Peer Review File
Article information: https://dx.doi.org/10.21037/jtd-22-813

Reviewer A

Comment 1: Please expand on the definition of sarcopenia for the reader: is there a range of scores from the mass/strength/physical performance metrics; i.e., what is considered "low" for each of these?

Reply 1: Thank you for your comments. We have added sentences in the methods section.

Changes in the text: Page 10, lines 1,2.
The cut-off criterion for low muscle mass was < 7.0 kg/m2 for men and < 5.7 kg/m2 for women.

Page 10, lines 4-8.
Handgrip strength was measured in the standing position with full elbow extension using an electronic dynamometer (HG-251; N-Force, Wakayama, Japan). Measurements were performed twice for each hand, with the largest grip strength value being used for analysis. The cut-off criterion for low muscle strength was defined as < 28.0 kg for men and < 18.0 kg for women.

Page 10, lines 10,11.
The cut-off criterion for low physical performance was defined as < 1.0 m/s for both sexes.

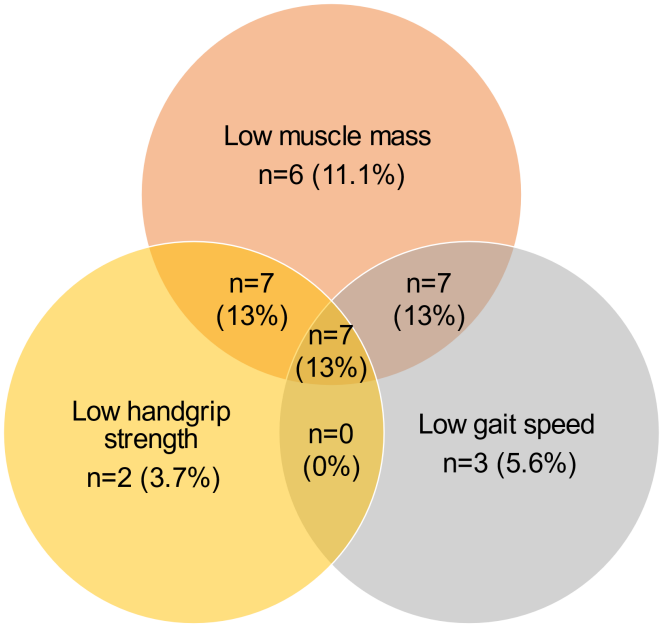
Comment 2: Along the same lines, please include in Table 1 the numbers and percentages of patients who met the definition of sarcopenia. We know 7 patients were low for all three (i.e., severe sarcopenia), but what criteria did the other 14 patients with sarcopenia meet?

Reply 2: Thank you for your comments. We added the numbers and percentages of patients who met the definition of low handgrip strength, low gait speed, and low muscle mass. In addition, we added a Venn diagram in the Supplementary Appendics (Supplementary Figure S1)

Changes in the text: Table1.

Page 13, lines 14,15
Supplementary Figure S1 shows the relationships among low muscle mass, low handgrip strength, and low gait speed.

Supplementary Material. Page 3 (Supplementary Figure S1)



Comment 3: Why were patients who needed supplemental oxygen at rest excluded? A substantial proportion of patients with IPF need O2 at rest; including them would have made results more applicable to the broader population.

Reply 3: Thank you for your comments. To clarify the clinical impact of SARC-F scores of patients with IPF, we excluded patients who required supplemental oxygen therapy at rest since they find it difficult to perform the 6-minute walk test or pulmonary function tests. However, as you have pointed out, this exclusion criterion limited the examination of the prevalence of sarcopenia or the utility of the SARC-F in patients with IPF. Therefore, we have revised the methods and limitations sections.

Changes in the text: Page 8, lines 1-3.
The inclusion criteria included having provided written informed consent and the ability to perform a 6-min walk test (6 MWT) for assessing the exercise capacity in patients with IPF.

Page 8, lines 5-7
To clarify the clinical impact of SARC-F scores in patients with IPF, we excluded patients who required supplemental oxygen therapy at rest given their difficulty to perform 6-MWT or pulmonary function tests.

