Efficacy and safety of inhaled alpha-1-antitrypsin by patients with severe alpha-1-antitrypsin deficiency and frequent exacerbations of Chronic Obstructive Pulmonary Disease.

Jan Stolk*, Naveh Tov, Kenneth R. Chapman, Pablo Fernandez, William MacNee, Nicholas S. Hopkinson, Eeva Piitulainen, Niels Seersholm, Claus F. Vogelmeier, Robert Bals, Gerry McElvaney, and Robert A. Stockley.

* Corresponding author

Dr. J. Stolk, Department of Pulmonology, Leiden University Medical Center, Leiden, NL

Email: j.stolk@lumc.nl

Supplementary Material

Study procedures:

Treatment began at the baseline/randomisation visit and continued for a period of 50 weeks. During this period, either AAT inhalation 80 mg in phosphate buffered saline, or 0·9% saline (placebo), was administered by eFlow® twice daily (i.e., a total daily dose of 160 mg of AAT or an equivalent volume of placebo, which was identical to the study drug but without AAT). Each session lasted for approximately 8 to 15 minutes. If the patient was receiving regular bronchodilator therapy, this was administered prior to inhalation of the study drug. Both the AAT and placebo were supplied in sterile, single-use glass vials containing 4 mL of a ready-to-use solution.

Study patients were allowed free use of concomitant medications for the treatment of the underlying disease, including antibiotics and steroids (inhaled and systemic), as well as any additional therapy approved by the principal investigator of the study site.

The eDiary consisted of daily questions for the patient designed to evaluate, among other parameters, change in concomitant medications, breathlessness (dyspnoea), sputum volume and colour, and well-being. The completed daily data were then automatically transmitted to a secure internet database via a transmission device. Scoring of Anthonisen symptoms in the electronic diary is that of the manual Bronkotest (Middlesex, UK). Dyspnea was scored on a 4-point scale (2= normal or usual for me, 3= worse than usual, 4= much worse than usual, and 1= better than usual), daily sputum volume was scored as none =(0), 1= (up to a teaspoonful), 2= (up to a tablespoonful),3= (up to an egg-cupful), 4= (more than an egg-cupful), and sputum color was assessed using the Bronkotest color chart (1–2 being mucoid and 3, 4, and 5 being increasing purulence). The daily diary printout was initially reviewed independently by two physicians (RAS and PF), and where any differences were observed the episode was reviewed jointly to obtain a consensus. Exacerbations were identified when one or more major Anthonisen criteria (AT) deteriorated from a baseline period of stability for more than 2 consecutive days.

The method for identifying the start and end of exacerbations was as those published by (Vijayasaratha and Stockley 2008). In addition, the duration of deterioration for each AT symptom from the patient's baseline or "usual" state during the exacerbation period was noted. Symptoms relapsing within 7 days of a previously identified exacerbation were judged to be part of the same exacerbation event.

Pulmonary function tests (PFTs) were performed at the screening visit (baseline), and then at weeks 2, 4, 20, and 50, as well as at termination (week 54), or at the time of early discontinuation if possible. Post bronchodilator FEV_1 and slow vital capacity (SVC) were measured by trained technicians with the patient in a sitting position, in compliance with the current ATS/ERS recommendations, and the highest value from three consecutive measurements was recorded in the CRF.

Serum samples obtained at baseline and weeks 2, 4, 20, 50, and 54, were analysed using a validated semi-quantitative electrochemiluminescence (ECL) bridging assay for the detection of AAT-reactive IgG antibodies (Charles River Laboratories Preclinical Services, Montreal, Canada) according to the guidelines. Briefly, serum IgG was purified on a Protein G spin column and then incubated with AAT conjugated to Ruthenium to permit formation of anti-drug antibodies (ADAs). The mixture was then added to ECL plates preincubated with AAT conjugated to biotin. Any ADAs cross-link to the ruthenium and biotin can be detected by ECL.

Outcomes

The primary outcome defined for the study was time from randomisation to the first moderate or severe event-based exacerbation. A moderate event-based exacerbation was defined as requiring a course of treatment with antibiotics and/or systemic corticosteroids and a severe exacerbation as an episode requiring hospitalisation. For patients taking routine antibiotics, any increase in the current dose of antibiotics or change of the type of antibiotics was deemed to indicate a moderate exacerbation.

