Monte Carlo External Control Arm Generation Utilizing Real-World Patient Data and Deep Learning Quantitative CT Metrics Demonstrates Treatment Effect in the ATLAS IPF Trial

Kiril R. Kirov[1], Elliott Bussell[1], Muhunthan Thillai[1,2], Felix A. Woodhead[3], Howard M. Lazarus[3], Craig S. Conoscenti[3], Simon L.F. Walsh[1,4]

[1]Qureight Ltd, Cambridge UK, [2]Royal Papworth Hospital, Cambridge UK, [3]Avalyn Pharma Inc, Cambridge, MA USA, [4]National Heart and Lung Institute, Imperial College UK

Efficacy testing of antifibrotic treatments for idiopathic pulmonary fibrosis (IPF) poses challenges without a placebo control arm trial design. This study explores a Monte Carlo Cross-validation (MCCV) approach for generating external control arms (ECAG) using real-world hospital data in an IPF trial of inhaled pirfenidone (AP01).

Trial

The ATLAS Trial (ANZCTR ACTRN12618001838202) evaluated the safety and tolerability of AP01 in patients with IPF at 25 sites in Australia, New Zealand, the Czech Republic, Poland, the Netherlands, and the UK (*Thorax 2023; 78:882–889. doi:10.1136/thorax-2022-219391)*. Patients were randomised 1:1 to receive nebulised AP01 50mg once a day or 100mg twice a day for up to 72 weeks. At Week 48 of treatment, 28 patients from each dose had lung function tests and comprise the two treatment cohorts for which external control arms are generated. Both cohorts have similar baseline characteristics (Table 1).

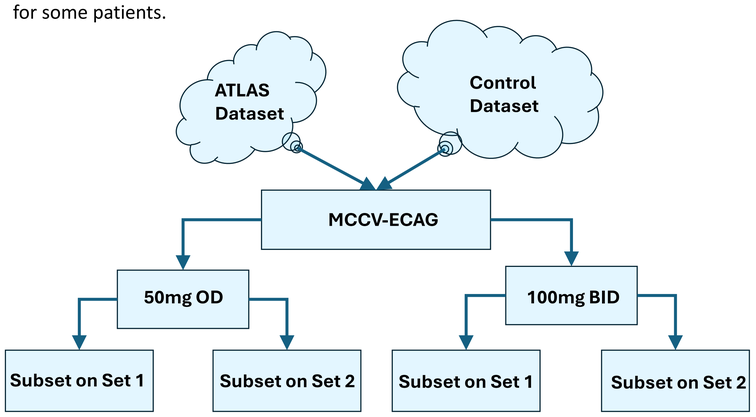
Table 1. Baseline characteristics of ATLAS patients who have completed 48 weeks of treatment with AP01 and have lung function measurements.

	50mg OD	100mg BID
Number of patients	28	28
Sex, M/F	20/8	22/6
Age, years	72.9 (7.1)	71.5 (7.9)
FVC, L	2.58 (0.55)	2.62 (0.53)
TLCO, mmol/min/kPa	3.61 (1.28)	3.45 (1.32)
FEV1/FVC, %	82.6 (5.9)	82.5 (6.9)
FVC, % predicted	71.1 (11.8)	72.5 (9.9)
TLCO, % predicted	49.1 (12.5)	50.4 (11.3)
Lung Volume, L*	4.0 (0.9)	4.2 (1.0)
Fibrosis Volume, %*	14.7 (7.2)	12.8 (8.7)

Mean values and standard deviation (in brackets). FVC – forced vital capacity, TLCO – transfer factor for carbon monoxide, FEV1 – forced expiratory volume in 1s, OD – once a day, BID – twice a day. *$N_{50mg\ OD}$ = 27, $N_{100mg\ BID}$ = 25. **Lung Volume (L) and Fibrosis Volume (%) are derived from HRCT scans using Qureight's Lung8 and Fibr8 neural network semantic segmentation algorithms.

External Control Arm Generation

Figure 1 provides an overview of the MCCV-ECAG experiments. Each branch in the figure involved 10,000 draws without replacement from Qureight's collection of real-world IPF data of treatment-naive patients in the UK. Control arms were generated in a 1:1 treatment-control ratio. The quality of matching on baseline characteristics was assessed using chi-squared, t-, Mann-Whitney, and multivariate kernel tests. Set 2 external control arms were drawn from an approximately 40% smaller data pool due to missing HRCT scans for some patients.



Set 1 – age, sex, height, FVC % predicted, FEV1/FVC % predicted, TLCO % predicted
Set 2 – Set 1 + HRCT Lung Volume (L) and Fibrosis Volume (%)

Figure 1. External control arm generation tree.

External Control Arm Generation Results

MCCV-ECAG produced well-matched 1:1 control arms for both treatment groups on both sets of baseline characteristics. The mean p-values for the 1,000 best-matched control arms of each MCCV-ECAG branch in Figure 1 were in the range 0.44 (SD = 0.04) - 0.58 (SD=0.05)]. An example of the similarity between the treatment and control distributions achieved by the MCCV-ECAG procedure is shown in Figure 2.

