Peer Review File
Article information: https://dx.doi.org/10.21037/jtd-22-1820

Reviewer A

This is a thorough and well-executed study of treatment patterns for patients with RA-ILD in a multi-site national ILD registry (CARE-PF). There are several important and interesting findings in this evaluation, the most compelling from my perspective is the finding that in a well-resourced country and a registry from well-resourced ILD clinics found that only 27% of these RA-ILD patients met their pre-specified criteria for "treatment." Additionally, it is striking in these data that fully 60 patients were trialed on treatment who did not continue therapy, although the registry-based data collection doesn't allow the investigators to determine why a patient wasn't continued, one can surmise that these treatment regimens are 1) difficult to tolerate and 2) treating clinicians may not have significant buy-in that these are "effective therapies."

The authors have performed an appropriate adjusted analysis and couch their findings appropriately within the limitations of retrospective registry data.

I agree with the authors that their findings are compelling and support the urgent need for randomization for drug efficacy and even placebo-controlled trials in RA-ILD to understand the efficacy of immunosuppression given the outstanding issues in this field.

The authors appropriately highlight that confounding bias, lack of baseline DMARD regimen data, and their small sample size (44 on treatment) limits most conclusions about treatment effect on these clinical outcomes. However, these analyses bring up interesting hypotheses about treatment options and highlight the urgent need to improve RA-ILD treatment uptake and study these outcomes prospectively with randomization to answer these important questions. The authors should be commended for their work to answer these important questions.

Reply to Reviewer A: Thank you – we appreciate your positive view of our work.

Reviewer B

Congratulations on completing your study. I am glad to see you are investigating the approach to (and efficacy of) therapy for patients with RA-ILD. As you point out in the manuscript, there are very, very few data to guide the field.

I have only 2 very minor comments about your paper:

Reviewer B Comment 1: Page 14, line 319-20: "We identified no differences…" but the log-rank p is 0.037, so there are statistically significant differences for this analysis. Of course, like many other analyses in this study, there may be too few subjects to tease things apart any further, but this should be corrected.

Reply Comment 1: Thank you for your observation, we will correct this in the manuscript.

Changes in the text Comment 1: We have modified the text as advised on pages 14-15 lines 320-327 and updated the p-value to match figure 2b. as 0.0370. It now states: "There was a trend towards improved transplant-free survival among the subgroups of patients with NSIP treated or untreated compared to UIP patients who were treated or untreated (p=0.0370) (shown in Fig. 2b)."

Reviewer B Comment 2: Recognizing results will not permit determination of median survival in the NSIP group(s), readers recognize median survival, and I think it would be helpful to include those data in the text.

Reply Comment 2: Thank you for this suggestion, we will add this to the text of the manuscript for the UIP group.

Changes in the text Comment 2: The following has been added to the manuscript on page 15, lines 327: "Median survival time for UIP was 7.98 years (95%CI IQR 5.40-11.49)."

Reviewer B Comment 3: Table 4. For the labels, I would edit to FVC% decline and DLCO% decline…I recognize the title for the table says decline, but I would change the table heading, as all values in the table are positive, and seeing positive values under a heading labeled "change" could be misinterpreted as improvement.

Reply Comment 3: This is an important clarification that has been pointed out, and we have made the appropriate change to Table 4.

Changes in the text Comment 3: Table 4 headings have now been changed. "FVC change" now reads "FVC % decline" and "DLCO % change" now reads "DLCO % decline".

Reviewer C

The authors report on clinical outcomes of rheumatic interstitial lung disease using a large registry. There are several concerns.

Reviewer C Comment 1: Similar studies (Eur Respir J 2010; 35: 1322–1328, Am J Respir Crit Care Med 2007;705–711) have been conducted in the past to show that RA-UIP has a poor prognosis. I think the novelty and clinical usefulness of this study are not well described.

Reply Comment 1: The main objectives of this study were to characterize treatment of RA-ILD and to determine if HRCT patterns might influence treatment decisions by clinicians. Of additional interest was to determine potential associations of treatment with outcomes and prognosis, which we did not identify in this cohort. Our findings for UIP patterns having a worse prognosis than NSIP is consistent with prior reports. Of further importance in this particular study is that the poor outcomes associated with UIP pattern strengthens the robustness of our clinician determination of HRCT pattern and that we report poor prognosis among those with NSIP pattern.

