Author’s response to reviews

Title: Disease progression in Idiopathic Pulmonary Fibrosis with mild physiological impairment: Analysis from the Australian IPF Registry

Authors:

Helen Jo (helen.jo@sydney.edu.au)
Ian Glaspole (IGHLaspole@lasv.com.au)
Yuben Moodley (yuben.moodley@uwa.edu.au)
Sally Chapman (sally.chapman@health.sa.gov.au)
Samantha Ellis (s.ellis@alfred.org.au)
Nicole Goh (n_goh@yahoo.com)
Peter Hopkins (peter_hopkins@health.qld.gov.au)
Gregory Keir (gregory_keir@health.qld.gov.au)
Annabelle Mahar (annabelle.mahar@sswahs.nsw.gov.au)
Wendy Cooper (wendy.cooper@sswahs.nsw.gov.au)
Paul Reynolds (paul.reynolds@adelaide.edu.au)
Haydn Walters (haydn.walters@utas.edu.au)
Christopher Zappala (chris.zappala@health.qld.gov.au)
Christopher Grainge (christopher.grainge@hnehealth.nsw.gov.au)
Heather Allan (heather@lungfoundation.com.au)
Sacha Macansh (sacha@lungfoundation.com.au)
Tamera Corte (tameracorte@mac.com)

Version: 2 Date: 17 Dec 2017

Author’s response to reviews:

Dear Elisabetta Balestro, Spyros Papiris and Justin Oldham,
Thankyou for reviewing our paper “Disease progression in Idiopathic Pulmonary Fibrosis with mild physiological impairment: Analysis from the Australian IPF Registry” for publication in BMC Pulmonary Medicine. Your thoughts and insights were exceedingly useful and I agree with many of the comments made.

Overall, we have edited the paper and analysis for focus on the FVC≥80% as the primary criteria for mild physiological impairment. We have kept DLco, CPI and GAP as secondary analysis but changed the DLco criteria to 55% as I agree, the threshold of 40% chosen was more in keeping with severe disease. Please find a detailed review of the changes and response to comments below.

Overall, we feel that the comments and changes made have improved the manuscript and we thankyou again for your ongoing consideration.

Sincerely,

Helen Jo

Response to review is greater detail:

Spyros A Papiris (Reviewer 1):

This is an interesting trial focusing on mild impairment patients. The study could have significant clinical implications. The authors are advised to ameliorate their work by addressing the following:
1) The main concern regards the fact that the authors have chosen as main DLCO value for mild impairment that of ≥40%. The authors of course provide also exploratory analysis for other criteria such as 50%, 35% and 30%. Taken into consideration the fact that the GAP index uses as cut off points the values of 55% for mild, 36-55% for moderate and less than 35% for severe disease the authors are kindly advised to reconsider the selection of the criteria for mild disease based on a DLCO ≥40%.

Thankyou for this comment. We agree with your assessment and have changed the DLco criteria to 55%.

2) Furthermore in the discussion section when they conclude that their findings regarding the annual decline of FVC% between those patients with mild and more severe impairment mirror the results of the post hoc analysis of the pirfenidone and Nintedanib trials, they should comment on the fact that all pirfenidone and Nintedanib trials had excluded patients with FVC <50% and DLCO < 30-35%. Could the authors please provide some comments on that as well as on how many patients with FVC <50% and/or DLCO < 30-35% did the "more severe" group of the Australian IPF registry include?

I agree that is an important point to consider and we have amended the discussion to read:

“Registry patients also have a wide spectrum of IPF disease severity with 68 patients below the threshold for inclusion in clinical trials (FVC<50%, n=9; DLco <30% n=47, and n=6 for both), and are managed by respiratory physicians throughout Australia, not limited to tertiary referral centres.”

Justin M Oldham, MD, MS (Reviewer 2):
My only major concern is the stratification points chosen in this analysis. A relative balance between "mild" and "moderate-to-severe" cohorts seemed to drive the chosen stratification points for FVC, DLCO and CPI rather than more intuitive stratification points driven by previous data/publications. Brett Ley's GAP paper shows that mortality in those with DLCO<50% is quite high relative to those >50%, which resulted in a 55% cut point for DLCO in the GAP score. Using this sort of data-driven stratification point irrespective of how many patients fall into each resulting group would make for a cleaner study in my mind. Alternatively, focusing the analysis on FVC-stratified groups with secondary analyses for GAP and CPI would alleviate the DLCO issue altogether. A primary focus on FVC would make sense b/c the inability to obtain anti-fibrotic therapy is often based on the FVC % predicted (at least in the US) rather than DLCO. This approach would make your choice of FVC as the stratification variable in Tables 1 and 4 more consistent with the overall approach.

Thankyou for this comment and we agree entirely. We have therefore changed the paper to focus on the FVC criteria as it is clinically used, but kept the secondary analysis of DLco, CPI and GAP. We have also changed the threshold to 55% as suggested.

Other minor concerns:

Methods

- Please briefly explain why a random intercept and slope mixed model was chosen. Was this based on an information criterion (Aikake/Baysian?)

