*** **	2.0	1.0	**1.0*** **
70th percentile	37.0	28.5	**1.1*** **	49.1	35.5	**1.1*** **	2.0	2.0	**0.9*** **
75th percentile	42.0	32.5	**1.1*** **	58.0	41.5	**1.1*** **	2.0	2.0	**1.0*** **
80th percentile	49.0	37.0	**1.0*** **	71.0	48.5	**1.2*** **	2.0	2.0	**0.9*** **
85th percentile	62.5	43.0	**1.0*** **	91.0	60.5	**1.1*** **	3.0	2.0	**0.8*** **
90th percentile	81.0	53.0	**0.9*** **	123.0	80.5	**1.1*** **	3.7	3.0	**1.2*** **
95th percentile	99.0	71.1	**0.9*** **	208.4	131.3	**1.0*** **	5.0	4.0	**1.1*** **
Std. dev	27.2	20.0	**0.9*** **	145.9	69.0	**1.1*** **	2.0	1.3	**1.1*** **
Mean	34.0	26.3	**1.1*** **	70.5	43.2	**1.2*** **	2.0	1.6	**1.2*** **

Note: Cells include medians of the actigraphy-derived metrics. *N* indicates the number of patients. *p-value < 0.05, **p-value < 0.01, ***p-value < 0.005, where all p-values are two-sided. Significant differences (p-value < 0.05) are highlighted in bold.
Abbreviations: FC II and FC III, WHO functional classes II and III; NA, not available; Std. dev, standard deviation.

walk for more than 3 min in one bout. On the contrary, actigraphy provides a continuous, objective, remote, quantifiable, ubiquitous, real-world (i.e., outside of a clinical environment) assessment of patient's ambulatory behavior and physical activity. This allows actigraphy to capture different dimensions of physical activity, reflecting daily variability in walking speeds, durations, and contexts that the 6MWD does not. For example, standard deviation of step rate effectively reflects variability in walking intensity, while mean bout distance is related to patient's endurance. This type of granular and comprehensive information on patient's physical activity can help detect subtle changes in disease progression or treatment response that may not be captured by traditional momentary assessments. The moderate correlations indicate that actigraphy offers complementary information to the 6MWT rather than simply replicating the test, including its limitations and weaknesses. Additionally, the actigraphy-derived metrics are less susceptible to individual factors such as adherence to the instructions of the 6MWT, willingness to cover the longest possible distance in 6 min, and effort exerted by patients during the 6MWT, thus potentially making them a more robust measure of daily physical activity. Neither does actigraphy interfere with patient's lifestyle and physical activity. The low burden and passive nature make actigraphy a well-suited candidate for remote monitoring, an attribute that proved valuable during the COVID-19 pandemic. Altogether, these specific characteristics of actigraphy can significantly reduce patient burden while increasing the richness of the data collected.

Several studies have explored the utility of actigraphy for monitoring exercise capacity in patients with pulmonary hypertension by estimating the time spent in non-sedentary activities each day [8, 13, 14]. To increase accuracy of such measurements, non-sedentary activities have been divided into subcategories (e.g., light, moderate, and vigorous activities) [32, 33] based on activity counts continuously generated by wearable actigraphy devices over 1-min adjacent and non-overlapping time epochs [12]. Despite being commonly accepted and used in clinical research, the measurements of time spent in non-sedentary activities suffers from several limitations. First, the thresholds used to categorize different activities are not tailored to the patient populations of interest. For example, thresholds suggested by Freedson et al. [32] were obtained in a healthy and young sample of individuals

(mean age = 23.9 years, $n = 50$), while Koster et al. [33] analyzed a larger and relatively older population (mean age = 78.4 years, $n = 62$) in which no study participant was reported to be diagnosed with PAH. Second, the thresholds across different studies have been obtained under a variety of conditions, for example, walking on a treadmill at different speeds [32] versus a free-living environment [33]. Third, different approaches have been adopted to delineate the boundaries between the activity categories. For example, Freedson et al. [32] linked activity counts to steady-state oxygen consumption in exercising individuals while Koster et al. [33] analyzed actigraphy data generated by two distinct actigraphy devices worn simultaneously by the study participants and used the thresholds established for one device to derive similar thresholds for the second device. Fourthly, definitions of different activity categories, as reported in the literature, do not necessarily reflect individual's perception of his/her exercise capacity. Lastly, the metrics are typically computed on each day separately and then aggregated across days over a pre-defined time period (which may vary from, for example, 7 days [34–36] to 14 days [8]). This approach requires specification of the minimum volume of data collected for each day to be deemed valid for aggregation, and imbalances in the volume of data collected within a day and across days may pose a significant challenge for data analysis and interpretation of the obtained results (e.g., comparison between mornings and evenings, weekdays and weekends).