Changes in the text Comment 1: None.

Reviewer C Comment 2: I got the impression that, due to the limitations of existing treatments, we may need to consider treatment strategies centered on antifibrotic agents in the future. Do you plan to use the data after the approval of antifibrotic agents?

Reply Comment 2: We agree that real-world retrospective data on antifibrotic use in RA-ILD will be of great importance. However, nintedanib has only recently (less than one year ago) become available in Canada for use in non-IPF progressive pulmonary fibrosis including the PPF subset of RA-ILD. Therefore, it will require several years of additional data collection for us to analyze these data in our current registry.

Changes in the text Comment 2: None.

Reviewer D

Thank you for this interesting and well-written manuscript. Multi-center register data on RA-ILD patients are valuable, and this study had a relatively large sample size. The study was interesting, although concern arose about the study setting, reliability and generalizability of the data.

Reviewer D Comment 1: Data collection, reliability of the data: I would like more information about the data collection: was all information on clinical features, radiological patterns and medical treatment presented in this study gathered from CARE-PF registry, or did you utilize some other sources as well, e.g. medical records? If all information was based on registry data, how often are the registry data updated? Was the information about the medical treatment of each participant reliable and up-to-date?

Reply Comment 1: Registry data are collected at time of enrolment, including all relevant clinical data preceding and following the date of enrolment. For the purposes of this study, a detailed chart review was performed on every registry patient to ascertain treatment details and outcomes. We have cited the original "methods paper" describing the CARE PF registry as a reference for further details on the registry protocol. We further state in the methods section of this submission that a detailed and standardized chart review was under-taken to ascertain medical therapies, doses, data of use and outcomes - please see page 12, line 260-261 in the manuscript where it previously stated: "Treatment data were ascertained via a standardized retrospective chart review at each participating site." These data are considered reliable and are completed up to date at the time of censor.

Changes in the text Comment 1: The following sentences was added to the Methods sub-section Study Population page 10 lines 211-214: "Registry data are collected at time of enrolment, including all relevant clinical data preceding and following the date of enrolment. For the purposes of this study, a detailed chart review was performed on every registry patient to ascertain treatment details and outcomes."

Reviewer D Comment 2: Generalizability: What do you estimate, how large a proportion of RA-ILD patients end up in the CARE-PF registry in your centers?

Reply Comment 2: There are no robust epidemiologic population-based data to estimate the proportion of total patients with RA-ILD per population at risk, those seen by other Respirologists not involved in CARE-PF, and those not enrolled in the registry but still seen at each registry site. Similar to other registries, we are unfortunately unable to speculate with any reasonable precision and acknowledge that our study population is primarily generalizable to the types of patients who are cared for by specialists and subspecialists.

Changes in the text Comment 2: The following line was added to page 19, lines 433-435: "Additionally, we are not able to accurately ascertain the proportion of patients in Canada with RA-ILD who are cared for at a CARE-PF site or subsequently in the registry."

Reviewer D Comment 3: Question regarding the reliability of HRCT patterns: It is confusing that the authors referred to contemporaneous radiologic definitions (line 229), but still reported that the radiologic classification of study subjects was based on the clinician's interpretation. I would like authors to clarify whether the HRCT scans were re-evaluated especially for this study, or whether the determination of the radiological pattern was solely dependent on the clinician's expertise and HRCT interpretation skills. Were radiologists' reports of participants' HRCT scans available

for clinicians treating the study subjects?

Reply Comment 3: The clinicians caring for and enrolling patients with ILD in the CARE PF registry each have either > 20 years of ILD experience or have at least 2 years of ILD experience following a dedicated ILD subspecialty fellowship. All CARE-PF investigators are therefore considered ILD experts, and many contributing to guidelines for entities such as IPF, HP and the IIPs including NSIP. All sites have access to formal MDD and all CARE PF diagnoses are established according to contemporaneous best practices. Thus, our clinician determinations of HRCT pattern are generally of high quality and accuracy. The HRCT pattern is indicated on the diagnosis case-report form at the time of registry enrolment. The clinician has access to both the HRCT images and radiology report when indicating the CT pattern. CTs are also reviewed in a formal MDD in cases with diagnostic and/or therapeutic uncertainty. CTs were not specifically re-reviewed for the purposes of this study as this is a real-world cohort that best reflects typical clinical practice. In support of our clinician determination of HRCT patterns, our survival outcomes are consistent with those previously reported in the literature, suggesting accuracy in clinician interpretation and pattern selection for UIP and NSIP.