Page 19, lines 5-7
Fourth, we excluded patients on long-term oxygen treatment at rest. Therefore, our findings may not be representative of the overall population of patients with IPF.

Comment 4: For the SGRQ, CAT and SARC-F, please include in the text possible score ranges and what higher scores connote.

Reply 4: Thank you for your comments. We have added score ranges.

Changes in the text: Page 11, lines 1-6.
The CAT comprises eight items related to respiratory disorder symptoms and their impact. Patients are asked to respond to all items using an identical 0–5 response scale. The total score ranges from 0 to 40, with a score of 0 indicating no impairment. The SGRQ is a specific questionnaire for respiratory diseases that contains three domains, which are all scored from 0 to 100. The score is positively correlated with the impairment of health-related quality of life.

Page 11, lines 7-11.
The HADS is composed of 14 items; among them, seven comprise the anxiety subscale while the remaining seven comprise the depression subscale. Each item is rated on a 0–3 scale. Accordingly, the total subscale score ranges from 0 (no distress) to 21 (maximum distress), with higher scores indicating more severe distress.

Page 12, lines 2,3.
The SARC-F scores range from 0 (best) to 10 (worst). Patients with a total score ≥ 4 were considered to be at risk of sarcopenia (8).

Comment 5: The VIF (variance inflation factor) is not a commonly used metric. Please include explanatory text: what is it? How is it used? How is it to be interpreted?

Reply 5: Thank you for your comments. The VIF is used to detect multicollinearities in linear models (Cheng J, Sun J, Yao K, Xu M, Cao Y. A variable selection method based on mutual information and variance inflation factor. Spectrochim Acta A Mol Biomol Spectrosc. 2022 ;268:120652. doi: 10.1016/j.saa.2021.120652.) However, we decided not to use multivariate analysis in our study.

Comment 6: Page 12, lines 1-9: please report correlations as strengths (weak, moderate strong) according to convention.

Reply 6: Thank you for your comments. We added correlations as strengths.

Changes in the text: Page 13; line18. Page 14; lines 1-7.
The SARC-F score was moderately correlated with the pulmonary function tests, ($P < 0.001$), partial pressure of oxygen at rest ($P < 0.001$), and CAT scores ($P < 0.001$); strongly correlated with the SGRQ score; and weakly correlated with the HADS score ($P < 0.05$). Similar to the SARC-F score, the CAT scores were strongly and weakly correlated with the SGRQ ($P < 0.001$) and HADS scores ($P < 0.05$), respectively

Comment 7: Gender should be included in multivariate models.

Reply 7: Thank you for your comments. Univariate linear regression analysis revealed that sex was not a predictive factor for daily step count. We have added this result in Table 3. As aforementioned, we decided not to perform multivariate analysis.

Comment 8: Stepwise selection is not appropriate. Please choose relevant clinical variables to include in the model, leave in the model and report beta coefficients, 95% CIs and/or p values for all variables in the model.

Reply 8: Thank you for your comments. As you have mentioned, stepwise selection without deliberation was not appropriate. We examined each clinical variable in univariate analyses (Table 3). Only the distance walked during the 6 MWT (Standardized

β 0.33, P= 0.011) and SARC-F score (Standardized β -0.39, P = 0.005) were significant predictors for daily step count. Distance walked during the 6 MWT was correlated with the SAEC-F score (r = -0.62, P < 0.001). Therefore, there was little utility of conducting multivariate analysis.

Changes in the text: Page 15; lines16-18, Page 16; lines 1,2.

Linear regression analyses of the daily step count

Table 3 shows the results of the linear regression analyses for the daily step counts. Distance walked in the 6 MWT (Standardized β 0.33, P = 0.011) and the SARC-F score (Standardized β -0.39, P = 0.005), but not the CAT or SQRQ scores, were significant predictors for daily step count.

Comment 9: There is substantial collinearity among the PROs. It would probably be best not to include SARC-F and SGRQ or SARC-F and CAT in any of the same models.

Reply 9: Thank you for your comments. As you have pointed out, multivariate analysis is not appropriate since there may be collinearity among the PROs. We have deleted these results.

