

Baseline Risk Evaluation Predicts Likelihood of Clinical Worsening: A Sample Size Re-analysis from Freedom-EV

R. James White¹, Youlan Rao², CQ Deng², Andrew Nelsen², Raymond L Benza³

1 University of Rochester Medical Center, Rochester NY; 2 United Therapeutics, Research Triangle Park, NC, 3Allegheny Health Network, Pittsburgh, PA

Background: Multiparametric risk analyses like the Reveal 2 or French non-invasive (French) risk scores predict outcomes for patients with pulmonary arterial hypertension (PAH). Freedom-EV was a global, event driven study of oral treprostinil in participants who had recently started monotherapy as treatment for PAH. In Freedom-EV, we prospectively planned to analyze the change in the French non-invasive risk score, and, *post-hoc*, we also analyzed the change in Reveal 2 score. For this *post-hoc* analysis, we hypothesized that participants who had clinical worsening (CW) events started with a higher baseline risk score than those who did not. We hypothesized that enriching a study selecting only those with higher risk scores would allow a smaller number of participants.

Methods: We calculated the French and Reveal 2 risk scores for four groups: active treatment with CW events, active without events, placebo with events, placebo without events. We used Fisher's exact analysis to compare categorical French risk scores and a non-parametric ANCOVA to analyze continuous Reveal 2 scores. We re-analyzed sample size enrolling only participants with 0 or 1 French low risk factors with an underlying hazard ratio of 0.625 (observed among higher risk participants in Freedom-EV) and a placebo median event time of 20 months. We assumed accrual during 36 months (total study duration 42 months) with a 19% active dropout rate and 10% placebo dropout.

Results: Both scores demonstrated statistically higher risk among those with CW events as compared to those without. Our analyses suggest that we would have 90% power to demonstrate the same hazard ratio with 191 CW events and 430 higher risk participants (French score of 0 or 1 low risk factors) as compared to the 205 CW events and 690 participants actually enrolled.

Conclusion: For event driven studies, enriching for those with higher baseline risk may be reasonable.

	Active-CW event	Active-no event	Placebo-CW event	Placebo-no event
Baseline Reveal 2 Median** [Min, Max]	7 [3, 13]	6 [2, 13]	7 [2, 12]	5 [1, 13]
N (for Reveal)	85	252	120	213
Baseline Number of French Low Risk Factors##				
0 Low Risk	31/85 (36%)	54/252 (21%)	35/120 (29%)	24/214 (11%)
1 Low Risk	36/85 (42%)	76/252 (30%)	50/120 (42%)	60/214 (28%)
2 Low Risk	15/85 (18%)	87/252 (34%)	20/120 (17%)	74/214 (35%)
All 3 Low Risk	3/85 (4%)	35/252 (14%)	15/120 (12%)	56/214 (26%)

** Reveal 2 difference between groups, ANCOVA, $p < 0.001$, ## French difference between groups, Fisher's exact, $p < 0.001$.