



Published in final edited form as:

*Circulation*. 2012 July 17; 126(3): 349–356. doi:10.1161/CIRCULATIONAHA.112.105890.

## Validation of 6-Minute Walk Distance as a Surrogate End Point in Pulmonary Arterial Hypertension Trials

Nicole B. Gabler, PhD, MPH, MHA, Benjamin French, PhD, Brian L. Strom, MD, MPH, Harold I. Palevsky, MD, Darren B. Taichman, MD, PhD, Steven M. Kawut, MD, MS, and Scott D. Halpern, MD, PhD

Center for Clinical Epidemiology and Biostatistics and Department of Biostatistics and Epidemiology (N.B.G., B.F., B.L.S., S.M.K., S.D.H.); Leonard Davis Institute of Health Economics (B.F., B.L.S., S.D.H.); Pulmonary, Allergy, and Critical Care Division (H.I.P., D.B.T., S.M.K., S.D.H.); and Pulmonary Vascular Disease Program of the Penn Cardiovascular Institute (H.I.P., D.B.T., S.M.K., S.D.H.), Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

### Abstract

**Background**—Nearly all available treatments for pulmonary arterial hypertension have been approved based on change in 6-minute walk distance (6MWD) as a clinically important end point, but its validity as a surrogate end point has never been shown. We aimed to validate the difference in 6MWD against the probability of a clinical event in pulmonary arterial hypertension trials.

**Methods and Results**—First, to determine whether 6MWD between baseline and 12 weeks mediated the relationship between treatment assignment and development of clinical events, we conducted a pooled analysis of patient-level data from the 10 randomized placebo-controlled trials previously submitted to the US Food and Drug Administration ( $n=2404$  patients). Second, to identify a threshold effect for the 6MWD that indicated a statistically significant reduction in clinical events, we conducted a meta-regression among 21 drug/dose-level combinations.

6MWD accounted for 22.1% (95% confidence interval, 12.1%–31.1%) of the treatment effect ( $P<0.001$ ). The meta-analysis showed an average difference in 6MWD of 22.4 m (95% confidence interval, 17.4–27.5 m), favoring active treatment over placebo. Active treatment decreased the probability of a clinical event (summary odds ratio, 0.44; 95% confidence interval, 0.33–0.57). The meta-regression revealed a significant threshold effect of 41.8 m.

**Conclusions**—Our results suggest that 6MWD does not explain a large proportion of the treatment effect, has only modest validity as a surrogate end point for clinical events, and may not be a sufficient surrogate end point. Further research is necessary to determine whether the threshold value of 41.8 m is valid for long-term outcomes or whether it differs among trials using background therapy or lacking placebo controls entirely.

## Keywords

hypertension; pulmonary; meta-analysis; statistics; trials

Pulmonary arterial hypertension (PAH) is a progressive disease that leads to right-sided heart failure and death.<sup>1,2</sup> Seven drugs developed in the past 20 years have been shown to improve 6-minute walk distance (6MWD) in patients with PAH and have been approved for use in the United States on that basis. Although 6MWD is viewed by regulatory agencies as a clinically important end point in its own right, several studies have shown that patients who achieve greater improvements in 6MWD (or reach certain absolute values of 6MWD) have better clinical outcomes<sup>3–5</sup>; however, such observational data are insufficient to determine whether 6MWD is a valid surrogate end point for clinical events.<sup>6–9</sup>

Determining the validity of 6MWD as a surrogate end point in PAH trials is particularly timely in light of conflicting results of recent meta-analyses.<sup>10–13</sup> Although some of these contradictions may be related to inadequate sample size or follow-up or to intrinsic limitations of study-level meta-analyses, these differences also suggest the possibility that 6MWD is a poor surrogate. In the present study, we used 2 complementary approaches to validate 6MWD.<sup>14–17</sup> First, we used patient-level data from all available phase 3 randomized clinical trials submitted for drug approval to assess whether changes in 6MWD (Δ6MWD) mediate the relationship between treatment assignment and clinical outcomes. Second, we quantified how treatment effects on Δ6MWD predicted treatment effects on patient-centered outcomes, with the goal of determining whether a threshold Δ6MWD exists beyond which investigators could reliably predict that superior clinical outcomes would follow in future trials.

## Methods

### Study Population

Through a contract with the US Food and Drug Administration (FDA) to one author (S.D.H.), we obtained deidentified individual patient data for all participants in phase 3 placebo-controlled randomized trials submitted to the FDA through 2008 that tested prostanoids, endothelin receptor antagonists, or phosphodiesterase inhibitors. Eleven clinical trials examined 7 agents (ambrisentan, bosentan, sitaxsentan, iloprost, treprostinil, sildenafil, and tadalafil). We excluded 1 study (BREATHE-2, the Bosentan Randomized trial of Endothelin Antagonist Therapy for PAH) because it only included 33 participants and was not a phase 3 trial. Additional trial details are available elsewhere.<sup>18–25</sup> We purposely included trials of treatments that were not approved by the FDA (eg, sitaxsentan), a result of either unacceptable toxicity of effective dosages or dosages that were determined to be ineffective. Both the mediator and threshold analyses would be predictably biased if we only included FDA-approved treatments, and the resulting conclusions would not be useful for future trial design or regulatory decisions. All included trials reported similar methodology, including outcome assessment and variable collection at 12-week follow-up.

## Clinical Events

Clinical events included any of the following before the end of the trial: Death, lung transplantation, atrial septostomy, hospitalization because of worsening PAH, withdrawal for worsening right-sided heart failure, or addition of other PAH medications. We did not consider a deterioration in 6MWD to represent a clinical event because it was the surrogate we were attempting to validate. Further details are provided in online-only Data Supplement Table I.

## Six-Minute Walk Distance

Change in 6MWD was calculated as the difference, in meters, between the distance walked at baseline and 12 weeks. Baseline 6MWD was recorded at or within 2 weeks of randomization. In all analyses, patients who were missing a 12-week 6MWD because they died during the trial (n=45, 2%) were assigned a value of 0. We chose this value to reflect the fact that deaths are extremely important clinical events and to be consistent with the metric used in the majority of the trials included in our analysis. We used multiple imputation<sup>26</sup> to designate values for patients who survived but were nonetheless missing 12-week 6MWD (n=182, 8%). The imputation model included variables associated with clinical events: Baseline 6MWD, age, sex, weight, race, height, diagnosis category (idiopathic, connective tissue disease, HIV infection/anorexigen use, or congenital heart disease), 6MWD at 4- or 6-week follow-up, New York Heart Association (NYHA) functional classification, warfarin use, baseline sodium, cardiac output, and mean pulmonary arterial pressure. Five data sets were imputed. All imputations were completed in SAS version 9.2.