The approach adopted in our study does not use any threshold. Moreover, it does not differentiate between different types of physical activities and does not aggregate across them. The approach solely focuses on a single but most prevalent type of physical activity human beings perform in their everyday lives, namely, walking. Furthermore, the approach described in our study does not aim to characterize the volume of walking bouts but instead the characteristics such as step rate, bout distance, and bout duration and, thus, is more representative of patients' physical activity than those that measure time spent in a given type of physical activity. This approach also does not introduce the requirement of a minimum volume of actigraphy and/or walking bout data collected per day [8, 35]. Rather, it assumes that a 2-week period is sufficient to construct the distributions of walking bout characteristics that accurately capture and represent patients' walking behavior in everyday life. Note, however, that testing this assumption explicitly warrants further research.

The negative correlations between the actigraphy-derived metrics and the PAH-SYMPACT Physical Impact domain score indicate that patients who report greater physical impacts caused by PAH tend to have lower levels of physical activity, as captured by actigraphy. This relationship emphasizes the utility of actigraphy in capturing the real-life physical limitations experienced by PAH patients, which might not be fully captured in highly controlled clinical settings. Similarly, the negative correlations between the actigraphy-derived metrics and the Borg dyspnea index further support the validity of the actigraphy-derived metrics. Patients experiencing higher levels of dyspnea tend to engage in less physical activity, which is accurately captured by lower step rates, shorter walking bouts, and reduced bout distances. This finding highlights the sensitivity of actigraphy metrics to changes in symptoms that affect daily physical activity levels.

With respect to test-retest reliability, we found strong reproducibility of the examined actigraphy-derived measures. Ensuring high reliability of the actigraphy-derived measures is crucial, as it verifies the device's capacity to reliably measure exercise capacity. Consistency is essential for monitoring disease progression and evaluating the effectiveness of therapies. The 2-week time period used to estimate patients' representative walking behavior aimed to capture regular activity patterns while reducing the impact of temporary changes. This time period was also deemed sufficient to offer a thorough assessment of daily activity.

There are some limitations in the analyses reported. First, as the data analyses were exploratory and no specific hypothesis was set a prior, no correction for multiple testing was performed. Even if performed, any choice of such a correction would remain questionable, as one would need to decide on the total number of tests performed. This number could have reflected the total number of actigraphy-derived measures associated either with a single characteristic of walking bouts (i.e., $N = 21$) or, alternatively, with the three of them combined (i.e., $n = 3 \times 21 = 63$). Secondly, correction for multiple tests assumes their independence when in fact, the three characteristics of the walking bouts and the actigraphy-derived measures of any given characteristic are not believed to be independent.

The three examined characteristics of walking bouts could have varied across different seasons. For example, Jehn et al. [37] demonstrated modulation of daily total step count and symptoms severity in adults with PAH by environmental factors. Specifically, daily total step count significantly decreased with an increase in ambient temperature and humidity, whereas the reverse was the case for symptoms severity in PAH. Patients in the TRACE study were enrolled in nine countries in the Western Europe and across different states in the USA throughout a year regardless of the season [8]. Although the collected actigraphy data could have been stratified and analyzed per season, variability in the environmental factors across the four seasons, ten countries and different geographies inside those countries would have likely artificially inflated variance in the examined actigraphy-derived measures. Potential analysis of the effect of seasonality on the obtained results was further complicated by the fact that the TRACE study was not designed to thoroughly test such an effect. Specifically, patients participated in TRACE for at maximum two seasons (i.e., 24 weeks), thus eliminating the opportunity to directly assess intra-patient changes in the volume and characteristics of walking bouts across the four seasons.

The characteristics of walking bouts were pooled across weekdays and weekends. Similarly, no distinction was made between different parts of a day (i.e., morning, afternoon, and evening) when analyzing the data. It is plausible that the volume and characteristics of walking bouts could have varied throughout a day and a week due to changes in daily routine and context [7, 35, 38]. For example, individuals typically walk more in the mornings and afternoons while fulfilling their professional responsibilities and duties as compared to the evenings when

they rest. Alternatively, individuals typically participate in social and community activities more frequently during the weekends than during the weekdays. To reduce variability in the volume and characteristics of walking bouts over time, the actigraphy-derived measures were estimated using a 2-week analysis time window. This choice agrees well with similar studies conducted previously [7, 8, 34, 36]. Given the good patient compliance in wearing the accelerometer in daily life in the TRACE study [8], the volume of collected actigraphy data within a 2-week analysis window was deemed sufficient to accurately estimate the characteristics of patient's walking bouts. Similarity of the findings obtained in the analysis of convergent validity across the three clinical outcome measures and three time points confirms the above assumption.