We chose a random intercept model a priori to reflect that people with IPF can have different starting FVC values. We used random slope analysis to allow for variation in FVC trajectories to allow for heterogeneous disease courses.
- What type of variance-covariance correlation structure was chosen for the mixed model (exchangeable, autoregressive, independent, etc)?

We used an unstructured correlation structure. We have amended the manuscript to include this detail:

“An unstructured linear mixed model for changes in FVC % predicted per year was fitted with random intercepts and slopes to compare the disease trajectory for patients with mild compared with moderate-severe physiological impairment, over the entire time of patient follow up in the registry.”

- Please explain how time points were treated in your mixed model? It's unlikely that all patients underwent follow-up PFTs at the same time point, so it should be mentioned how alignment of data was performed (ie 1-year intervals, 6-month intervals, etc).

Thankyou for this comment. There was no data alignment performed. The actual dates of all PFTs performed were used for calculation in the mixed model as PFTs occurred at any time point during the patients followup, reflecting different practices and availability around Australia.

- Why was progression-free survival only modeled over 12 months in your logistic regression analysis? Your KM's and mixed model plots show 5-year follow-up and your median survival is almost 3 years so it seems valuable data may be lost by limiting this analysis to 12-months. More importantly, it's unclear to me why logistic regression for binary death/progression was used in Table 4 rather than Cox regression, using time to death/progression.

Thankyou four your comment. We decided to focus on the 12 month outcome for this analysis as we felt that clinically, the indecision on whether to treat patients with mild physiological impairment is in the first 6-12 months, with many physicians electing to monitor mild patients for a few months before initiating treatment (Maher et al. ATS 2017). We wanted to highlight
that even within 12 months in patients with mild disease, there can be progression, and if physicians/funding bodies still wanted to limit access, additional, informative criteria may be useful to select patients that should have early treatment.

Results

- Please consider using "mild" and "moderate-to-severe" for your groupings throughout the manuscript, as there likely exists at least two groups with progressively worse survival within the current "more severe" group (the data for GAP increasing mortality risk for GAP stage II to GAP stage III supports this).

Thank you, we agree with this comment and the wording has been changed throughout the manuscript.

- For Table 1, please mention how many patients in the "mild" and "moderate-severe" groups (using FVC cut off) received anti-fibrotic therapy. If there was a substantial number, please include this as a variable in your survival and longitudinal FVC analyses, as this could arguably influence the results of both.

This has been added to the table and there were basically equal numbers in both groups (~25%). While I have not included antifibrotics in the multivariable model, I have analysed this and it makes no difference to the results. As this is not a paper focusing on the therapeutic effect of antifibrotics (as it is retrospective and non-randomised etc) we have decided not to include it in the univariable or multivariable analysis.

- Exclusion of patients unable to perform the DLCO maneuver is not appropriate in my mind because these patients often cannot perform the test due to advanced disease. This finding is informative (as supported by the GAP paper) so should not be treated the same as other missing data. Depending on how you decide to proceed with DLCO modeling, please conduct a sensitivity analysis (ok for the supplement) to determine whether exclusion of patients unable to
perform DLCO influenced your results. Imputation of missing DLCO data to the lowest quartile mean or lowest observed value seems like reasonable ways to account for these data in the sensitivity analysis. This obviously doesn't need to be done if the study becomes more narrowly focused on FVC.

Given that any patient with IPF anywhere in Australia can be included in the Registry, it is difficult to know whether the DLco was not performed as it was not available vs unable to perform. We agree that a sensitivity analysis with imputation of missing DLco data would a reasonable analysis. Given however that we have primarily focussed on FVC, we have decided not to perform this.

- Please check your results for table 2a. A HR of 0.03 for FVC (% predicted) seems a rather low (i.e. for every 1-unit increase in FVC % predicted, you get a 33-fold reduction in mortality risk?) Next, the HR for DLCO(% predicted) is 0, which seems incorrect. Next, the HR for 6MWT is 1.0 with CI that crosses 1 but a p-value of 0.001, suggesting this was calculated incorrectly.

Thankyou for picking this up. You make a very valid point. The issues with the HR were with the units used (FVC 70% was imputed as 0.7), giving ridiculous HRs. We have changed and clarified the unit changes for all the HRs in all the tables. (FVC*, DLCO*, CPI*, 6MWD^, SGRQ score#, UCSD SOBQ score*)

*For every 10 unit change; ^ for every 50m change; #for every 4 point change

- In Table 2b, please reverse your baseline group in your Cox model as it currently appears as though your mild groups have higher mortality risk than your severe groups (i.e. I read that FVC >80% increases the risk of death by 3-fold compared to FVC <80%, GAP stage 1 increases the risk of death 7-fold compared to GAP stage 2-3, etc).
Thank you for this observation. I agree that it was confusing and we have therefore reversed the baseline group in the table 2b.

- For table 3, please include the confidence intervals and p-value for each group's point estimate to show that the decline observed in each group was statistically different than 0 despite being no different than each other.