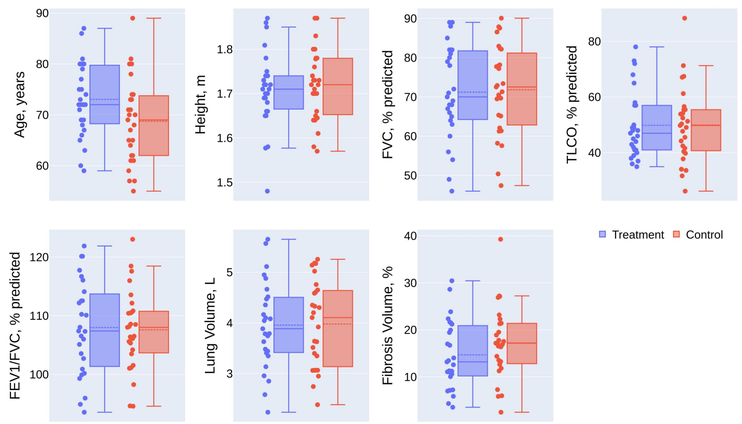


Figure 2. Comparison of the baseline characteristics of the 50mg AP01 cohort after 48 weeks of treatment with the characteristics of an external control arm generated by Monte Carlo cross-validation. Mean values are represented with a dashed line. The chi-squared p-value for matching on sex is 1.0. The weighted mean p-value for matching on all eight baseline features is 0.69.

Lung Function Decline in the ATLAS Cohort

While the mean FVC % predicted change of the 50mg OD group showed an overall decline, the mean change of the 100mg BID group was positive for the first six months of treatment (Figure 3). At Week 48 of treatment, the mean change in FVC (% predicted) for the low- and high-dose groups was -4.57% (SD=5.61%, N=28) and 0.035% (SD=7.87, N=28), respectively.

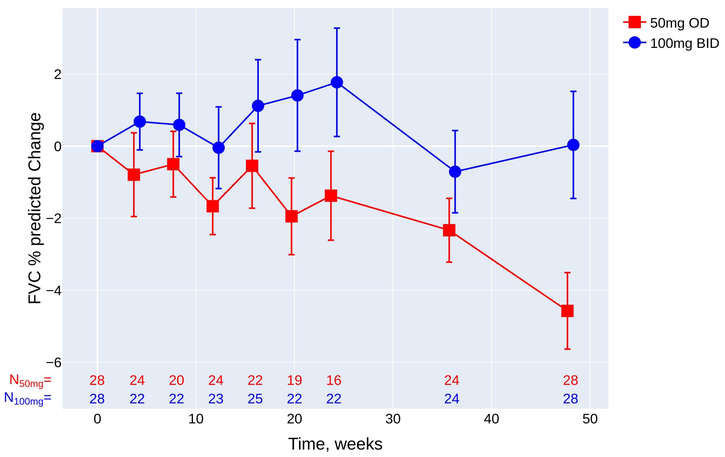


Figure 3. Longitudinal change of the mean FVC % predicted of the two ATLAS treatment groups. The error bars signify the standard error. The number of patients at each follow-up visit is indicated at the bottom of the plot for both trial groups.

Efficacy Testing

For efficacy testing, the mean change of FVC (% predicted) at Week 48 of treatment was compared with the distribution of mean FVC (% predicted) changes of the 1,000 best-matched control arms generated by MCCV-ECAG for each trial group and baseline feature set (Figure 4 and Table 2).

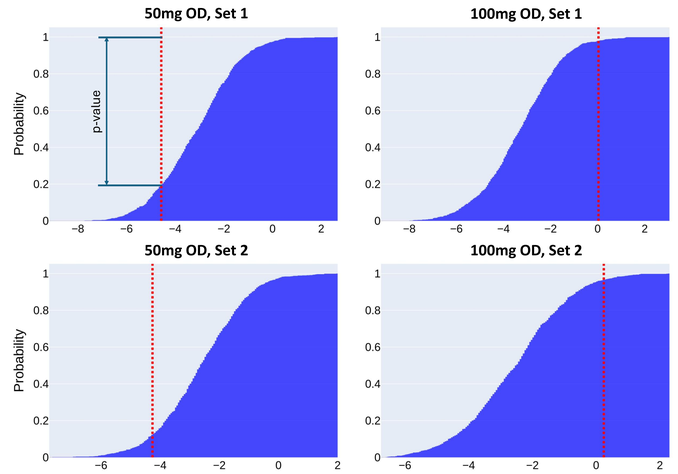


Figure 4. Cumulative distribution functions of mean control FVC % predicted change at 48 weeks of follow-up derived using MCCV-ECAG based on feature Set 1 (top row) and Set 2 (bottom row). The mean FVC % predicted change of the 50 and 100mg AP01 groups at 48 weeks of treatment is overlayed (red dashed lines) for comparison.

Table 2. Right-tailed efficacy test results at 48 weeks of treatment. The p-value of the test is the fraction of the 1,000 best-matched MCCV-ACAG control arms which have a mean FVC % predicted change greater than that of the treatment group. α = 0.05.

Treatment Group	Feature Set	$N_{patients}$	Efficacy p-value
50mg OD	Set 1	28	0.803
	Set 2	20	0.901
100mg BID	Set 1	28	0.021
	Set 2	25	0.035

Conclusions

- MCCV-ACAG enables the generation of well-matched external control arms from real-world data sources.
- The Monte Carlo approach demonstrates that groups of patients with no statistically significant differences in core baseline parameters can have very different changes in FVC % predicted after 48 weeks of follow-up.
- The efficacy testing based on MCCV-ECAG has shown a statistically significant treatment effect for 100mg BID of AP01.
- This work presents the first application of quantitative HRCT in creating external control arms in IPF and signals a potential paradigm shift in the investigation of novel drug candidates in clinical trials.

Conflicts of Interest

Kirov, Bussell, Thillai, and Walsh are employees of Qureight Ltd. Woodhead, Lazarus, and Conoscenti are employees of Avalyn Pharma Inc. This work is funded by Avalyn Pharma Inc and Qureight Ltd.

ATS 2025