In agreement with a meta-analysis of data from 18 randomized clinical trials in PAH [39], male patients in the TRACE study demonstrated better performance in the 6MWT at baseline compared to their female counterparts (Supporting Information Table 17). However, this difference did not reach statistical significance. In contrast, patients in WHO functional class II significantly outperformed those in functional class III in the 6MWT (Supporting Information Table 17). Similar observations were made for the examined actigraphy-derived measures (Supporting Information Table 15), with the aforementioned effects being more pronounced in the higher percentiles. Furthermore, the 6MWT outcome was significantly modulated by patient age, with older patients performing worse than their younger counterparts (Supporting Information Figure 15). Given the significant correlations between the 6MWD and the examined actigraphy-derived measures (Table 2), the actigraphy-derived measures also significantly correlated with patient age (Supporting Information Table 16).

Despite the limitations, the work is aligned to the US Food and Drug Administration guidelines. As per the guidelines, proving construct validity, which includes known-group and discriminant validity, is essential for the approval of clinical outcome assessments [40]. Construct validity guarantees that the instrument precisely assesses the intended concept of interest. Here we demonstrate good construct validity in terms of both known-group and discriminant validity, which combined with convergent validity, confirm the effectiveness of actigraphy as a reliable and valid method for tracking exercise capacity in PAH patients. Our future work will look further into relating the findings to known clinical anchors and will focus on determining which metric should be selected, or in developing a composite score, for use as a potential endpoint in clinical trials.

5 | Conclusion

The results of this study demonstrate that actigraphy is an effective approach to continuous monitoring of exercise capacity in adult patients with PAH. Given a relatively low price of wrist-worn accelerometer devices, the ease of use, the continuous and ubiquitous nature of data collection, and the low burden imposed on patients, deployment of such devices in clinical practice has a potential to improve monitoring of patient's well-being over time, to accelerate clinical trial programs and to enable the conduct of decentralized clinical trials [13].

Author Contributions

Rana Zia Ur Rehman and **Dzmitry A. Kaliukhovich:** study design, data analysis, interpretation, writing of the article. **Federico Parisi, Noman Ashraf, Meenakshi Chatterjee, Nikolay V. Manyakov, Tarik Yardibi, Mona Selej, Tommaso Mansi, Oscar M. Carrasco-Zevallos,** and **Preston Dunnmon:** study design, interpretation, writing of the article. **Anna R. Hemnes** and **Robert P. Frantz:** interpretation, writing of the article. All authors approve the article and for it to be submitted via third party.

Acknowledgments

Medical writing support, including submission of the article, was provided by Jessica Beake, PhD, of Beake Medicom Ltd, UK, and was funded by Janssen Pharmaceutical Companies of Johnson & Johnson. RZUR is the guarantor of the manuscript and takes responsibility for the content of the manuscript, including the data and analysis.

Ethics Statement

Ethical approval was received from independent ethics committees/institutional review boards, and the study was conducted in compliance with the Declaration of Helsinki. All patients provided written informed consent.

Conflicts of Interest

RZUR, FP, NA, MC, NVM, TY, MS, TM, PD, OMCZ, and DAK were, at the time of study conduct and data analysis, employees of Janssen Pharmaceutical Companies of Johnson & Johnson and may own shares in the company. RPF reports personal fees from Janssen Pharmaceutical Companies of Johnson & Johnson, outside the submitted work; and Mayo Clinic may financially benefit from the algorithm described in future. ARH has no competing interests to report.

Data Availability Statement

The data-sharing policy of Janssen Pharmaceutical Companies of Johnson & Johnson is available at https://www.janssen.com/clinical-trials/transparency. As noted on this site, requests for access to the study data can be submitted through the Yale Open Data Access (YODA) Project site at http://yoda.yale.edu.

References

1. R. L. Benza, D. P. Miller, M. Gomberg-Maitland, et al., "Predicting Survival in Pulmonary Arterial Hypertension: Insights From the Registry to Evaluate Early and Long-Term Pulmonary Arterial Hypertension Disease Management (REVEAL)," *Circulation* 122 (2010): 164–172.

2. R. J. Oudiz, "Death in Pulmonary Arterial Hypertension," *American Journal of Respiratory and Critical Care Medicine* 188 (2013): 269–270.

3. S. Gu, H. Hu, and H. Dong, "Systematic Review of Health-related Quality of Life in Patients With Pulmonary Arterial Hypertension," *PharmacoEconomics* 34 (2016): 751–770.