Thank you for this observation. I have included the 95% CI compared to 0 for each group. The p value for all values compared to 0 were <0.001.

- For each of the tables, please include the number of missing data in the footer. Given the major limitation of missing data, it would be helpful to know how many patients were included in an analysis of any given variable. This is true especially for 6MWT data, which was highlighted as a significant predictor.

For each table we include an “n” value indicating the number of participants who have data for that variable available as I agree that missing data is inevitable in registry analysis.

- Please include the number of patients included the unadjusted and adjusted Cox/logistic analyses since missing data will reduce the number of patients included in a final adjusted model.

I have added a footnote in the multivariable analysis table to indicate the number with all data available for multivariable analysis (n=412 vs 416 for univariable analysis).

* Multivariable model includes age, gender, BMI and smoking status and includes 412 of 416 available for univariable model.
Please include a multivariable adjusted analysis for Table 4 to help address whether other variables with clear confounding potential influence the point estimates for any given variable of interest.

Thankyou for your comment. I have included a multivariable analysis (adjusted for age, gender, smoking and BMI) in the table below. Given the results were not particularly different to the univariable analysis, we did not include the changes in the manuscript, rather, we have stated the results for a multivariable analysis in the text:

On a multivariable logistic regression model including age, male, BMI and smoking status, nadir oxygen saturation during 6MWT remained a significant predictor of poorer outcome (OR 0.89, 95%CI 0.80, 0.98; p=0.024)*.

*note the OR has been changed as there was an error in the original calculation – was not restricted to FVC≥80%.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>OR</th>
<th>95% CI</th>
<th>p</th>
<th>n</th>
<th>OR</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>216</td>
<td>0.99</td>
<td>0.94</td>
<td>1.05</td>
<td>0.847</td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>216</td>
<td>0.86</td>
<td>0.35</td>
<td>2.09</td>
<td>0.739</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>212</td>
<td>1.00</td>
<td>0.92</td>
<td>1.10</td>
<td>0.954</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever smoker</td>
<td>216</td>
<td>0.93</td>
<td>0.34</td>
<td>2.50</td>
<td>0.879</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC % pred*</td>
<td>216</td>
<td>0.81</td>
<td>0.56</td>
<td>1.16</td>
<td>0.241</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLco %pred*</td>
<td>216</td>
<td>0.87</td>
<td>0.67</td>
<td>1.14</td>
<td>0.312</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPI</td>
<td>216</td>
<td>1.26</td>
<td>0.86</td>
<td>1.86</td>
<td>0.234</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAP stage</td>
<td>216</td>
<td>1.12</td>
<td>0.43</td>
<td>2.90</td>
<td>0.813</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 MWD</td>
<td>63</td>
<td>0.99</td>
<td>0.99</td>
<td>1.00</td>
<td>0.083</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpO2 at start</td>
<td>63</td>
<td>0.86</td>
<td>0.71</td>
<td>1.03</td>
<td>0.108</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpO2 at end</td>
<td>63</td>
<td>0.89</td>
<td>0.81</td>
<td>0.98</td>
<td>0.021</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SpO2 nadir: 59 0.89 0.81 0.98 0.024 57 0.89 0.024
SGRQ#: 199 1.02 0.93 1.12 0.688 195 1.02 0.631
UCSD-SOBQ*: 126 1.00 0.97 1.03 0.970 124 0.98 0.886
Cough severity*: 177 1.16 0.95 1.41 0.136 173 1.16 0.137

- For figure 3, please clarify which set of criteria were used to determine "mild" and "moderate-to-severe?" Additionally, please include data points and a fitted LOWESS curve (ok to put into supplement if similar to current Figure 3). The current mean curves for the two groups don't take into account patient dropout over time so it would be helpful to see a rough estimate of how many patients informed the mean at any given timepoint.

Thankyou for your observation. The current figure 3 is a graphical representation of the annual FVC% predicted fall per year calculated using linear mixed models (grouped by FVC 80%). I have created a LOWESS curve by group for the supplement, as well as a summary of the means including number of participants with values in that range. Given the fact that the mixed model takes into account the change in each person’s FVC rather than the mean FVCs at each time-point, we feel that the current figure 3 is more informative. I have however, added a comment to the current figure 3 to make this clearer.

“Figure 3. Decline in FVC % predicted in mild (FVC≥80%) compared with moderate-severe (FVC<80%) physiological impairment

Graphical representation of annual FVC% decline as calculated by unstructured linear mixed model with random intercept and slopes for mild and moderate-severe disease.”
Discussion

- In the limitations section please mention that this was a retrospective analysis of a prospective observational study, as these patients were not prospectively followed with this study in mind (similar to retrospective (post-hoc) analysis of prospectively followed clinical trial cohorts).

Thankyou for this comment. We have amended the manuscript to read as below:

“Our study has several limitations. Firstly, this was a retrospective analysis of a prospective observational study and all investigations were performed as part of the patients’ routine clinical care, resulting in some variability in the investigations performe