## Mediator Analysis

We used standard methodology to determine whether 6MWD mediates the relationship between treatment assignment and the development of clinical events at 12-week follow-up.<sup>15,16</sup> We defined treatment assignment as either active treatment or placebo. We conducted regression analyses to evaluate the following 4 hypotheses: (1) Treatment assignment has a significant effect on 6MWD from baseline to 12 weeks. (2) 6MWD has a significant effect on the odds of developing a clinical event. (3) Treatment assignment has a significant effect on the odds of developing a clinical event. (4) The effect of treatment assignment on the odds of developing a clinical event is attenuated when 6MWD is added to the model. It was necessary to reject the null for all 4 hypotheses to support 6MWD as a mediator/surrogate end point.

We used logistic or linear regression for binary or continuous outcomes, respectively. All regression models adjusted for study to account for study-level differences in treatment assignment and for baseline walk distance to account for patient-level differences in risk of clinical events. No other adjustments were made, because patients were randomly assigned to treatment or placebo.

After rejecting the null for the above 4 hypotheses, we determined the proportion of variability explained by the 6MWD in the relationship between treatment assignment and development of a clinical event.<sup>27,28</sup> We used a generalized linear model with a logit link to

quantify the relationship between treatment assignment and log odds of developing a clinical event, with adjustment for study and baseline walk (“reduced” model). We then added the 6MWD to the model (“full” model). The subsequent change in the treatment assignment coefficient between the reduced and full models provided the proportion of variability explained by 6MWD. Bootstrap resampling was used to create a confidence interval (CI) for the percent change. Estimates of percent change were obtained for each resampled data set, and the standard deviation of the estimates across 1000 resampled data sets was used as the standard error.<sup>29</sup>

In addition, we used a modified Sobel test to assess whether the amount of mediation was statistically significant; the modified test accounted for the fact that the surrogate (continuous) and the outcome (binary) were on different scales.<sup>16,30</sup> We evaluated the assumption of no effect modification between treatment and the mediator<sup>30</sup> by fitting a logistic regression model for clinical events with an interaction term between treatment assignment and 6MWD.

### Threshold Effect Analysis

We then conducted a trial-level meta-analysis and meta-regression to assess the relationship of the treatment effect on the mediator (6MWD) with the treatment effect on the odds of developing a clinical event at 12-week follow-up. This threshold analysis proceeded in 4 steps.

First, we estimated values for the exposure variable, placebo-adjusted study-level 6MWD, by conducting linear regressions within each trial. In these trial-specific patient-level regressions, the exposure variable was treatment (drug and dose), entered as indicator variables (with placebo as the referent). The outcome variable was 6MWD, and adjustment again was made for baseline 6MWD.

Second, we estimated values for the outcome, the placebo-adjusted study-level log odds of developing a clinical event between baseline and 12-week follow-up, using logistic regression analyses within each trial. In these trial-specific patient-level regressions, treatment (drug and dose, entered as indicator variables with placebo as the referent) was the exposure, clinical event (yes or no) was the outcome, and adjustment was made for baseline 6MWD. Exact logistic regression was used when the number of clinical events in a study was small.

Third, we used the 21 drug/dose combinations versus placebo across trials in a fixed-effects meta-regression that related estimated difference in 6MWD to the estimated log odds ratio (OR) for clinical events. The square of the inverse standard error was used as a weight to account for uncertainty in the estimated log OR. We determined the threshold effect by calculating 95% prediction bands around the meta-regression line. The threshold was calculated as the value of difference in 6MWD where the upper prediction band crossed the null value of 1.0 for relative odds of a clinical event. Prediction bands quantify the uncertainty in predicting the difference in clinical worsening in a single trial given a defined difference in 6MWD. Linearity of the final regression model was assessed via standard regression diagnostics.

Finally, we sought to determine whether study-level patient characteristics confounded the association between 6MWD and relative odds of clinical events. Potential confounders were chosen a priori on the basis of known differences in treatment response by race and sex<sup>31</sup> and by diagnosis and NYHA functional classification.<sup>32</sup> Therefore, covariates in our regression model were race (percentage black), sex (percentage female), PAH diagnostic category (percentage connective-tissue related), and NYHA functional classification (percentage class III or IV). If a potential confounder altered the coefficient for the treatment variable by 10%, it was retained in the final model.

We conducted 2 secondary analyses. First, we excluded the PHIRST (Pulmonary Arterial Hypertension and Response to Tadalafil) study, in which patients were permitted to use background therapy with other approved PAH-specific agents. Second, we removed patients who were NYHA class IV at randomization, because it is unlikely that these patients will be included in future clinical trials. All regressions and meta-regressions were conducted in R version 2.13 (R Development Core Team, Vienna, Austria).

The present study was determined to be exempt by the Institutional Review Board of the University of Pennsylvania (approval #814001). All coauthors had access to the study data, take responsibility for the analysis, and had authority over manuscript preparation and the decision to submit for publication.

## Results

The 10 trials included 2404 patients; 1563 (65%) were allocated to active treatment. Participants' median age was 50 years (range, 10–90 years), 22% were male, and 5% were black (Table 1). A total of 581 patients (24%) had a diagnosis of PAH caused by connective tissue disease, and 1349 (56%) were categorized as NYHA classification III or IV. Forty-five patients (2%) died between baseline and 12-week follow-up. An additional 153 patients (6%) experienced other clinical events; 83 of these did not have 12-week walk distance. Mean baseline walk distance was 341 m (SD, 85.7 m). Demographic, anthropometric, laboratory, and hemodynamic values were similar between groups defined by treatment allocation.

Characteristics of the 10 trials are presented in Table 2. Study-level percentages of female patients and patients diagnosed with connective tissue disease–related PAH were consistent across studies. Percentages for NYHA classification III/IV and black race showed more variation across studies.