Comment 10: Page 15, lines 1-2 "…SARC-F is useful for assessing the quality of life and daily activity in…" This is incorrect and a vast over-extension of results. True, SARC-F scores are associated with measures of QOL and activity, but I'm sure the authors would agree that SARC-F is no substitute for QOL or activity. It's scores tell us something about QOL and activity in patients with IPF – and this supports the validity of SARC-F as a measure of ?? overall well-being ?? – but SARC-F scores only tell so much about these other outcomes.

Reply 10: Thank you for your comments. We agree with your comments. We corrected phrases and sentences.

Changes in the text: Page 16, lines 12,13.

To our knowledge, this is the first study to show that the SARC-F scores were correlated with measures of quality of life and daily activity in outpatients with IPF.

Comment 11: I have the same issue with "Our results indicate that the SARC-F can

be a useful and convenient questionnaire to assess pulmonary function…" This is not correct. I would say your results support the validity of the SARC-F as a measure of sarcopenia (given the association between gold standard, objective sarcopenia as you've measured and SARC-F scores) and that SARC-F scores are associated with other clinically relevant metrics…thus its usefulness in this population.

Reply 11: Thank you for your comments. We agree with your comments. We deleted some sentences, and corrected phrases and sentences.

Changes in the text: Page 19, lines 8,9
In conclusion, SARC-F scores were associated with health status and daily activity in patients with IPF.

Comment 12: Page 15, line 17: Please avoid every using the sentence stating a questionnaire is validated…I recognize this is commonly done, but it really has no meaning. Scores from a questionnaire can be valid for assessing a construct in a given population, and that is how it should be stated. Validation is not a threshold phenomenon; it's an ongoing process in which hypotheses are tested and knowledge is build around what scores from a questionnaire can and can't tell us about the population under study.

Reply 12: Thank you for your comments. We agree with your comments. According to your suggestions, we deleted some sentences.

Changes in the text: Page 17, line 7
SARC-F is widely used to screen for sarcopenia (9-11).

Comment 13: Given statements above, please revise conclusion.

Reply 13: According to your suggestions, we revised some sentences.

Changes in the text: Page 4, lines 8,9
Conclusions: SARC-F scores were correlated with health status and daily activity in patients with IPF. Further studies are warranted to validate the utility of the SARC-F in patients with idiopathic pulmonary fibrosis.

Page 19, lines 8,9.
In conclusion, SARC-F scores were associated with health status and daily activity in patients with IPF.

Comment 14: Figure 1, panel A: please show 3 groups: no sarcopenia, severe sarcopenia, sarcopenia (can assess for difference with ANOVA and follow with p value corrected pairwise comparisons if you wish)

Reply 14: Thank you for your comments. We compared the robust group, non-severe sarcopenia group, and severe sarcopenia group using ANOVA. There were significant differences in the SARC-F score ($P < 0.001$), but not the CAT score ($P = 0.106$) and daily step count ($P = 0.124$), among the robust, non-severe sarcopenia, and severe sarcopenia groups. This could be attributed to our small sample size. The utility of distinguishing between non-severe and severe sarcopenia in patients with IPF remains unclear in clinical practice..