The time to first moderate/severe and mild/moderate/severe exacerbation was derived from the patient's eDiary and updated after a blind review by two experts in the field (RAS and PF). The definition of a mild exacerbation was defined as requiring increased doses of inhaled long-acting β2 adrenergic agonist or inhaled dose of corticosteroids. They reconciled the eDiary data with the concomitant medications and reports of treatment emergent adverse events (TEAEs) in order to determine the start date and severity of the symptom-based exacerbation. Any course of treatment without an associated symptom change was deemed not to have been a true exacerbation.

Statistical analysis of spirometry (dropouts)

Multiple imputations were used to analyse the effect of study dropouts. In multiple imputations, a missing data value is imputed multiple times to account for the uncertainty in the imputed values, to generate multiple complete data sets which are then analysed according to the original method, and the results of analysis are combined to generate a single p-value. Imputations can be done in different ways, to represent different assumptions.

For example, under the Missing at Random (MAR) assumption, missing data are imputed based on the observed data according to the assigned arm, namely, missing data for placebo patients are imputed based on observed values for placebo patients and missing data for AAT patients are imputed based on observed values for AAT patients. There are different ways to implement multiple imputations under the Missing-Not-At-Random (MNAR) assumption. A conservative way is the Jump-to-Reference, which assumes that the effect of the drug is lost once a patient terminated the study and hence all missing data, regardless of the assigned arm, is imputed according to observed data of the placebo group. If the different analysis gives consistent results in terms of trends then the results are robust for missing data

Patients who were excluded from analysis had no FEV_1 measurement after randomization. Using an imputation method would actually increase the difference between the groups.

Safety outcomes

Adverse events were recorded according to GCP. An external Data Safety Monitoring Board was established to supervise the study through periodic reviews of blinded patient data. Members of the board included physicians with expertise in respiratory indications, specifically in AATD and clinical trials in the target population, as well as a biostatistician with expertise in clinical development. In addition, an analysis of AE relative rates was performed in order to evaluate whether the presence of ADAs was related to TEAEs. The FEV_1 and related spirometry results were measured per protocol as safety parameters.and were analysed post-hoc on the ITT population as efficacy endpoints under the assumption of missing-at-random. FEV_1 (ml) change from baseline at week 2, week 4, week 20, and week 50 were analysed using Mixed Model of Repeated Measures (MMRM), comparing the least squares means for the changes from baseline to week 50 and the overall effect between the two treatment groups with an unstructured covariance matrix. The treatment effect was adjusted for visit and the interaction between treatment and visit with the following parameters: FEV_1 at baseline, age, oxygen use, BMI, and SGRQ at baseline. Advanced statistical methods were used to address the

robustness of the FEV_1 results in light of the high dropout rate, including a conservative analysis which assumes that all dropouts (regardless of treatment group) behave like placebo after withdrawal (Jump-to-Reference) (Molenberghs and Kenward 2007).

Results:

Table S1 and S2 summarizes the number and rate of event-based exacerbations. Severe exacerbations were relatively infrequent, and there were few patients in either group who experienced a severe exacerbation. The number of patients with mild and moderate exacerbations and the rate of events were similar between the two arms of the study. There were 14 event-based exacerbations in 13 AAT treated patients (15.3% of the ITT population) and 9 events in 6 placebo patients (7.4% of the ITT population). The first exacerbation in the study was severe in 7 AAT patients (8.2% of the ITT population) but none of the severe event-based exacerbations recorded by the placebo patients was the first event. After the first event, the number of severe event-based exacerbations was similar: there were 7 severe event-based exacerbations recorded in 6 AAT treated patients and 9 exacerbations in 6 placebo treated patients. Details are shown in Table S2. As can be seen in Table S3, dyspnoea contributed to the majority of symptom-based exacerbations. This is the most frequent cardinal symptom experienced during the first symptom-based exacerbation. Moreover, dyspnoea was the main symptom that led the exacerbation to be treated, as can be seen in the type II and III exacerbations.