### Does 6MWD Mediate the Relationship Between Treatment Assignment and Clinical Events?

The 4 criteria necessary to establish 6MWD as a mediator of the relationship between treatment assignment and development of a clinical event are listed in Table 3. For each, we found a statistically significant result in the required direction. First, assignment to active treatment versus placebo led to greater differences in the 6MWD (mean difference in 6MWD, 22.4 m; 95% CI, 15.9–28.9 m). Second, greater differences in 6MWD significantly reduced the odds of clinical events at 12-week follow-up (OR for a 10 m

increase in 6MWD, 0.89; 95% CI, 0.87–0.91). Third, assignment to active treatment versus placebo significantly decreased the relative odds of clinical events at 12 weeks (OR, 0.43; 95% CI, 0.31–0.59). Finally, the effect of treatment assignment on the development of a clinical event was attenuated with the addition of 6MWD to the model (OR, 0.52; 95% CI, 0.37–0.73). The proportion of the effect of treatment on the odds of developing a clinical event at 12 weeks that was explained by 6MWD was 22.1% (95% CI, 12.1%–31.1%). Additionally, the modified Sobel test confirmed the statistical significance of the mediation ( $Z=4.77$ ,  $P<0.001$ ). There was no significant interaction between treatment and 6MWD (estimate [95% CI], 0.002 [–0.002 to 0.006],  $P=0.25$ ).

### Does a Threshold Effect Exist for 6MWD?

Compared with placebo, nearly all drug/dose combinations resulted in a greater 6MWD at 12-week follow-up (Table 4). The summary results indicate an average difference in 6MWD of 22.4 m (95% CI, 17.4–27.5 m) for assignment to active treatment relative to placebo. The effect on reduction of clinical events was consistent across drug/dose combinations. Relative to placebo, all 21 drug/dose combinations lowered the odds of developing a clinical event at 12-week follow-up (summary OR, 0.44; 95% CI, 0.33–0.57).

The Figure illustrates the results of our meta-regression and threshold analysis. The upper prediction interval crossed the null value for relative odds of a clinical event at a difference in 6MWD of 41.8 m. This value indicates the minimal summary difference in 6MWD that corresponds to a statistically significant reduction in clinical events. Models that included race, sex, diagnosis category, and NYHA functional classification showed no evidence of confounding, and these variables were not included in the final meta-regression model.

In a sensitivity analysis, we removed the 4 drug/dose combinations from the PHIRST study, because this study was the only one to allow concomitant background therapy. The exclusion of this study resulted in a smaller threshold value of 25.7 m for the difference in 6MWD. In an additional analysis, removal of NYHA class IV patients ( $n=127$ ) did not appreciably change results. Further details are provided in the online-only Data Supplement.

## Discussion

This study provides the first rigorous examination of the validity of 6MWD from baseline to 12 weeks as a surrogate end point in trials of PAH therapies. 6MWD met all criteria as a mediator of the relationship between treatment and development of a clinical event at 12 weeks; however, the proportion of this relationship explained by 6MWD was modest at 22%, which suggests that 6MWD may not be an adequate surrogate. A threshold effect of 41.8 m was also identified, which means that if a drug improved 6MWD over 12 weeks by 41.8 m more than did placebo, investigators could predict, with 95% confidence, that the drug would reduce the clinical event rate over 12 weeks.

We found that only 4 of the 21 drug-dose combinations produced effects on 6MWD that could be said, with conventional degrees of certainty, to be associated with clinical improvements (Figure). If the lower threshold value of 25.7 m is used, then 5 of the

remaining 17 drug-dose combinations would be considered to produce statistically significant effects on clinical outcomes in the absence of background therapy. Of these 9 total drug-dose combinations that met the lower threshold value, 1 involved a drug that is not approved by the FDA (sitaxsentan) and 2 involved dosages that are not included in FDA labeling (sildenafil 40 mg 3 times per day and sildenafil 80 mg 3 times per day). Three drugs (iloprost, tadalafil, and treprostinil) did not meet either of the threshold values.

It is essential to explore the validity of surrogate end points, because valid surrogates provide efficient mechanisms for early-phase studies of new interventions. Specifically, trials that use validated surrogate end points can be conducted more quickly, with smaller sample sizes, fewer risks to subjects, and reduced research costs, than trials that use true clinical end points.<sup>33,34</sup> However, only if surrogate end points are validated will they clearly provide these virtues; in the absence of validation, there is considerable risk, as shown famously in the Cardiac Arrhythmia Suppression Trial (CAST),<sup>35</sup> of falsely concluding the effectiveness of a new intervention.

The present data also add to a growing body of literature on the use of 6MWD as an outcome measure in other settings. Using different methods, 6MWD has been evaluated in idiopathic pulmonary fibrosis,<sup>36</sup> chronic obstructive pulmonary disease,<sup>37</sup> and cardiac rehabilitation.<sup>38</sup> In the present study, we found that the proportion of the effect of treatment on preventing clinical events explained by the change in 6MWD was 22.1%, which falls well below the 50% to 75% threshold for a valid surrogate described by Freedman et al,<sup>27</sup> although some consider this an overly stringent criterion.<sup>15</sup> Clearly, PAH treatments have unmeasured effects on the outcome that are not fully captured by the change in 6MWD.<sup>39</sup> Thus, the finding of true but modest mediation by 6MWD suggests that it may not be sufficient to use on its own. Incorporation in a combination surrogate measure with hemodynamic or other assessments might improve its performance and warrants further study.

The threshold value we identified of 41.8 m for the difference in 6MWD may be considered for use as the level of improvement in 6MWD necessary to reliably conclude that the intervention will confer clinical benefits in future trials. Using different methods and patients from a single study (the SUPER study [Sildenafil Use in Pulmonary Arterial Hypertension]), Gilbert et al<sup>40</sup> estimated a minimally clinically important difference of 41 m that correlated with patient-reported improvement. Similarly, in untreated PAH patients, Paciocco et al<sup>41</sup> found that each increase in 50 m walked was associated with an 18% reduction in mortality. The consistency of findings across these studies using different methods supports the robustness of this result. However, 6MWD remains an inadequate surrogate end point given the modest degree of mediation of the treatment effect.