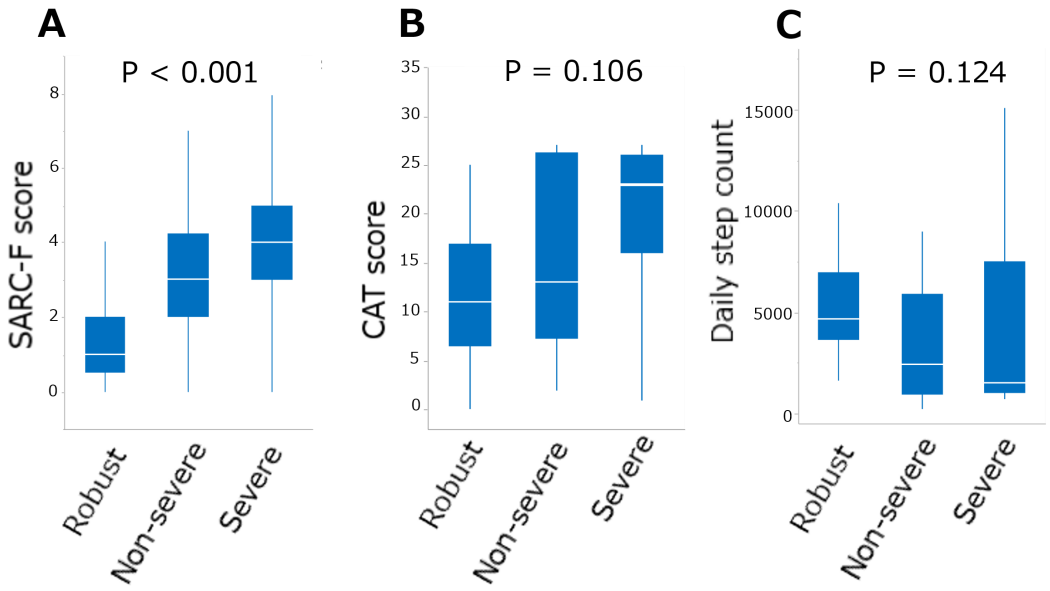
Changes in the text: Page 12, lines 11-13
Comparisons of the SARC-F score, CAT score, and daily step count among the robust group, non-severe sarcopenia group, and severe sarcopenia group were analyzed using an analysis of variance test.

Page 15, lines 3-6
Supplementary Figure S2 (A) (B) shows the comparisons of the SARC-F and CAT scores among the robust, non-severe sarcopenia, and severe sarcopenia groups. There were significant among-group differences in the SARC-F scores ($P < 0.001$) but not in the CAT scores ($P = 0.106$).

Page 15, lines 12-14
Supplementary Figure S2 (C) shows the comparisons of the daily step count among the robust, non-severe sarcopenia, and severe sarcopenia groups. There was no significant among-group difference in the daily step count ($P = 0.124$).

Supplementary Figure S2. Comparisons of the SARC-F score, CAT score, and daily step count among the robust, non-severe sarcopenia, and severe sarcopenia groups.
Comparisons of the SARC-F score (A), CAT score (B), and daily step count (C) among the robust, non-severe sarcopenia, and severe sarcopenia groups are shown. There was a

significant among-group difference in the SARC-F score (P < 0.001) but not in the CAT score (P = 0.106) and daily step count (P = 0.124).
Abbreviations: SARC-F, strength, assistance in walking, rising from a chair, climbing stairs, and falls questionnaire; CAT, chronic obstructive pulmonary disease assessment test.



Page 18, lines 6-10.

As shown in Supplementary Figure S2, there were significant differences in the SARC-F score (P < 0.001), but not the CAT score (P = 0.106) and daily step count (P = 0.124), among the robust, non-severe sarcopenia, and severe sarcopenia groups. This could be attributed to our small sample size. The utility of distinguishing between non-severe and severe sarcopenia in patients with IPF remains unclear in clinical practice.

[COMMENTS 15]

Table 1: Pack-years is packs per day times years smoked. I assume what is reported is packs per day times days in a year…this is incorrect.

ReplyThank you for this insightful comment. We have revised the pack-years.

Changes in the text: Table 1.

Comment 16: Table 2. FVC% r CAT correlation appears to not be significant (r

crosses 0), but p value is reported as <0.05. For all boxes with "<0.05", please change to actual p value.

Reply 16: Thank you for finding this error. We have revised Table 2.

Reviewer B

Comment 1: The final paragraph of the introduction is unclear and should be re-worded to more clearly convey the objectives of this manuscript. For instance, it is not clear from the results that assessing "mental status" of patients with IPF was a primary focus of this manuscript.

Reply 1: We agree with your comments. We have revised the final paragraph of the introduction.

Changes in the text: Page 7, lines 6-8
It remains unclear whether the SARC-F score can predict the quality of life and daily activity in patients with IPF. We aimed to confirm the association of SARC-F scores with the measurements of quality of life and activity.

Comment 2: Why were patients on oxygen excluded from analysis? A justification for this decision should be included.