Table S1: Time to First Moderate/Severe Event-Based Exacerbation

Time to First Moderate/Severe Event-Based Exacerbation (days)	Intention to treat Population		Per protocol Population	
	AAT N = 85	Placebo N = 83	AAT N = 68	Placebo N = 69
N (%)	63 (74.1)	60 (72.2)	59 (79.4)	51 (73.9)
Mean (SD)	103.8 (86.9)	113.0 (84.2)	100.7 (82.5)	119.8 (84.6)
Median (min, max)	83 (8, 332)	100 (10, 367)	79.5 (8, 318)	104 (10, 367)
Kaplan-Meier Estimates				
25th , 75th percentile	40, 211	57, 282	38, 180	63, 282
Median (95% CI)	112 (76, 142)	140 (104, 174)	109 (72, 142)	140 (104, 191)
# Censored Observations	22	23	14	18
P value from Log Rank comparison between treatments adjusted for country	0.0952		0.0693	

Table S2: Exacerbations experienced during the study

Severity of Event Based Exacerbations	Number of Exacerbations per Patient												Yearly Rate (Mean)		
	AAT						Placebo						p value calculated using the pooled method		
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	AAT	Placebo	p-value
Mild	85	0·48	1·18	0	0	7	83	0·54	1·31	0	0	9	0·66	0·60	0·797
Moderate	85	1·61	1·75	0	1	8	83	1·70	1·61	0	1	7	2·08	1·95	0·706
Severe	85	0·16	0·40	0	0	2	83	0·11	0·41	0	0	2	0·38	0·11	0·045
Moderate/Severe	85	1·78	1·78	0	1	8	83	1·81	1·74	0	1	7	2·46	2·06	0·281
Mild/Moderate/Severe	85	2·26	2·19	0	2	9	83	2·35	2·38	0	2	11	3·12	2·67	0·312

Table S3: Severity of the First Symptom-Based Exacerbation

Severity of the First Exacerbation by Anthonisen Type	AAT N=85		Placebo N=83	
	ALL	**Treated Exacerbations (Event based)**	**ALL**	**Treated Exacerbations (Event based)**
None	12 (14.1%)	-	12 (14.5%)	-
All Type III (1 symptom deteriorated)	34 (40.0%)	10 (29.4%)	33 (39.8%)	11 (33.3%)
Dyspnoea	25 (73.5%)	8 (80.0%)	27 (81.8%)	10 (90.9%)
Sputum volume	3 (8.8%)	1 (10.0%)	5 (15.2%)	1 (9.1%)
Sputum colour	6 (17.6%)	1 (10.0%)	1 (3.0%)	0 (0.0%)
All Type II (2 symptoms deteriorated)	23 (27.1%)	12 (52.2%)	12 (14.5%)	4 (33.3%)
Dyspnoea + Sputum volume	6 (26.1%)	5 (41.7%)	2 (16.7%)	1 (25.0%)
Dyspnoea + Sputum colour	11 (47.8%)	6 (50.0%)	4 (33.3%)	2 (50.0%)
Sputum volume + Sputum colour	6 (26.1%)	1 (8.3%)	6 (50.0%)	1 (25.0%)
Type I (all 3 major symptoms deteriorated)	16 (18.8%)	13 (81.3%)	26 (31.3%)	20 (76.9%)

The percentages in the Table are calculated and explained by examples on how to read the Table. For example, in the AAT group, (n = 85), 12 patients had no exacerbation, 12/85 = 14.1%. In the 27 patients who had change in Dyspnoea in the placebo group of 33 patients with a type III first exacerbation, 27/33 = 81.8

Spirometry outcome

Post hoc analysis of FEV_1 and FEV_1/SVC results demonstrated that patients inhaling AAT had an improvement of lung function (Table S4 and Figure S1). The decrease in FEV_1 from baseline to the study end tended to be less in the AAT group. By taking into account FEV_1 measured at all time points, MMRM analysis showed an overall effect of +18·6 ml FEV_1 (CI 95% -8.2, 45.4) for the AAT group and –29·4 ml (CI 95% -55.3, -3.5) for the placebo group by the end of the study, representing a 48·0 ml difference over 50 weeks, (p = 0·0124). A similar statistically significant difference in the AAT-treated patients was observed regardless of the imputation method used to deal with missing data (detailed above in the section statistical analysis of spirometry). At the study termination visit, 4 weeks after last dose, the FEV_1 in both groups had declined further (Figure S1), but the AAT group maintained a difference of 40 ml compared to placebo.