Confidence in the results of the present study stems from the large sample size used and the fact that the treatment effects on changes in 6MWD and on clinical events were consistent across all drug doses.<sup>42</sup> Nonetheless, the present study has limitations. First, as with any meta-analysis, the findings are subject to errors in the conduct, data entry, or analysis of the primary data. Second, these primary trials included mostly women and whites and used relatively short follow-up periods. We were unable to evaluate clinical end points that

occurred after 12 weeks, and our results should not be generalized beyond this time period. However, the demographics represented are consistent with the broader epidemiology of PAH, and the included trials all provided similar follow-up, clinical event definitions, and outcome measurement, which makes them well suited for our meta-analytic approach.

All of the trials that we examined were placebo controlled, which is an additional strength; however, 1 study allowed concomitant background therapy. Removal of the PHIRST trial reduced our threshold value, which suggests that future trials using background therapies may be subject to larger threshold values than those reported here. In other words, in the presence of an effective PAH therapy, a randomized controlled trial of a new therapy may need to produce larger differences in the 6MWD to provide confidence that the results correspond to differences in clinical outcomes. We were unable to further explore the role of background therapy because of small sample sizes. Future research, therefore, is needed, particularly in light of ethical questions surrounding the future conduct of placebo-controlled trials in PAH.<sup>43</sup>

Finally, we used statistical techniques to help address a clinical question. Although our threshold estimate provides a clear idea of what 6MWD would be needed to indicate that the benefits of a treatment will be greater than zero, the inference of a “clinically important” difference in outcomes might warrant selection of a different threshold. Individual patients may show clinical improvement without reaching a certain 6MWD threshold value, and patients who improve their 6MWD may not necessarily exhibit clinical improvement. However, the population-level threshold values we have established will be useful in the design of future clinical trials.

In conclusion, we used 2 complementary approaches to examine the validity of 6MWD as a surrogate end point in clinical trials of PAH therapies. We were able to identify significant threshold effects of 6MWD that can be used to guide future randomized controlled trials, and we found that 6MWD is a mediator in the relationship between treatment and clinical outcome. However, because 6MWD does not explain a large proportion of this treatment effect, it may not be sufficient to use it as a lone surrogate end point. Further studies are needed to identify combination surrogate end points that may have superior characteristics and to determine whether this threshold value we identified applies to trials that use background therapy or do not use placebo controls at all.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We are grateful to Maximilian Herlim and to Ziyue Liu, PhD, for their invaluable help preparing the data for analysis and to Drs Norman Stockbridge and Salma Lemtouni at the FDA for providing us with the data to conduct this study.

### Sources of Funding

This work was supported by an American Thoracic Society/Pfizer research grant in pulmonary hypertension (Dr Halpern). Dr Kawut was supported by K24 HL103844. Neither the FDA nor the funding source had a role in the



design of this study or in the decision to submit it for publication. The FDA did review the study before submission as a condition of the original contract but did not request any changes to our text.

#### Disclosures

Dr Gabler has participated in unrelated projects funded by Pfizer, Inc. Dr Strom has served as a consultant for Abbott, Amgen, AstraZeneca, BMS, Boehringer Ingelheim, GlaxoSmithKline, Novartis, NPS Pharma, Nuvo Research, Orexigen, Pfizer, Teva, and Vivus and has funded research from AstraZeneca, BMS, Pfizer, Shire, and Takeda. He has received contributions to Penn's pharmacoepidemiology training program from Abbott, Amgen, Hoffman, LaRoche, Novartis, Pfizer, Sanofi Pasteur, and Wyeth. Dr Palevsky has received consulting fees, advisory board fees, speaking fees, and/or research funding from Actelion, Bayer, GeNO, Gilead, GlaxoSmithKline, Pfizer, United Therapeutics, and Lung Rx. These roles in no way impacted on this analysis of the results of previously published studies. Dr Kawut has received consulting fees, advisory board fees, speaking fees, unrestricted educational grants, and/or research funding from Pfizer, Actelion, Bayer, Ikaria, Novartis, Merck, Gilead, United Therapeutics, and Lung Rx. Dr Taichman has received institutional research funding from Actelion.