Reply 2: Thank you for your comments. To clarify the clinical impact of SARC-F scores of patients with IPF, we excluded patients who needed supplemental oxygen therapy at rest. Such patients have difficulty to perform 6-minute walk test or pulmonary function tests. However, as you pointed out, this exclusion criterion was insufficient for examining the prevalence of sarcopenia or the usefulness of SARC-F questionnaires in IPF. Therefore, we revised the methods and limitations.

Changes in the text: Page 8, lines 1-3.
The inclusion criteria included having provided written informed consent and the ability to perform a 6-min walk test (6 MWT) for assessing the exercise capacity in patients with IPF.

Page 8, lines 5-7.
To clarify the clinical impact of SARC-F scores in patients with IPF, we excluded patients who required supplemental oxygen therapy at rest given their difficulty to perform 6-MWT or pulmonary function tests.

Page 19, lines 5-7.
Fourth, we excluded patients on long-term oxygen treatment at rest. Therefore, our findings may not be representative of the overall population of patients with IPF.

Comment 3: How the multivariable models were selected (the criteria for inclusion and exclusion of variables) should be provided.

Reply 3: Thank you for your comments. As you have mentioned, stepwise selection without deliberation was not appropriate. We examined each clinical variable in univariate analyses (Table 3). Only the distance walked during the 6 MWT (Standardized β 0.33, P= 0.011) and SARC-F score (Standardized β -0.39, P = 0.005) were significant predictors for daily step count. Distance walked during the 6 MWT was correlated with the SAEC-F score (r = -0.62, P < 0.001). Therefore, there was little utility of conducting multivariate analysis. We have deleted the results of multivariable models.

Changes in the text: P15; lines 16-18, Page16;lines 1,2.
Linear regression analyses of the daily step count
Table 3 shows the results of the linear regression analyses for the daily step counts. Distance walked in the 6 MWT (Standardized β 0.33, P = 0.011) and the SARC-F score (Standardized β -0.39, P = 0.005), but not the CAT or SQRQ scores, were significant predictors for daily step count.

Comment 4: When reporting results of linear regression in the body of the manuscript, it would be more appropriate to report a measure of the strength of association (such as beta coefficients) in addition to the p-value to provide context to readers.

Reply 4: Thank you for your comments. We have revised sentences and Table 3.

Comment 5: Likewise, when comparing the discrimination of SARC-F and CAT to predict sarcopenia it would be helpful to report the median and interquartile range

of SARC-F and CAT scores by groups. Further, though receiver operator characteristic curves are discussed in the methods, they are not presented in the manuscript and would be helpful. Finally, given the small sample size, it may be best to not report the statistical comparison of AUC between SARC-F and CAT scores to predict sarcopenia.

Reply 5: Thank you for your comments. We have made some revisions according to your insightful comments.

Changes in the text: Page 14, lines 12-14.
The median SARC-F scores in the sarcopenia and non-sarcopenia groups were 3 [2–4] and 1 [1–2], respectively. Additionally, the median CAT scores in the sarcopenia and non-sarcopenia groups were 18 [9–26] and 11 [7–17], respectively.

Page 14; lines 17,18, Page 15; lines 1,2
We performed ROC curve analyses of sarcopenia according to the SARC-F and CAT scores. The areas under the curve (AUC) values of the SARC-F and CAT scores were 0.77 (95% confidence interval [CI], 0.63–0.92, P = 0.001) and 0.66 (95% CI, 0.50–0.63, P = 0.62), respectively.

Page 18. lines 13-15
Given our small sample size, we could not calculate the cut-off SARC-F score for detecting sarcopenia or compare the AUC values of the SARC-F and CAT scores for sarcopenia.

Comment 6: When reporting daily step count in patients with sarcopenia, including a median and interquartile range between groups would be helpful.

Reply 6: Thank you for your comments. We added sentences.

Changes in the text: Page 15, lines 9-11.
There were significant between-group differences in the daily step counts (sarcopenia group, 2138 [986–5490]; non-sarcopenic group, 4652 [3626–6880]; P = 0.005; Figure 1C).

Comment 7: The median smoking pack-years reported of 600 seems unlikely and

may reflect an error.

Reply 7: Thank you for finding this error. We have revised the pack-years.