Table S4: MMRM Analysis of Lung Function – All Covariates*, Without Early Termination Visit Data (ITT Population **)

| Lung Function | Change from Baseline to Week 50 | | | Overall treatment effect | | P-Value (Mixed Linear Model - Overall Treatment Effect) |
| | Least Squares Mean (95% CI) | | | Least Squares Mean (95% CI) | | |
	AAT (N = 85)**	Placebo (N = 83)**	P-Value	AAT (N = 85)**	Placebo (N = 83)**	
FEV$_1$ (ml)**	-8·8 (-52·8, 35·1)	-63·6 (-103·5, -23·7)	0·0704	18·6 (-8·2, 45·4)	-29·4 (-55·3, -3·5)	0·0124
FEV$_1$ percent of predicted value (%)	-0·05 (-1·39, 1·28)	-1·66 (-2·87, -0·45)	0·0796	0·65 (-0·24, 1·54)	-0·70 (-1·57, 0·16)	0·0339
FEV$_1$ / SVC ratio	0·67 (-0·32, 1·66)	-1·11 (-1·99, -0·23)	0·0092	0·73 (-0·04, 1·51)	-0·95 (-1·70, -0·21)	0·0025
DLco (%)	-3·09 (-4·90, -1·28)	-3·35 (-5·01, -1·70)	0·8335	-2·27 (-3·60, -0·93)	-1·70 (-3·01, -0·39)	0·5527

* All covariates are visit and the interaction between treatment and visit with the following parameters: FEV$_1$ at baseline, age, oxygen use, BMI, and SGRQ at baseline.

** Patients with no FEV$_1$ measurement after baseline visit (6 patients in AAT and 2 in Placebo arms) did not contribute to the model.

Figure S1: FEV$_1$ (ml) mean changes from baseline up to week 54 (4 weeks after last dose) is shown for both groups. Data at each time point is shown as mean +/- SD
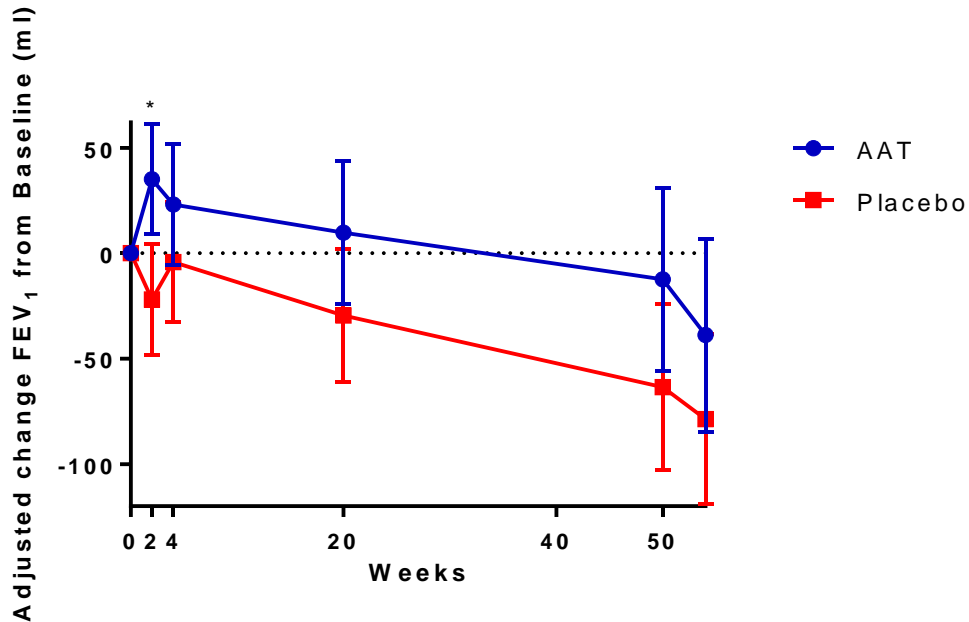
* Significant difference between the treatment arms in week 2 by MMRM. p = 0.0461

Several missing data imputation methods were applied to the post hoc analysis data, including worst case scenario models (Figure S2). It is important to note that J2R imputation provides a worst-case scenario for withdrawn subjects in the AAT treatment group because it considers the unrealistic assumption that the therapeutic effect of AAT ceases immediately, and therefore "dilutes" the clinical effect of the treatment. Even in this scenario, the overall effect was still close to significance (p = 0.0871), as shown in Figure S2 . The J2R imputation is the most conservative approach for the handling of data that satisfy the assumption of missing not at random.