## References

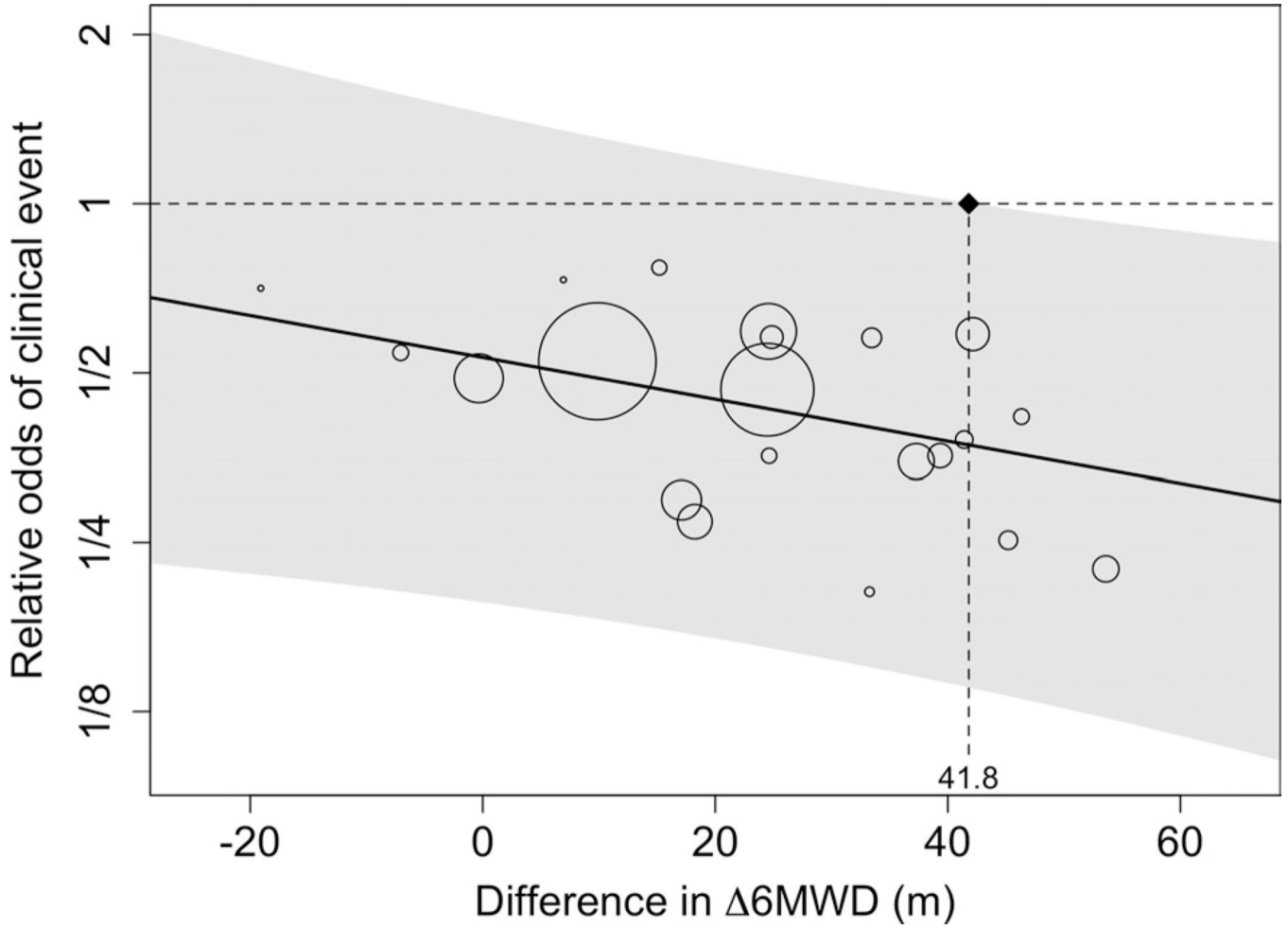
1. Taichman DB, Mandel J. Epidemiology of pulmonary arterial hypertension. *Clin Chest Med.* 2007; 28:1–22. [PubMed: 17338925]
2. Humbert M, Sitbon O, Chaouat A, Bertocchi M, Habib G, Gressin V, Yaici A, Weitzenblum E, Cordier JF, Chabot F, Dromer C, Pison C, Reynaud-Gaubert M, Haloun A, Laurent M, Hachulla E, Cottin V, Degano B, Jais X, Montani D, Souza R, Simonneau G. Survival in patients with idiopathic, familial, and anorexigen-associated pulmonary arterial hypertension in the modern management era. *Circulation.* 2010; 122:156–163. [PubMed: 20585011]
3. Provencher S, Sitbon O, Humbert M, Cabrol S, Jais X, Simonneau G. Long-term outcome with first-line bosentan therapy in idiopathic pulmonary arterial hypertension. *Eur Heart J.* 2006; 27:589–595. [PubMed: 16431875]
4. Sitbon O, Humbert M, Nunes H, Parent F, Garcia G, Herve P, Rainisio M, Simonneau G. Long-term intravenous epoprostenol infusion in primary pulmonary hypertension: prognostic factors and survival. *J Am Coll Cardiol.* 2002; 40:780–788. [PubMed: 12204511]
5. Miyamoto S, Nagaya N, Satoh T, Kyotani S, Sakamaki F, Fujita M, Nakanishi N, Miyatake K. Clinical correlates and prognostic significance of six-minute walk test in patients with primary pulmonary hypertension: comparison with cardiopulmonary exercise testing. *Am J Respir Crit Care Med.* 2000; 161:487–492. [PubMed: 10673190]
6. Baker SG, Kramer BS. A perfect correlate does not a surrogate make. *BMC Med Res Methodol.* 2003; 3:16. [PubMed: 12962545]
7. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med.* 1989; 8:431–440. [PubMed: 2727467]
8. Ventetuolo CE, Benza RL, Peacock AJ, Zamanian RT, Badesch DB, Kawut SM. Surrogate and combined end points in pulmonary arterial hypertension. *Proc Am Thorac Soc.* 2008; 5:617–622. [PubMed: 18625754]
9. Snow JL, Kawut SM. Surrogate end points in pulmonary arterial hypertension: assessing the response to therapy. *Clin Chest Med.* 2007; 28:75–89. [PubMed: 17338929]
10. Macchia A, Marchioli R, Marfisi R, Scarano M, Levantesi G, Tavazzi L, Tognoni G. A meta-analysis of trials of pulmonary hypertension: a clinical condition looking for drugs and research methodology. *Am Heart J.* 2007; 153:1037–1047. [PubMed: 17540207]
11. Helman DL, Brown AW, Jackson JL, Shorr AF. Analyzing the short-term effect of placebo therapy in pulmonary arterial hypertension: potential implications for the design of future clinical trials. *Chest.* 2007; 132:764–777. [PubMed: 17505027]
12. Galie N, Manes A, Negro L, Palazzini M, Bacchi-Reggiani ML, Branzi A. A meta-analysis of randomized controlled trials in pulmonary arterial hypertension. *Eur Heart J.* 2009; 30:394–403. [PubMed: 19155250]
13. Macchia A, Marchioli R, Tognoni G, Scarano M, Marfisi R, Tavazzi L, Rich S. Systematic review of trials using vasodilators in pulmonary arterial hypertension: why a new approach is needed. *Am Heart J.* 2010; 159:245–257. [PubMed: 20152223]

