Re: TAR-20-138: resubmission: Manuscript

Dear Editors,

We are pleased to resubmit our manuscript with the updated title of "The Value of Imaging and Clinical Outcomes in a Phase 2 Clinical Trial of a Lysophosphatidic Acid Receptor Antagonist in Idiopathic Pulmonary Fibrosis".

We appreciate three reviewers' insightful comments to improve this manuscript. We carefully revised according to the reviewers comment.

Thank you for the opportunity of resubmitting this manuscript.

Reviewer(s)' Comments to Author:

Reviewer: 1

Comments to the Author
This is a prospective study utilising AI technology to assess quantitative changes in HRCT images for fibrosis and correlate these with clinical and physiological measurements within a 26 week prospective blinded randomised controlled trial of BMS daily for IPF.

The introduction is well written and easy to understand. The rationale for the study is explained and the aims are clearly set out. There is good use of appropriate referencing.

The methods, patient group and study endpoints are clearly explained.

Results:
Baseline characteristics are representative of an IPF cohort with no differences in each arm
The results are clearly explained in the text with tables and figures complementing the text well.
The authors have presented the data that quantification fibrosis score correlate well with lung function but not breathlessness score at baseline and correlate with lung function and breathlessness score at 26 weeks.
Whilst the mean fibrosis quantification scores do not differ between groups at 26 weeks the authors have demonstrated a greater proportion of patients in treatment arms showing reduction in fibrosis scores as seen in figure 1
The discussion discusses the limitations of the study and compares their results with other published studies.
Whilst the RCT had to stop due to hepatotoxicity reasons the results from this limited number of patients still provides interesting findings related to quantitative imaging changes in IPF and their correlation with lung function and breathlessness and provides important insight in it utility as a biomarker in future studies.

This is a well designed and conducted study and despite limited recruitment as the study was terminated early - it provides important findings regarding the utility of quantitative imaging as a biomarker for future studies

Response: We appreciate the comments from Reviewer 1.

Reviewer: 2

Comments to the Author
A very nicely formulated study. I would add to the limitations part that this is an analysis of data of pre-existing study thus all the limitations, biases and confounders will be naturally part of this new one. Second, you described in detail certain radiological parameters and their significance as stand alone or in summation , it worthwhile looking in other possible significance in variables such as QILD/QLF for example , as those might be more significant as biologically IPF does not behave as a simple add or substract formula.

Response: We added a limitation in Discussion in the third sentence of the fifth paragraph.

"Imaging outcome of QLF was part of the secondary endpoint outcome in a clinical trial. The changes in radiological outcome is dependent upon the available and analyzable HRCT examinations in the trial."

The unit of QLF score is already in the percentage scale by adjusted by the sum of whole lung size. Similarly, the unit of QILD score is also in the percentage scale (i.e. QILD= QLF+QHC+QGG, where QHC is quantitative honeycomb score and QGG is quantitative ground glass score). If we make a ratio, we may lose the physical meaning of disease pathology. Unlike the nonspecific interstitial pneumonia (NSIP) there are not many pathological studies on ground glass opacity in IPF population. Since most of IPF population are either UIP or probably UIP patterns in HRCT, we would like to keep our focus on QLF changes.

Reviewer: 3

Comments to the Author
The authors have produced a well-written manuscript exploring the utility of a computer-based quantitative HRCT as an efficacy endpoint in a prospective IPF study. This is a topical area of great interest as the authors have eluded to, as new and more reliable biomarkers are needed to monitor and detect early changes of disease progression.

Although the BMS trial was cut short, the purpose of this study was to evaluate the utility of computer-based quantitative CT. Within the confines of a shortened study, the authors were able to demonstrate the value of quantitative HRCT as an efficacy endpoint for new treatment and disease monitoring for IPF. Limitations of the study were noted.

Minor points to clarify:
1) 7 out of 137 were excluded due to improper scanning - were these dictated by a central panel/radiologist associated with the study?

Response: Seven out of 137 excluded cases were based on the guideline of technical quality control (QC) by a central imaging center and confirmed by a central radiologist. We revised the last sentence in the first paragraph under Automated Quantitative CT Image Analysis as follows**:**
"Of these 137 patients, an additional 7 patients were excluded due to HRCT scans with motion artifact, inconsistent positioning, and/or improper Digital Imaging and Communications in Medicine (DICOM) formatting. The motion artifact and inconsistent positioning in prone and supine images were confirmed by a central radiologist (N=6). Improper DICOM formatting was checked by standard technical quality control in the imaging analyses (N=1), where DICOM formatting is necessary condition in quantification of CT images."

The original sentence was: "Of these 137 patients, an additional 7 patients were excluded due to HRCT scans with motion artifact, inconsistent positioning, and/or improper Digital Imaging and Communications in Medicine (DICOM) formatting."

2) Was QILD the only quantitative method used in this study? Or did the authors have data from other methods to compare given the lack of consensus on the optimal methodology?

Response: Yes, QLF and QILD scores are only quantitative method used in this study. The author did not have data from other methods. We acknowledge that there are several methods in quantifying the ILD patterns. Most of currently available methods are used the similar approaches using a set of texture features and machine learning. The individual methods are already included in the reference number 11-14, 25, 26, and 29-33.

3) Apart from UCSD-SOBQ, were there any other QOL or physiological score data (e.g. MMRC, ST George Respiratory questionnaires, GAP score) that could be used to further support QILD scores as a tool in patient-reported outcomes?
Response: There are not easily accessible other QOL and physiological scores or patient-reported outcome. To calculate GAP score, we need additional information of types of missing reasons whether the missing was due to technical failure, or an enrolled subject's disagreement, or other random reasons versus an enrolled subject's severity that it could not be measured. If it is for the latter reason (i.e. no random), the risk of GAP score increases by one point. Currently, we do not have additional information of reasons for missing functional measurements. We add a sentence below in the last paragraph in Discussion just before the Conclusion:

"In future studies, it would be desirable to include systematically collected Patient-Report Outcomes, reasons for missing pulmonary functional tests, and a performance measurement of 6 minute walk."

Thank you for the opportunity of resubmitting this manuscript.