Figure S2: Comparison of FEV$_1$ analysis with imputation methods

FEV1 (L) - results summary overall
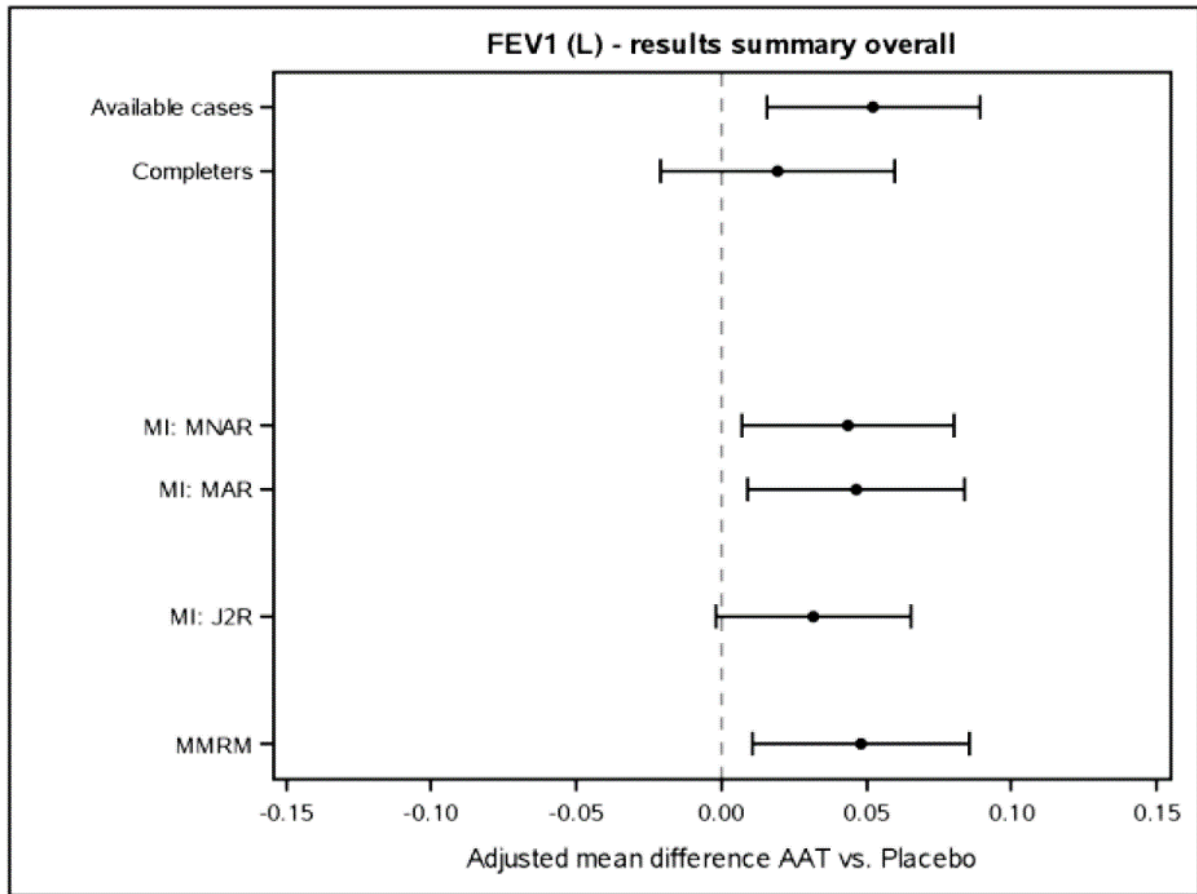
Adjusted for baseline FEV$_1$ value, age, country, BMI, SGRQ total score and oxygen use, MMRM: Mixed Models for Repeated Measures; MI: Multiple Imputations; MAR: Missing at Random; MNAR: Missing Not at Random; J2R: Jump to Reference

Figure S3 presents the results from the post hoc analysis for FEV$_1$ without the data for the first four weeks where there was a "hockey stick effect" in the FEV$_1$ of patients in the AAT treated group. It should be noted that there are only two post baseline values available; at Week 16 post 4 weeks (study week 20) and Week 46 post the 4 week baseline (study week 50). The MMRM was run with the change from "the average baseline value" at Week 16 and Week 46 and the following covariates; age, country, BMI, SGRQ total score, oxygen use, FEV$_1$ baseline value (defined above), visit, treatment group, and a visit by treatment group interaction term.