14. Johnson KR, Freemantle N, Anthony DM, Lassere MND. LDL-cholesterol differences predicted survival benefit in statin trials by the surrogate threshold effect (STE). *J Clin Epidemiol.* 2009; 62:328–336. [PubMed: 18834708]
15. Buyse M, Molenberghs G. Criteria for the validation of surrogate end-points in randomized experiments. *Biometrics.* 1998; 54:1014–1029. [PubMed: 9840970]
16. MacKinnon DP, Dwyer JH. Estimating mediated effects in prevention studies. *Eval Rev.* 1993; 17:144–158.
17. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat.* 2006; 5:173–186. [PubMed: 17080751]
18. Simonneau G, Barst RJ, Galie N, Naeije R, Rich S, Bourge RC, Keogh A, Oudiz R, Frost A, Blackburn SD, Crow JW, Rubin LJ. Continuous subcutaneous infusion of treprostinil, a prostacyclin analogue, in patients with pulmonary arterial hypertension: a double-blind, randomized, placebo-controlled trial. *Am J Respir Crit Care Med.* 2002; 165:800–804. [PubMed: 11897647]
19. Galie N, Olschewski H, Oudiz RJ, Torres F, Frost A, Ghofrani HA, Badesch DB, McGoon MD, McLaughlin VV, Roecker EB, Gerber MJ, Dufton C, Wiens BL, Rubin LJ. 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. [PubMed: 18506008]
20. Rubin LJ, Badesch DB, Barst RJ, Galie N, Black CM, Keogh A, Pulido T, Frost A, Roux S, Leconte I, Landzberg M, Simonneau G. Bosentan therapy for pulmonary arterial hypertension. *N Engl J Med.* 2002; 346:896–903. [PubMed: 11907289]
21. Olschewski H, Simonneau G, Galie N, Higenbottam T, Naeije R, Rubin LJ, Nikkho S, Speich R, Hoepfer MM, Behr J, Winkler J, Sitbon O, Popov W, Ghofrani HA, Manes A, Kiely DG, Ewert R, Meyer A, Corris PA, Delcroix M, Gomez-Sanchez M, Siedentop H, Seeger W. the Aerosolized Iloprost Randomized Study Group. Inhaled iloprost for severe pulmonary hypertension. *N Engl J Med.* 2002; 347:322–329. [PubMed: 12151469]
22. Barst RJ, Langleben D, Frost A, Horn EM, Oudiz R, Shapiro S, McLaughlin V, Hill N, Tapson VF, Robbins IM, Zwicke D, Duncan B, Dixon RAF, Frumkin LR. Sitaxsentan therapy for pulmonary arterial hypertension. *Am J Respir Crit Care Med.* 2004; 169:441–447. [PubMed: 14630619]
23. Barst RJ, Langleben D, Badesch D, Frost A, Lawrence EC, Shapiro S, Naeije R, Galie N. Treatment of pulmonary arterial hypertension with the selective endothelin-A receptor antagonist sitaxsentan. *J Am Coll Cardiol.* 2006; 47:2049–2056. [PubMed: 16697324]
24. Galie N, Ghofrani HA, Torbicki A, Barst RJ, Rubin LJ, Badesch D, Fleming T, Parpia T, Burgess G, Branzi A, Grimminger F, Kurzyna M, Simonneau G. Sildenafil Use in Pulmonary Arterial Hypertension Study Group. Sildenafil citrate therapy for pulmonary arterial hypertension. *N Engl J Med.* 2005; 353:2148–2157. [PubMed: 16291984]
25. Galie N, Brundage BH, Ghofrani HA, Oudiz RJ, Simonneau G, Safdar Z, Shapiro S, White RJ, Chan M, Beardsworth A, Frumkin L, Barst RJ. Tadalafil therapy for pulmonary arterial hypertension. *Circulation.* 2009; 119:2894–2903. [PubMed: 19470885]
26. Klebanoff MA, Cole SR. Use of multiple imputation in the epidemiologic literature. *Am J Epidemiol.* 2008; 168:355–357. [PubMed: 18591202]
27. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate end-points for chronic diseases. *Stat Med.* 1992; 11:167–178. [PubMed: 1579756]
28. Huang J, Huang B. Evaluating the proportion of treatment effect explained by a continuous surrogate marker in logistic or probit regression models. *Stat Biopharm Res.* 2010; 2:229–238. [PubMed: 20577652]
29. Efron, B.; Tibshirani, RJ. *An Introduction to the Bootstrap.* New York, NY: Chapman and Hall; 1993.
30. Jasti S, Dudley WN, Goldwater E. SAS macros for testing statistical mediation in data with binary mediators or outcomes. *Nurs Res.* 2008; 57:118–122. [PubMed: 18347484]

31. Gabler NB, French B, Strom BL, Liu Z, Palevsky HI, Taichman DB, Kawut SM, Halpern SD. Race and sex differences in response to endo-thelin receptor antagonists for pulmonary arterial hypertension. *Chest*. 2012; 141:20–26. [PubMed: 21940766]
32. Kuhn KP, Byrne DW, Arbogast PG, Doyle TP, Loyd JE, Robbins IM. Outcome in 91 consecutive patients with pulmonary arterial hypertension receiving epoprostenol. *Am J Respir Crit Care Med*. 2003; 167:580–586. [PubMed: 12446266]
33. Lassere MN. The biomarker-surrogacy evaluation schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. *Stat Methods Med Res*. 2008; 17:303–340. [PubMed: 17925313]
34. Rasekaba T, Lee AL, Naughton MT, Williams TJ, Holland AE. The six-minute walk test: a useful metric for the cardiopulmonary patient. *Intern Med J*. 2009; 39:495–501. [PubMed: 19732197]
35. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, Arensberg D, Baker A, Friedman L, Greene HL, Huther ML, Richardson DW. CAST Investigators. Mortality and morbidity in patients receiving encainide, flecainide, or placebo: the Cardiac Arrhythmia Suppression Trial. *N Engl J Med*. 1991; 324:781–788. [PubMed: 1900101]
36. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, Lancaster L, Noble PW, Sahn SA, Szwarcberg J, Thomeer M, Valeyre D, King TE Jr. Six-minute-walk test in idiopathic pulmonary fibrosis: test validation and minimal clinically important difference. *Am J Respir Crit Care Med*. 2011; 183:1231–1237. [PubMed: 21131468]
37. Redelmeier DA, Bayoumi AM, Goldstein RS, Guyatt GH. Interpreting small differences in functional status: the six minute walk test in chronic lung disease patients. *Am J Respir Crit Care Med*. 1997; 155:1278–1282. [PubMed: 9105067]
38. Hamilton DM, Haennel RG. Validity and reliability of the 6-minute walk test in a cardiac rehabilitation population. *J Cardiopulm Rehabil*. 2000; 20:156–164. [PubMed: 10860197]
39. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled. *Ann Intern Med*. 1996; 125:605–613. [PubMed: 8815760]
40. Gilbert C, Brown MC, Cappelleri JC, Carlsson M, McKenna SP. Estimating a minimally important difference in pulmonary arterial hypertension following treatment with sildenafil. *Chest*. 2009; 135:137–142. [PubMed: 18812447]
41. Paciocco G, Martinez FJ, Bossone E, Pielsticker E, Gillespie B, Rubenfire M. Oxygen desaturation on the six-minute walk test and mortality in untreated primary pulmonary hypertension. *Eur Respir J*. 2001; 17:647–652. [PubMed: 11401059]
42. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*. 2000; 1:49–67. [PubMed: 12933525]
43. Halpern SD, Doyle R, Kawut SM. The ethics of randomized clinical trials in pulmonary arterial hypertension. *Proc Am Thorac Soc*. 2008; 5:631–635. [PubMed: 18625756]