4. M. Humbert, G. Kovacs, M. M. Hoeper, et al., "2022 ESC/ERS Guidelines for the Diagnosis and Treatment of Pulmonary Hypertension," *European Respiratory Journal* 61 (2023): 2200879.

5. J. Moutchia, R. L. McClelland, N. Al-Naamani, et al., "Minimal Clinically Important Difference in the 6-Minute-walk Distance for Patients With Pulmonary Arterial Hypertension," *American Journal of Respiratory and Critical Care Medicine* 207 (2023): 1070–1079.

6. N. B. Gabler, B. French, B. L. Strom, et al., "Validation of 6-Minute Walk Distance as a Surrogate End Point in Pulmonary Arterial Hypertension Trials," *Circulation* 126 (2012): 349–356.

7. S. Ulrich, M. Fischler, R. Speich, and K. E. Bloch, "Wrist Actigraphy Predicts Outcome in Patients With Pulmonary Hypertension," *Respiration* 86 (2012): 45–51.

8. L. S. Howard, S. Rosenkranz, R. P. Frantz, et al., "Assessing Daily Life Physical Activity by Actigraphy in Pulmonary Arterial Hypertension," *Chest* 163 (2023): 407–418.

9. A. Godfrey, R. Conway, D. Meagher, and G. OLaighin, "Direct Measurement of Human Movement by Accelerometry," *Medical Engineering & Physics* 30 (2008): 1364–1386.

10. A. Channa, N. Popescu, J. Skibinska, and R. Burget, "The Rise of Wearable Devices During the COVID-19 Pandemic: A Systematic Review," *Sensors* 21 (2021): 5787.

11. Digital Medicine Society. Library of Digital Endpoints. https://dimesociety.org/get-involved/library-of-digital-endpoints/ (2023).

12. A. Neishabouri, J. Nguyen, J. Samuelsson, et al., "Quantification of Acceleration as Activity Counts in ActiGraph Wearable," *Scientific Reports* 12 (2022): 11958.

13. S. D. Nathan, K. R. Flaherty, M. K. Glassberg, et al., "A Randomized, Double-blind, Placebo-controlled Study of Pulsed, Inhaled Nitric Oxide in Subjects at Risk of Pulmonary Hypertension Associated With Pulmonary Fibrosis," *Chest* 158 (2020): 637–645.

14. C. S. King, K. R. Flaherty, M. K. Glassberg, et al., "A Phase-2 Exploratory Randomized Controlled Trial of INOpulse in Patients with Fibrotic Interstitial Lung Disease Requiring Oxygen," *Annals of the American Thoracic Society* 19 (2022): 594–602.

15. European Medicines Agency. Qualification Opinion for Stride Velocity 95th Centile as Primary Endpoint in Studies in Ambulatory Duchenne Muscular Dystrophy Studies, https://www.ema.europa.eu/en/documents/scientific-guideline/qualification-opinion-stride-velocity-95th-centile-primary-endpoint-studies-ambulatory-duchenne-muscular-dystrophy-studies_en.pdf (2023).

16. D. Lachant, A. Light, M. Lachant, et al., "Peak Steps: Capacity for Activity Improves After Adding Approved Therapy in Pulmonary Arterial Hypertension," *Pulmonary Circulation* 13 (2023): e12285.

17. D. Lachant, A. Light, K. Hannon, F. Abbas, M. Lachant, and R. J. White, "Comparison of Chest- and Wrist-Based Actigraphy in Pulmonary Arterial Hypertension," *European Heart Journal - Digital Health* 3 (2022): 90–97.

18. O. Sitbon, R. Channick, K. M. Chin, et al., "Selexipag for the Treatment of Pulmonary Arterial Hypertension," *New England Journal of Medicine* 373 (2015): 2522–2533.

19. N. Galiè, M. Humbert, J.-L. Vachiery, et al., "2015 ESC/ERS Guidelines for the Diagnosis and Treatment of Pulmonary Hypertension," *European Respiratory Journal* 46 (2015): 903–975.

20. N. Galiè, M. Humbert, J. L. Vachiery, et al., "2015 ESC/ERS Guidelines for the Diagnosis and Treatment of Pulmonary Hypertension," *European Heart Journal* 37 (2016): 67–119.

21. K. M. Chin, M. Gomberg-Maitland, R. N. Channick, et al., "Psychometric Validation of the Pulmonary Arterial Hypertension-Symptoms and Impact (PAH-SYMPACT) Questionnaire," *Chest* 154 (2018): 848–861.

