Official ATS/ERS/JRS/ALAT Clinical Practice Guidelines: Treatment of Idiopathic Pulmonary Fibrosis

An Update of the 2011 Clinical Practice Guideline

Online Supplement

Ganesh Raghu, Bram Rochwerg, Yuan Zhang, Carlos A. Cuello Garcia, Arata Azuma, Juergen Behr, Jan L. Brozek, Harold R. Collard, William Cunningham*, Sakae Homma, Takeshi Johkoh, Fernando J. Martinez, Jeffrey Myers, Shandra L. Protzko, Luca Richeldi, David Rind, Moises Selman, Arthur Theodore, Athol U. Wells, Henk Hoogsteden and Holger J Schünemann

Table of Contents

Question: Should patients with IPF be treated with imatinib, a tyrosine kinase inhibitor?

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	imatinib	no imatinib	Relative (95% CI)	Absolute (95% CI)		
Mortality (follow up: range 52 to 96 weeks; assessed with: all cause mortality)												
1	randomized trial	not serious	not serious	not serious	very serious [1,2]	none	8/59 (13.6%)	10/60 (16.7%)	**RR 0.81** (0.35 to 1.92)	32 fewer per 1000 (from 108 fewer to 153 more)	⊕⊕◯◯ LOW	CRITICAL
Disease progression (follow up: range 52 to 96 weeks; assessed with: FVC change from baseline in liters; better indicated by higher differences)												
1	randomized trial	not serious	not serious	not serious	serious [3]	none	59	60	–	MD **0.01 L lower** (0.13 lower to 0.11 higher)	⊕⊕⊕◯ MODERATE	CRITICAL
Adverse events (follow up: range 52 to 96 weeks)												
1	randomized trial	not serious	not serious	not serious	not serious	none	56/59 (94.9%)	37/60 (61.7%)	**RR 1.54** (1.25 to 1.9)	333 more per 1000 (from 154 more to 555 more)	⊕⊕⊕⊕ HIGH	CRITICAL
Serious Adverse Outcome (follow up: range 52 to 96 weeks)												
1	randomized trial	not serious	not serious	not serious	very serious [1,4]	none	18/59 (28.8%)	19/60 (30.0%)	**RR 0.96** (0.55 to 1.68)	12 fewer per 1000 (from 135 fewer to 204 more)	⊕⊕◯◯ LOW	CRITICAL

MD – mean difference, RR – relative risk

1. Confidence interval that does not exclude an appreciable benefit or an appreciable harm
2. Only 18 events
3. Only 119 patients
4. Only 37 events

Question: Should patients with IPF be treated with anticoagulation?

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	anticoa-gulant	no anticoa-gulant	Relative (95% CI)	Absolute (95% CI)		
Mortality (follow up: range 48-52 weeks)												
1	randomized trial	serious [1]	not serious	not serious	serious [2]	none	14/72 (19.4%)	3/73 (4.1%)	**RR 4.73** (1.42 to 15.77)	153 more per 1000 (from 17 more to 607 more)	⊕⊕◯◯ LOW	CRITICAL
Disease progression (follow up: mean 48 weeks; assessed with: FVC change (L); better indicated by higher values)												
1	randomized trial	serious [1]	not serious	not serious	serious [3]	none	72	73	-	MD **0.04 L lower** (0.12 lower to 0.04 higher)	⊕⊕◯◯ LOW	CRITICAL
Disease progression (follow up: mean 48 weeks; assessed with: FVC decline lower than or equal 10%, where smaller decline is a desirable effect)												
1	randomized trial	serious [1]	not serious	not serious	serious [3]	none	68/72 (94.4%)	64/73 (87.7%)	**RR 1.08** (0.97 to 1.19)	70 more per 1000 (from 26 fewer to 167 more)	⊕⊕◯◯ LOW	CRITICAL
Adverse events (follow up: mean 48 weeks)												
1	randomized trial	serious [1]	not serious	not serious	serious [4]	none	65/72 (90.3%)	61/73 (83.6%)	**RR 1.08** (0.95 to 1.23)	67 more per 1000 (from 42 fewer to 192 more)	⊕⊕◯◯ LOW	CRITICAL
Serious adverse events (follow up: mean 48 weeks)												
1	randomized trial	serious [1]	not serious	not serious	serious [5]	none	21/72 (29.2%)	12/73 (16.4%)	**RR 1.77** (0.94 to 3.33)	127 more per 1000 (from 10 fewer to 383 more)	⊕⊕◯◯ LOW	CRITICAL