Figure S3: Spirometry: MMRM, FEV1 (L) With Covariates*, Week 0-4 Mean As Baseline, Without Early Discontinuation Visit (ITT Population)
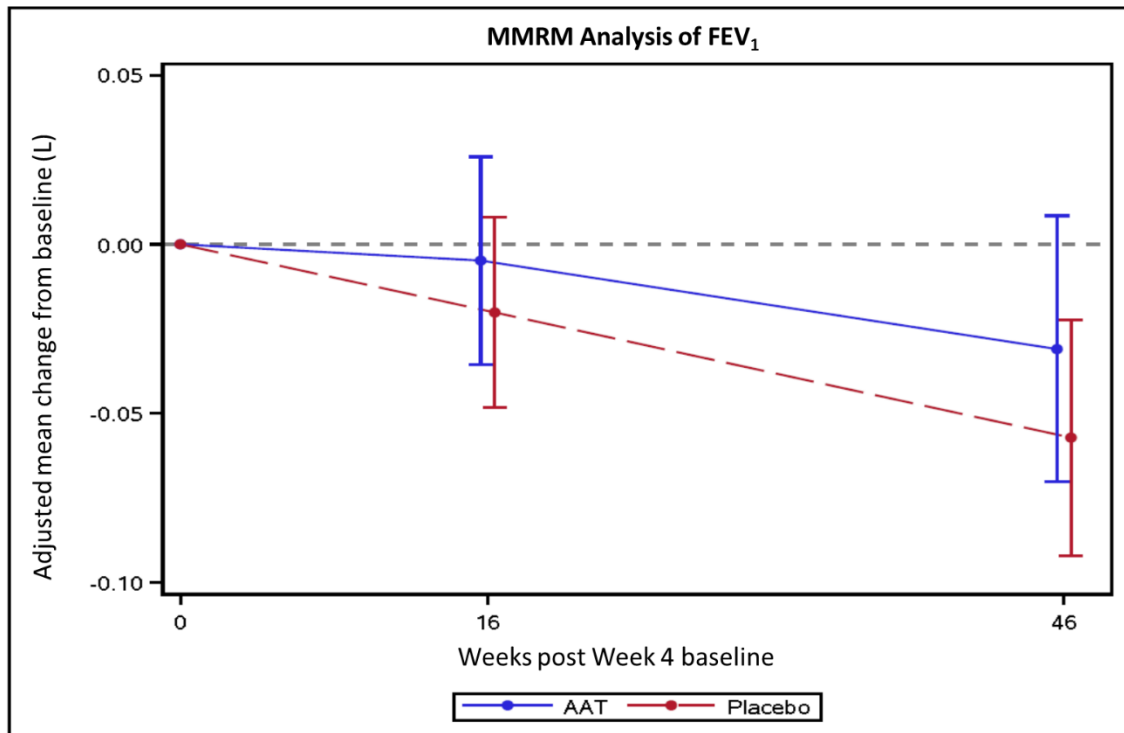
Adjusted for *Covariates: baseline value, age, country, BMI, SGRQ total score and oxygen use

Table S4: Health Statistics of the ITT population

	AAT	Placebo
	N = 85	N = 83
Number of days hospitalized		
Number of patients	22	12
Mean ± SD	7·7 ± 1·35	7·1 ± 1·34
Median (min-max)	5·5 (4-9)	5 (2·5-8)
Number of days used systemic steroids (≥ 10 days)*		
Number of patients	26	22
Mean days of steroid use ± SD	13.8 ± 6.3	32.4 ± 57.2
Median (min-max)	12 (10-51)	14 (10-294)
Number of days used systemic antibiotics (≥10 days)		
Number of patients	54	44
Mean days of antibiotic use ± SD	14·6 ± 10.7	17.6 ± 37.5
Median (min-max)	12 (8-100)	11 (10-349)

Table S4 shows that the mean number of hospitalization days during the study was similar in the AAT arm compared to Placebo (7.7 days vs. 7.1 days, respectively). However both the mean number of systemic steroid usage days and the mean number of days of antibiotic use were lower in the AAT group than the placebo group though not statistically different (13.8 vs·32.4 and 14.6 vs 17.6 days, respectively, $p > 0.05$).