### CLINICAL PERSPECTIVE

This study shows that the change in 6-minute walk distance (6MWD) satisfies the statistical criteria as a mediator between drug therapy and clinical outcomes in randomized clinical trials. Thresholds in 6MWD change were identified such that if future drugs produced such changes, it could be inferred that these drugs would produce clinical effects as well. Higher thresholds in 6MWD may need to be used when testing new agents in the presence of background pulmonary arterial hypertension therapies; however, our results indicate that 6MWD is likely not adequate for use as a surrogate end point in pulmonary arterial hypertension clinical trials, because only modest proportions of the effects of drugs on true clinical outcomes are explained by changes in 6MWD. Further research is needed to identify more robust surrogate end points or combinations.



**Figure.** Results of the meta-regression analysis showing the relationship between changes in 6-minute walk distance (6MWD) between baseline and 12-week follow-up at the drug/ dose level by study on the odds of a clinical event at 12 weeks. The circles each represent a drug/ dose combination, with sizes proportionate to study weights (detailed in Table 4 and based on inverse variance weighting). The shaded gray area corresponds to the bounds of the 95% prediction intervals. The threshold value is indicated on the horizontal axis at 41.8 m.

**Table 1**

## Characteristics of Study Participants

Characteristic	Active Treatment (n=1563)	Placebo (n=841)
Age, y	50 (38–61)	49 (37–60)
Male, n (%)	335 (21)	192 (23)
Race, n (%)		
White	1244 (80)	678 (81)
Black	86 (6)	39 (5)
Other	221 (14)	118 (14)
Height, cm	163 (157–169)	163 (157–170)
Weight, kg	69.4 (59.0–82.1)	70.1 (60.6–83.9)
BMI, kg/m <sup>2</sup>	25.5 (22.5–30.0)	26.1 (22.9–30.4)
PAH diagnosis, n (%)		
Idiopathic	946 (62)	508 (62)
Connective tissue disease	388 (25)	193 (24)
HIV infection/anorexigen use	41 (3)	18 (2)
Congenital heart disease	155 (10)	97 (12)
NYHA functional classification, n (%)		
I/II	628 (41)	389 (47)
III/IV	917 (59)	432 (53)
Baseline hemodynamics		
Mean right atrial pressure, mm Hg	8.0 (5.0–12.0)	8.0 (5.0–12.0)
Mean pulmonary arterial pressure, mm Hg	52.0 (43.0–62.0)	54.0 (45.0–64.5)
Cardiac output, L/min	4.0 (3.2–5.1)	3.9 (3.2–4.9)
Cardiac index, L/min/m <sup>2</sup>	2.4 (1.9–3.1)	2.3 (1.9–3.1)
Pulmonary capillary wedge pressure, mm Hg	9.0 (6.0–12.0)	9.0 (6.0–12.0)
Pulmonary vascular resistance, Wood units	10.9 (7.1–16.8)	11.2 (7.5–16.0)
Baseline laboratory values		
Hemoglobin, g/dL	14.7 (13.4–16.0)	14.6 (13.3–16.0)
Sodium, mEq/L	140 (138–142)	140 (138–142)
Warfarin use, n (%)	860 (59)	470 (64)
Baseline 6MWD, m	356 (287–408)	352 (276–410)
Study, n (%)		
ARIES-1	134 (9)	67 (8)
ARIES-2	127 (8)	65 (8)
BREATHE-1	145 (9)	69 (8)
AIR	101 (6)	101 (12)

Characteristic	Active Treatment (n=1563)	Placebo (n=841)
SUPER	204 (13)	65 (8)
STRIDE-1	118 (8)	60 (7)
STRIDE-2	123 (8)	62 (7)
STRIDE-4	64 (4)	34 (4)
PHIRST	314 (20)	81 (10)
Treprostinil	233 (15)	237 (28)

BMI indicates body mass index; PAH, pulmonary arterial hypertension; HIV, human immunodeficiency virus; NYHA, New York Heart Association; 6MWD, 6-minute walk distance; ARIES, Ambrisentan in Pulmonary Arterial Hypertension, Randomized, Double-Blind, Placebo-Controlled, Multicenter, Efficacy Studies; AIR, Aerosolized Iloprost Randomized; BREATHE, Bosentan: Randomized Trial of Endo-thelin receptor Antagonist Therapy; STRIDE, Sitaxsentan To Relieve Impaired Exercise; SUPER, Sildenafil Use in Pulmonary Hypertension; and PHIRST, Pulmonary Arterial Hypertension and Response to Tadalafil.

Summaries provided as median (quartile 1–quartile 3) unless otherwise indicated by n (%).