FVC – Forced vital capacity, MD – mean difference, RR – relative risk

1. The included study (Noth 2012) was stopped early due to the anticoagulant group 4.7 more times the risk of death than the placebo group.
2. Only 17 events
3. Wide confidence interval that does not exclude an appreciable benefit or no effect, also only 145 patients included.
4. Confidence interval does not exclude an appreciable harm or no effect
5. Only 33 events; confidence interval does not exclude an appreciable harm or no effect

Question: Should patients with IPF be treated with combination Prednisone, Azathioprine, and N-acetylcysteine?

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	NAC/Imuran/Prednisone	Placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality												
1	Randomized trial	Serious [1]	Not serious	Not serious	Very serious [2]	None	8/77 (10.4%)	1/78 (1.3%)	**RR 8.1** (1.04 to 63.26)	91 more per 1000 (from 1 more to 798 more)	⊕◯◯◯ VERY LOW	CRITICAL
Adverse Event												
1	Randomized trial	Serious [1]	Not serious	Serious [3]	Serious [4]	None	68/77 (88.3%)	61/78 (78.2%)	**RR 1.13** (0.98 to 1.3)	102 more per 1000 (from 16 fewer to 235 more)	⊕◯◯◯ VERY LOW	CRITICAL
Disease Progression (assessed with: change in FVC in liters; higher numbers are better)												
1	Randomized trial	Serious [1]	Not serious	Not serious	Not serious	None	77	78	-	Mean **0.01 L higher** (0.14 lower to 0.11 higher)	⊕⊕⊕◯ MODERATE	CRITICAL
Disease Progression (assessed with: DLCO in SI units (higher numbers are better))												
1	Randomized trial	Serious [1]	Not serious	Serious [5]	Not serious	None	77	78	-	MD **0.06 lower** (1.48 lower to 1.35 higher)	⊕⊕◯◯ LOW	CRITICAL
Quality of Life (assessed with: SGRQ (lower numbers are better))												
1	Randomized trial	Serious [1]	Not serious	Not serious	Serious [4]	None	77	78	-	MD **3.2 lower** (10.5 lower to 4.13 higher)	⊕⊕◯◯ LOW	CRITICAL

MD – mean difference, RR – relative risk

1. Study stopped early for harm.
2. Only 9 events; Very wide confidence intervals with low number of events and single study.
3. Varying degrees of adverse events. Some patient important, others less so.
4. Wide confidence interval that do not exclude benefit or harm
5. Unclear in terms of patient importance for this outcome.

Question: Should patients with IPF be treated with ambrisentan, a selective ER-A endothelin-receptor antagonist?