The usage of systemic corticosteroids was calculated only for periods where the drugs were taken for 10 days or more. It should be noted that the minimum period for antibiotic use was only 8 days because the patient started on a course of antibiotics 2 days before the start of the study .

We reanalyzed the safety data of the RCT using a subgroup analysis which compared safety results between "Less Severe" (LS) and "More Severe" (MS) obstructive lung disease sub-populations. Randomization is considered to be lost in such analyses, but the demographics of the subgroups do not suggest any large imbalance. Patients were grouped into the LS or MS subgroups based on the severity of airflow limitation at baseline and symptom load. According to the GOLD criteria (Vogelmeier, Criner et al.

2017), FEV_1 50% of predicted is the level that differentiates between severe and moderate airflow limitation. Our patients were symptomatic, and per that definition most of them had Saint George Respiratory Questionnaire (SGRQ) scores of >25 at baseline. A cutoff of 50 was chosen to subdivide symptomatic and highly symptomatic patients as this was the approximate median value of all patients in our study. In summary, the LS subgroup was defined as $FEV_1 > 50\%$ and $SGRQ \leq 50$ and the MS subgroup was defined as $FEV_1 \leq 50\%$ or $SGRQ > 50$

The results of the analysis (see Table S6, Table S7, and Table S8) show that the higher incidence of SAEs in inhaled AAT vs. placebo (manifested as an increase in AEs, SAEs, hospitalization, decreased time to first mild/moderate exacerbation, and increased dropout) is concentrated in the "MS" sub-population. In contrast, the "LS" subpopulation had a safety profile that was comparable to placebo.

Table S6: Subgroup Analysis of Serious Adverse Events

	Less Severe *		More Severe **	
	AAT	Placebo	AAT	Placebo
Patients with at least one SAE N/Subgroup	4/22 (18%)	3/13 (23%)	26/65 (40%)	12/68 (18%)
SAEs N	6	5	35	16

* $FEV_1 > 50\%$ and $SGRQ \leq 50$, ** $FEV_1 \leq 50\%$ or $SGRQ > 50$

Table S7: Hospitalization events, ITT populations*

	ITT			
	AAT (N=85) N (%)		**Placebo (N=81) N (%)**	
Hospitalization Events				
Number of Patients Hospitalized (%)	22 (25.3%)		12 (14.8 %)	
	Less Severe	More Severe	Less Severe	More Severe

	ITT			
	AAT **(N=85)** **N (%)**		**Placebo** **(N=81)** **N (%)**	
	3 (3.5)	19 (22.3)	3 (3.7)	9 (11.1)
Mean Length of Hospital Stay, Days ± SE	7.7 ± 1.35		7.1 ± 1.34	
	Less Severe	More Severe	Less Severe	More Severe
	5.33 ± 0.33	8.05 ± 1.55	6.67±3.84	5.78 ± 1.42
Median Length of Hospital Stay (Days) (Q1,Q3)	5.5 (4.0,9.0)		5 (2.5,8.0)	
	Less Severe	More Severe	Less Severe	More Severe
	5.00 (5.0,6.0)	6.00 (4.0,14.0)	5.00 (1.0,14.0)	5.00 (3.0, 7.0)

*The data was derived from ITT population listings
** Standard error and quartile calculations are not applicable for this one event.

Table S8: Subgroup Analysis of Early Withdrawals

	Less Severe*		**More Severe****	
Subgroup	AAT	Placebo	AAT	Placebo
N	22	13	65	68
Total Withdrawals (%)	6 (27.3)	2 (15.4)	29 (44.6)	11 (16.2)
Withdrawals due to AE (%)	1 (4.5)	1 (7.6)	14 (21.5)	5 (7.3)

* $FEV_1 > 50\%$ and $SGRQ \leq 50$, ** $FEV_1 \leq 50\%$ or $SGRQ > 50$

Safety events over study duration.