**Table 2**

Characteristics of Participating Studies and Drug/Dose Combinations

Study and Drug/Dose	n	Clinical Events, n	Study-Level Statistics			
			% Female	% Black	% NYHA III/IV	% CTD
ARIES-1		10	84	7.3	65	31
Ambrisentan 5 mg	67					
Ambrisentan 10 mg	67					
Placebo	67					
ARIES-2		20	74	0	54	32
Ambrisentan 2.5 mg	64					
Ambrisentan 5 mg	63					
Placebo	65					
BREATHE-1		9	79	7.1	100	22
Bosentan 125 mg	75					
Bosentan 250 mg	70					
Placebo	69					
AIR		41	67	1.5	100	23
Iloprost	101					
Placebo	101					
STRIDE-1		7	79	8.1	67	24
Sitaxsentan 100 mg	55					
Sitaxsentan 300 mg	63					
Placebo	60					
STRIDE-2		8	77	13.1	63	29
Sitaxsentan 50 mg	61					
Sitaxsentan 100 mg	62					
Placebo	62					
STRIDE-4		2	84	6.7	39	15
Sitaxsentan 50 mg	32					
Sitaxsentan 100 mg	32					
Placebo	34					



Study and Drug/Dose	n	Clinical Events, n	Study-Level Statistics			
			% Female	% Black	% NYHA III/IV	% CTD
SUPER		17	76	2.6	61	30
Sildenafil 20 mg	68					
Sildenafil 40 mg	65					
Sildenafil 80 mg	71					
Placebo	65					
PHIRST		39	78	9.7	67	24
Tadalafil 2.5 mg	79					
Tadalafil 10 mg	79					
Tadalafil 20 mg	81					
Tadalafil 40 mg	75					
Placebo	81					
Treprostinil		45	81	5.3	7	19
Treprostinil	233					
Placebo	237					

NYHA indicates New York Heart Association; CTD, connective tissue disease; ARIES, Ambrisentan in Pulmonary Arterial Hypertension, Randomized, Double-Blind, Placebo-Controlled, Multicenter, Efficacy Studies; BREATHE, Bosentan: Randomized Trial of Endothelin receptor Antagonist Therapy; AIR, Aerosolized Iloprost Randomized; STRIDE, Sitaxsentan To Relieve Impaired Exercise; SUPER, Sildenafil Use in Pulmonary Hypertension; and PHIRST, Pulmonary Arterial Hypertension and Response to Tadalafil.

**Table 3**

Criteria to Establish Change in 6MWD as a Mediator in Relationship Between Treatment Assignment and Development of a Clinical Event at 12-Week Follow-Up

Criteria	Results*
Treatment assignment has a significant effect on 6MWD from baseline to 12-wk follow-up	Mean difference, 22.4 (95% CI, 15.9–28.9)
6MWD has a significant effect on the odds of developing a clinical event	OR, 0.89 per 10 m (95% CI, 0.87–0.91)
Treatment assignment has a significant effect on the odds of developing a clinical event at 12-wk follow-up	OR, 0.43 (95% CI, 0.31–0.59)
The effect of treatment assignment on the odds of developing a clinical event (compare with above) is attenuated with the addition of 6MWD to the model	OR, 0.52 (95% CI, 0.37–0.73)

6MWD indicates 6-minute walk distance; CI, confidence interval; and OR, odds ratio.

\* All models include adjustment for study and baseline walk; assignment to placebo is the reference for all models that included treatment.

**Table 4**

## Drug/Dose-Specific Placebo-Adjusted Results

Study and Drug/Dose	Difference in 6MWD, m (95% CI)	OR for Clinical Events (95% CI)	Meta-Regression Weight
ARIES-1			
Ambrisentan 5 mg	24.9 (0.32–49.4)	0.58 (0.13–2.53)	1.8
Ambrisentan 10 mg	41.4 (15.6–67.2)	0.38 (0.07–2.04)	1.4
ARIES-2			
Ambrisentan 2.5 mg	37.3 (8.2–66.3)	0.35 (0.11–1.12)	2.8
Ambrisentan 5 mg	53.6 (24.4–82.8)	0.22 (0.06–0.88)	2.1
BREATHE-1			
Bosentan 125 mg	33.4 (8.6–58.3)	0.58 (0.12–2.80)	1.5
Bosentan 250 mg	46.3 (21.1–71.6)	0.42 (0.07–2.47)	1.2
AIR			
Iloprost	24.5 (–2.4–51.3)	0.47 (0.23–0.96)	7.3
STRIDE-1			
Sitaxsentan 100 mg	33.3 (9.1–57.4)	0.20 (0.02–1.99)	0.7
Sitaxsentan 300 mg	24.6 (–0.45–49.7)	0.36 (0.06–2.15)	1.2
STRIDE-2			
Sitaxsentan 50 mg	–7.1 (–27.6–13.5)	0.54 (0.09–3.20)	1.2
Sitaxsentan 100 mg	15.2 (–4.7–35.1)	0.77 (0.13–4.72)	1.2
STRIDE-4			
Sitaxsentan 50 mg	–19.1 (–49.9–11.8)	0.71 (0–4.42)*	<0.1
Sitaxsentan 100 mg	6.9 (–23.5–37.5)	0.73 (0–2.07)*	<0.1
SUPER			
Sildenafil 20 mg	39.3 (15.0–63.6)	0.36 (0.09–1.47)	1.9
Sildenafil 40 mg	45.2 (20.6–69.8)	0.25 (0.05–1.28)	1.5
Sildenafil 80 mg	42.1 (18.0–66.3)	0.59 (0.17–1.98)	2.6
PHIRST			
Tadalafil 2.5 mg	–0.3 (–20.7–20.0)	0.49 (0.18–1.34)	3.8
Tadalafil 10 mg	17.1 (–4.3–38.5)	0.30 (0.10–0.90)	3.1
Tadalafil 20 mg	24.6 (5.1–44.1)	0.59 (0.23–1.51)	4.4
Tadalafil 40 mg	18.2 (–2.2–38.6)	0.27 (0.08–0.90)	2.7
Treprostinil			
Treprostinil	9.9 (–6.7–26.4)	0.52 (0.27–0.99)	9.2
Summary <sup>†</sup>	22.4 (17.4–27.5)	0.44 (0.33–0.57)	

6MWD indicates 6-minute walk distance; CI, confidence interval; OR, odds ratio; ARIES, Ambrisentan in Pulmonary Arterial Hypertension, Randomized, Double-Blind, Placebo-Controlled, Multicenter, Efficacy Studies; BREATHE, Bosentan: Randomized Trial of Endothelin receptor Antagonist Therapy; AIR, Aerosolized Iloprost Randomized; STRIDE, Sitaxsentan To Relieve Impaired Exercise; SUPER, Sildenafil Use in Pulmonary Hypertension; and PHIRST, Pulmonary Arterial Hypertension and Response to Tadalafil.

Reference group is the placebo group in each study. All analyses are adjusted for baseline walk distance.

\* Obtained from exact logistic regression.

<sup>†</sup>Obtained from fixed-effects meta-analysis ( $P$  for heterogeneity=0.99).