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	ambrisentan	no ambrisentan	Relative (95% CI)	Absolute (95% CI)		
Mortality (follow up: median 52 weeks)												
1	randomized trial	serious [1]	not serious	not serious	serious [2]	none	26/329 (7.9%)	6/163 (3.7%)	**RR 2.15** (0.9 to 5.11)	42 more per 1000 (from 4 fewer to 151 more)	⊕⊕◯◯ LOW	CRITICAL
Mortality and/or disease progression[3] (follow up: median 52 weeks; assessed with: composite of death and/or disease progression)												
1	randomized trial	serious [1]	not serious	not serious	not serious	none	116/329 (35.3%)	34/163 (20.9%)	**RR 1.69** (1.21 to 2.36)	144 more per 1000 (from 44 more to 284 more)	⊕⊕⊕◯ MODERATE	CRITICAL
Disease progression (follow up: median 72 weeks; assessed with: change in FVC in %; better indicated by higher values)												
1	randomized trial	serious [1]	not serious	not serious	serious [4]	none	329	163	-	MD **3.2 lower** (7.39 lower to 0.99 higher)	⊕⊕◯◯ LOW	CRITICAL
Adverse events (follow up: median 52 weeks)												
1	randomized trial	serious [1]	not serious	not serious	not serious	none	278/329 (84.5%)	136/163 (83.4%)	**RR 1.01** (0.93 to 1.1)	8 more per 1000 (from 58 fewer to 83 more)	⊕⊕⊕◯ MODERATE	CRITICAL
Serious adverse events (follow up: median 52 weeks)												
1	randomized trial	serious [1]	not serious	not serious	serious [4]	none	73/329 (22.2%)	25/163 (15.3%)	**RR 1.45** (0.96 to 2.19)	69 more per 1000 (from 6 fewer to 183 more)	⊕⊕◯◯ LOW	CRITICAL

MD – mean difference, RR – relative risk

1. Stopped early for lack of benefit and high likelihood of increased risk of mortality
2. Only 32 events; confidence interval does not exclude an appreciable harm or no effect
3. Disease progression was defined as worsening pulmonary function tests or acute decompensation (unexplained rapid deterioration over 4 wk with increased dyspnea requiring hospitalization and oxygen supplementation > 5 L/min to maintain a resting oxygen saturation [arterial blood gas;SaO2]>90% or PaO2 >55 mmHg [sea level] or 50mmHg [above 1,400 m])
4. The confidence interval does not exclude an appreciable harm or no effect

Question: Should patients with IPF be treated with pirfenidone?

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pirfenidone	Placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality (follow up: 72 weeks)												
5	Randomized trials	Not serious	Not serious	Not serious	Serious [1]	None	41/804 (5.1%)	59/763 (7.7%)	**RR 0.7** (0.47 to 1.02)	23 fewer per 1000 (from 2 more to 41 fewer)	⊕⊕⊕◯ MODERATE	CRITICAL
Acute exacerbation (follow up: 72 weeks; assessed with worsening PFTs or hospitalization)												
4	Randomized trials	Serious [2]	Not serious	Not serious	Serious [3]	None	10/526 (1.9%)	14/486 (2.9%)	**RR 0.69** (0.2 to 2.42)	9 fewer per 1000 (from 23 fewer to 41 more)	⊕⊕◯◯ LOW	CRITICAL
Disease progression (follow up: 72 weeks; assessed with: change in FVC in liters (higher numbers are better))												
4	Randomized trials	Not serious [4]	Not serious	Not serious	Not serious	None	521	485	-	MD **0.23 higher** (0.06 higher to 0.41 higher)	⊕⊕⊕⊕ HIGH	CRITICAL
Disease Progression (assessed with: DLCO (Higher numbers better))												
4	Randomized trials [5]	Not serious	Not serious	Serious [6]	Not serious	None	526	See comment	-	See comment	⊕⊕⊕◯ MODERATE	CRITICAL
Oxygen saturation (higher numbers are better) (follow up: 9 months)												
2	Randomized trials	Not serious	Not serious	Serious [6]	Not serious	None	171	135	-	MD **0.53 higher** (1.01 lower to 2.06 higher)	⊕⊕⊕◯ MODERATE	IMPORTANT
Photosensitivity (follow up: 72 weeks)												
4	Randomized trials	Not serious	Not serious	Not serious	Not serious	None	130/526 (24.7%)	30/489 (6.1%)	**RR 5.3** (1.46 to 19.24)	264 more per 1000 (from 28 more to 1119 more) [1]	⊕⊕⊕⊕ HIGH	IMPORTANT
Anorexia												
5	Randomized trials	Not serious	Not serious	Not serious	Not serious	None	122/804 (15.2%)	36/766 (4.7%)	**RR 2.96** (2.06 to 4.27)	92 more per 1000 (from 50 more to 154 more)	⊕⊕⊕⊕ HIGH	IMPORTANT
Fatigue												
4	Randomized trials	Not serious	Not serious	Serious [7]	Not serious	None	178/695 (25.6%)	120/659 (18.2%)	**RR 1.42** (1 to 2.02)	76 more per 1000 (from 0 fewer to 186 more)	⊕⊕⊕◯ MODERATE	IMPORTANT
Stomach discomfort												
4	Randomized trials	Not serious	Not serious	Serious [7]	Not serious	None	54/526 (10.3%)	10/489 (2.0%)	**RR 4.2** (2.17 to 8.11)	65 more per 1000 (from 24 more to 145 more)	⊕⊕⊕◯ MODERATE	IMPORTANT

