Quality assessment for included studies

This study included two randomized controlled trials (RCTs) and 13 observational studies. We used the Cochrane risk of bias tool to assess the quality of RCTs, and for observational studies, the Newcastle-Ottawa Scale was adopted (Tables S1,S2).

Grading the quality of evidence

The Grades of Recommendation, Assessment, Development and Evaluation (GRADE) criteria was applied to assess quality of evidence for each outcome.

The quality of evidence for outcomes on forced vital capacity (FVC)% predicted, diffusing capacity of the lung for carbon monoxide (DLCO)% predicted and change in FVC was judged as low due to serious risk of bias, which was upgraded because all plausible confounding would reduce demonstrated effect. The quality of evidence for outcome on six-minute walking distance (6MWD) was judged as very low due to serious risk of bias and serious imprecision, which was upgraded because all plausible confounding would reduce demonstrated effect.

How results can be interpreted given the quality of the included studies?

Taken together, the majority of included studies was observational studies with relatively higher risk of bias than RCTs, although we excluded several observational studies with high risk of bias (assessed a score of ≤5 using the Newcastle-Ottawa Scale) in the meta-analysis. We applied the GRADE criteria to evaluate quality of each evidence, which revealed low quality of the primary results. Therefore, our confidence in the evidence was limited and these findings should be interpretated with caution. However, given limited synthetic analysis on anti-reflux therapy for idiopathic pulmonary fibrosis (IPF) currently, this study could provide a preliminarily systematic review with regard to anti-reflux therapy on pulmonary function in IPF. Additionally, several suggestions for future research were also proposed.

Table S1 Quality assessment for RCTs

Study	Random sequence generation? (selection bias)	Allocation concealment? (selection bias)	Blinding of participants, personnel, and outcome assessors? (performance and detection bias)	Incomplete outcome data? (attrition bias)	Selective reporting? (reporting bias)	Other bias?	Overall risk of bias
Dutta 2019, (23)	Yes	Yes	Yes	No	No	No	Low
Raghu 2018*, (45)	Yes	Yes	No	No	No	No	Low

Risk of bias was assessed using the Cochrane risk of bias tool. *, this RCT was unblinded for participants, personnel, and outcome assessors because the intervention in this RCT was surgery, which was difficult to reach blinding of participants. Meanwhile, the primary outcome, change in FVC, was measured in accordance with ATS/ERS standards. Therefore, the assessed overall risk of bias for this study was also low. RCT, randomized controlled trial; ATS, American Thoracic Society; ERS, European Respiratory Society.

https://dx.doi.org/10.21037/jtd-21-771

Table S2 Quality assessment for observational studies

Study	Selection				Comparability	Outcome			Total score
	Exposed cohort	Nonexposed cohort	Ascertainment of exposure	Outcome of interest		Assessment of outcome	Length of follow-up	Adequacy of follow-up	
Costabel 2018, (37)	*	*	–	–	**	*	*	*	7
Ghebremariam 2015, (26)	*	*	*	–	**	*	*	*	8
Jo 2019, (40)	*	*	–	*	–	*	*	–	5
Kreuter 2016, (38)	*	*	*	–	**	*	*	*	8
Kreuter 2017, (39)	*	*	*	–	**	*	*	*	8
Lee 2011, (28)	*	*	*	–	*	*	*	*	7
Lee 2013, (29)	*	*	*	–	**	*	*	–	7
Linden 2006, (42)	–	*	*	–	**	*	*	–	6
Liu 2017, (27)	*	*	–	–	**	*	*	*	7
Noth 2012, (41)	*	*	–	–	–	*	–	*	4
Raghu 2006, (30)	–	–	*	–	**	*	*	*	6
Raghu 2013[#], (43)	–	–	–	–	–	–	–	–	–
Raghu 2016, (44)	–	–	*	–	**	*	*	*	6

Risk of bias was assessed using the Newcastle-Ottawa Scale. A higher overall score corresponds to a lower risk of bias; a score of ≤5 (out of 9) indicates a high risk of bias. [#], this study was not assessed because it was published in abstract form without reporting detailed information.

https://dx.doi.org/10.21037/jtd-21-771