22. D. McCollister, S. Shaffer, D. B. Badesch, et al., "Development of the Pulmonary Arterial Hypertension-Symptoms and Impact (PAH-SYMPACT®) Questionnaire: A New Patient-Reported Outcome Instrument for PAH," *Respiratory Research* 17 (2016): 72.

23. J. Hwang, A. Fernandez, and A. Lu, "Application and Validation of Activity Monitors' Epoch Lengths and Placement Sites for Physical Activity Assessment in Exergaming," *Journal of Clinical Medicine* 7 (2018): 268.

24. K. M. Metcalf, B. I. Baquero, M. L. Coronado Garcia, et al., "Calibration of the Global Physical Activity Questionnaire to Accelerometry Measured Physical Activity and Sedentary Behavior," *BMC Public Health* 18 (2018): 412.

25. K. Valkenet and C. Veenhof, "Validity of Three Accelerometers to Investigate Lying, Sitting, Standing and Walking," *PLoS One* 14 (2019): e0217545.

26. R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin, "Automatic Sleep/Wake Identification From Wrist Activity," *Sleep* 15 (1992): 461–469.

27. R. P. Troiano, D. Berrigan, K. W. Dodd, L. C. Mâsse, T. Tilert, and M. Mcdowell, "Physical Activity in the United States Measured by Accelerometer," *Medicine and Science in Sports and Exercise* 40 (2008): 181–188.

28. C. Tudor-Locke, T. V. Barreira, J. M. Schuna, Jr., E. F. Mire, and P. T. Katzmarzyk, "Fully Automated Waist-Worn Accelerometer Algorithm for Detecting Children's Sleep-period Time Separate From 24-h Physical Activity or Sedentary Behaviors," *Applied Physiology, Nutrition, and Metabolism* 39 (2014): 53–57.

29. J. J. McGough and S. V. Faraone, "Estimating the Size of Treatment Effects: Moving Beyond p Values," *Psychiatry* 6 (2009): 21–29.

30. S. C. Mathai, M. A. Puhan, D. Lam, and R. A. Wise, "The Minimal Important Difference in the 6-minute Walk Test for Patients With Pulmonary Arterial Hypertension," *American Journal of Respiratory and Critical Care Medicine* 186 (2012): 428–433.

31. A. M. Hughes, J. Annis, A. R. Hemnes, et al., "Abstract 10474: Long-Term Daily Activity is Associated With Exercise Capacity and Quality of Life in Pulmonary Arterial Hypertension," *Circulation* 146 (2022): A10474-A10474.

32. P. S. Freedson, E. Melanson, and J. Sirard, "Calibration of the Computer Science and Applications, Inc. Accelerometer," *Medicine and Science in Sports and Exercise* 30 (1998): 777–781.

33. A. Koster, E. J. Shiroma, P. Caserotti, et al., "Comparison of Sedentary Estimates Between activPAL and hip- and Wrist-Worn Actigraph," *Medicine and Science in Sports and Exercise* 48 (2016): 1514–1522.

34. V. Mainguy, S. Provencher, F. Maltais, S. Malenfant, and D. Saey, "Assessment of Daily Life Physical Activities in Pulmonary Arterial Hypertension," *PLoS One* 6 (2011): e27993.

35. J. Minhas, H. Shou, S. Hershman, et al., "Physical Activity and its Association With Traditional Outcome Measures in Pulmonary Arterial Hypertension," *Annals of the American Thoracic Society* 19 (2022): 572–582.

36. S. Saxer, M. Lichtblau, C. Berlier, E. D. Hasler, E. I. Schwarz, and S. Ulrich, "Physical Activity in Incident Patients With Pulmonary Arterial and Chronic Thromboembolic Hypertension," *Lung* 197 (2019): 617–625.

37. M. Jehn, A. Gebhardt, U. Liebers, et al., "Heat Stress is Associated With Reduced Health Status in Pulmonary Arterial Hypertension: A Prospective Study Cohort," *Lung* 192 (2014): 619–624.

38. S. Sehgal, A. Chowdhury, F. Rabih, et al., "Counting Steps: A New Way to Monitor Patients With Pulmonary Arterial Hypertension," *Lung* 197 (2019): 501–508.

39. C. E. Ventetuolo, J. Moutchia, G. L. Baird, et al., "Baseline Sex Differences in Pulmonary Arterial Hypertension Randomized Clinical Trials," *Annals of the American Thoracic Society* 20 (2023): 58–66.

40. U.S Food and Drug Administration. Guidance for Industry. Patient-reported Outcome Measures: Use in Medical Product Development to Support Labeling Claim, https://www.fda.gov/media/77832/download (2009).

Supporting Information

Additional supporting information can be found online in the Supporting Information section.