MD – mean difference, RR – relative risk

1. Confidence interval does not exclude an appreciable benefit or no efect. Relatively wide confidence intervals. Even if at upper limit of CI, one would not tolerate cost/side effects of drug.
2. One trial stopped early (Azuma et al.) because of perceived benefit in regards to exacerbations. This trial was not included in the other outcomes and therefore only acute exacerbation was downgraded for risk of bias.
3. Only 24 events; confidence interval dies not exclude an appreciable benefit or an apreciable harm.
4. Data were imputed in studies 004 and 006.
5. It is not clear which patients had DLCO measured and the data provided in the primary publications do not allow for pooling of results.
6. The importance of this outcome measure for patients and the relation to patient important outcomes is uncertain.
7. The severity and duration of this outcome (and subsequently the impact on patients) was not clear.

Question: Should patients with IPF be treated with nintedanib, a tyrosine kinase inhibitor?

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**nintedanib**	**no nintedanib**	**Relative (95% CI)**	**Absolute (95% CI)**		
Mortality (follow up: range 52 to 96 weeks; assessed with: all cause mortality)												
3	randomized trials[1]	not serious	not serious	not serious	serious[2]	none	60/981 (6.1%)	42/508 (8.3%)	**RR 0.7** (0.47 to 1.03)	25 fewer per 1000 (from 2 more to 44 fewer)	⊕⊕⊕◯ MODERATE	CRITICAL
Disease progression (follow up: range 52 to 96 weeks; assessed with: FVC (mean observed change in liters; better indicated by higher differences))												
3	randomized trials	serious [3]	not serious	not serious	not serious	none	691	482	not estimable	MD **0.11 higher** (0.08 higher to 0.14 higher)	⊕⊕⊕◯ MODERATE	CRITICAL
Disease progression (follow up: median 52 weeks; assessed with: FVC decline less than or equal 10%, increased probability of lower decline is beneficial)												
3	randomized trials	not serious	not serious	not serious	serious [4]	none	664/977 (68.0%)	304/506 (60.1%)	**RR 1.15** (1.06 to 1.25)	90 more per 1000 (from 36 more to 150 more)	⊕⊕⊕◯ MODERATE	CRITICAL
Adverse events (follow up: range 52 to 96 weeks)												
3	randomized trials	not serious	not serious	not serious	not serious	none	927/981 (94.5%)	456/508 (89.8%)	**RR 1.06** (1.02 to 1.09)	54 more per 1000 (from 18 more to 81 more)	⊕⊕⊕⊕ HIGH	CRITICAL
Serious Adverse Effects (follow up: range 52 to 96 weeks)												
3	randomized trials	not serious	not serious	not serious	not serious	none	284/981 (29.0%)	153/508 (30.1%)	**RR 0.98** (0.83 to 1.16)	6 fewer per 1000 (from 48 more to 51 fewer)	⊕⊕⊕⊕ HIGH	CRITICAL

MD – mean difference, RR – relative risk

1. Two published reports: Richeldi 2011 counted as one RCT, and Richeldi 2014 with two RCTs.
2. Confidence interval does not exclude an appreciable benefit or no difference
3. Lost to follow-up not accounted in the final results
4. Confidence interval does not exclude an appreciable clinical benefit or no difference

Question: Should patients with IPF be treated with anti-acid medication?