After the first year, the rate per 100 days of safety events in the AAT treated group dropped significantly (from 0.29 to 0.11 withdrawals, 0.27 to 0.15 SAEs and 0.81 to 0.64 exacerbations) becoming similar to that of the placebo group. SAEs and severe exacerbations were balanced

between AAT and placebo. Moderate exacerbations were balanced between AAT and placebo throughout the study, and exhibited clear seasonality. Withdrawals were imbalanced in favor of placebo, with lower rate per year after the first year (see Figure S4 below)."
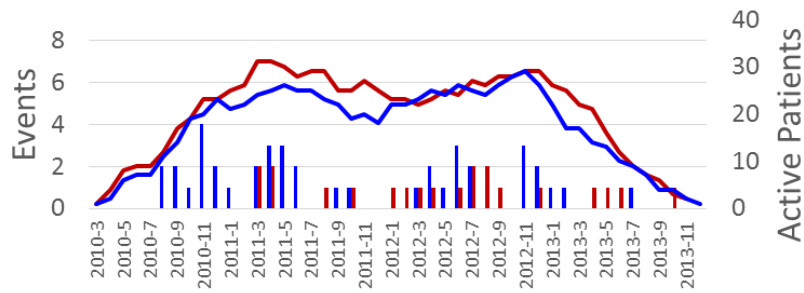
These changes observed between the first year and the rest of the study were influenced by small sample size, patients variability and the improvement in the handling of AAT and nebulizer, that resolved the complain about prolonged nebulization time.
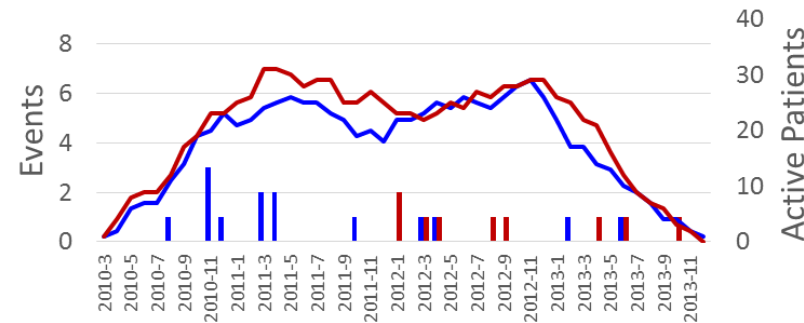
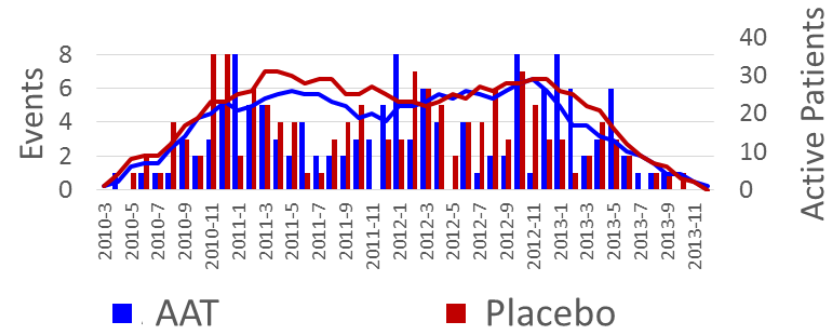**Figure S4: Safety events over study duration**

A. Withdrawals
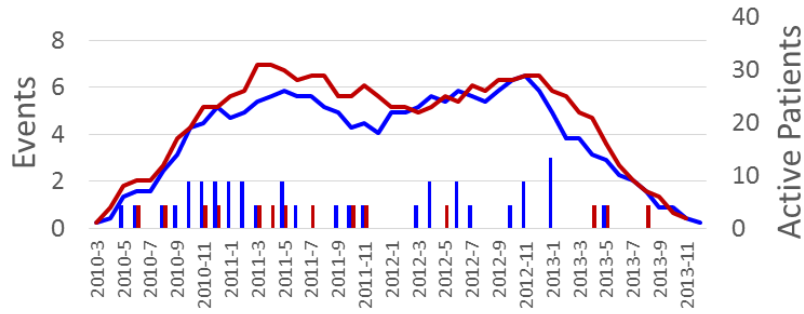
B. Serious Adverse Events

C. Severe Exacerbations

D. Moderate Exacerbations

■ AAT   ■ Placebo

References

Molenberghs, G. and M. G. Kenward (2007). Multiple Imputation. <u>Missing Data in Clinical Studies</u>, John Wiley & Sons, Ltd**: 105-117.**

Vijayasaratha, K. and R. A. Stockley (2008). "Reported and unreported exacerbations of COPD: analysis by diary cards." <u>CHEST Journal</u> **133**(1): 34-41.