Changes in the text Comment 3: The following sentence was added on page 11, lines 241-243: "All sites have access to formal multidisciplinary rounds discussion, HRCT images, and the radiology report when designating the HRCT pattern on the case-report form at the time of registry enrollment."

Reviewer D Comment 4: Prednisone is commonly used for RA-ILDs and some studies even suggest favourable treatment responses to corticosteroids. Previous research data has also shown that cellular NSIP pattern is associated with better steroid response and lower mortality compared with fibrotic NSIP pattern. For these reasons, the use of prednisone or other glucocorticoids should also have been evaluated in this study. Certainly, the authors are right that evaluation of prednisone doses and duration of use is challenging in a retrospective study setting. However, I would recommend authors to add information on the use of corticosteroids in a simplified manner (yes/no) regardless of dose or duration of treatment, and also mention whether other immunosuppressants were used with or without steroids. It is probable that the proportion of RA-ILD patients using medication for ILD would be much higher if steroids were included in analysis.

Reply Comment 4: We agree that better characterizing the role of prednisone to treat RA-ILD is an important question, one ideally informed by prospective clinical trials. There is a lack of robust evidence to suggest that prednisone has a favorable treatment profile in fibrotic RA-ILD. Unlike the medications we included for this study prednisone doses are frequently adjusted and despite rigorous chart review, it is very difficult to ascertain accurate dosing. For these reasons, we hesitate to include data on

prednisone, concerned that its inclusion would introduce important confounding and misclassification bias. Our reasoning is noted in the Methods section, pages 12 lines 258-260. We have added this to the limitations section and suggest it as a focus for future work in this area.

Changes in the text Comment 4: We had previously added the following line to the Discussion section in our last revision within the manuscript on page 19, lines 422-424: "We were unable to evaluate the effect of prednisone given the limitations for accuracy in dosing or duration of therapy and could not exclude important misclassification bias."

Reviewer D Comment 5: How did the authors choose the definitions for potentially therapeutic doses (lines 140-245)?

Reply Comment 5: There are no robust data to inform ideal therapeutic doses of immunosuppressive drugs to treat RA ILD, thus we chose a low threshold for doses in the interest of favouring inclusivity.

Changes in the text Comment 5: We had previously added the following line in the last revision to the Methods section of the manuscript on pages 11-12, lines 252-253 "These doses were chosen to favour inclusivity, recognizing that there are no robust studies informing the optimal dosing for this clinical indication."

Reviewer D Comment 6: Would it be possible to differentiate between cellular and fibrotic NSIP patterns and analyse these subclasses separately? The results regarding the outcome and treatment response might be different between these subgroups.

Reply Comment 6: We agree this would be of interest for future study, but unfortunately is not possible given the limitations of chest CT for making this distinction combined with the infrequent performance of surgical lung biopsies for patients enrolled in this study.

Changes in the text Comment 6: The following sentence was added to the Discussion on pages 19-20, lines 435-438: "This study utilized HRCT chest imaging to evaluate treatment response, however future study adding surgical lung biopsy results to further delineate cellular and fibrotic NSIP may be of interest."

Reviewers D Comment 7: Did you have information on patients having experienced an acute exacerbation of RA-ILD and the treatment of ILD in this subgroup of patients?

Reply Comment 7: We agree that this is an important outcome and would be of interest as a focal point as a separate study to further understand this important outcome in RA-ILD. However, we were not able to accurately ascertain this outcome in the registry given heterogeneity and low number of events in an already small cohort.

Changes in the text Comment 7: The following sentence was added to the Discussion on page 16, lines 365 to 367: "We were not able to evaluate response to treatment in those who experienced acute exacerbations as this outcome is difficult to accurately ascertain within the registry."