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	anti-acid treatment	no anti-acid	Relative (95% CI)	Absolute (95% CI)		
Mortality (follow up: range different for two groups: 694 days (325 to 1,213 days) for those taking GER medications and 624 days (from 292 to 1,134 days) for those not taking GER medications.)												
1	observational study [1]	serious [1]	not serious	not serious	serious [1]	None	96	108	**HR 0.47** (0.24 to 0.93)	not estimable	⊕◯◯◯ VERY LOW	CRITICAL
All-cause mortality (follow up: mean 30 weeks)												
1	observational study	serious [2]	not serious	not serious	serious [3]	None	124	118	11% vs 18%	not estimable	⊕◯◯◯ VERY LOW	CRITICAL
Acute Exacerbation (follow up: mean 30 weeks)												
1	observational study	serious [2]	not serious	not serious	serious [3]	None	124	118	0 vs 12%	not estimable	⊕◯◯◯ VERY LOW	CRITICAL
Hospitalization (follow up: mean 30 weeks)												
1	observational study	serious [2]	not serious	not serious	serious [3]	None	124	118	17% vs 30%	not estimable	⊕◯◯◯ VERY LOW	CRITICAL
Disease progression (measured with FVC change in liters (follow up: mean 30 weeks; Higher numbers indicate better outcome)												
1	observational study	serious [2]	not serious	not serious	serious [4]	None	124	118	not estimable	MD **0.07 higher** (0 higher to 0.14 higher)	⊕◯◯◯ VERY LOW	CRITICAL
Function (measured as change in 6-minute walk distance; follow up: mean 30 weeks; Higher numbers indicate better outcome)												
1	observational study	serious [2]	not serious	not serious	serious [4]	None	124	118	not estimable	MD **35.73 higher** (52.08 lower to 123.54 higher)	⊕◯◯◯ VERY LOW	CRITICAL
abnormal acid GER (follow up: mean 18 months)												
1	observational study	not serious	not serious	serious [5]	serious [6]	None	12/19 (63.2%)	40/46 (87.0%)	**RR 0.73** (0.51 to 1.04)	235 fewer per 1000 (from 35 more to 426 fewer)	⊕◯◯◯ VERY LOW	IMPORTANT

MD – mean difference, RR – relative risk

1. Lee 2011 was a study with 68 of 204 accepting gastroesophageal reflux treatment and the study might have risk of bias in selecting report of outcome because it only reported median survival time and HR (unadjusted and adjusted based on 25 covariates). With 204 participants, the study may lack power to conduct adjusted analysis for 25 covariates. This is also a retrospective analysis of 2 cohorts, no prospective collection of 24-hour pH and/or esophageal impedance testing, or anti-acid treatment information is collected.
2. The indication of antiacid treatment was based on the individual physician's decision. This study is in risk of selecting report of outcomes. For all-cause mortality, acute exacerbation, and all cause hospitalization, this study only reported percentages in 2 groups, which is based on survival analysis. The percentages in antireflux treatment and no treatment group were 11% vs 18%, 0 vs 12% and 17% vs 30%, respectively.
3. According to Lee 2013, 124 of them took anti-acid or H2-antihistamines, while 118 not, for all the three outcomes, there were only small numbers of events.
4. The total sample size was 242, with 124 of them taking anti-acid or H2 drugs, while 118 not taking. The confidence interval of mean difference included 0, and does not exclude “no difference”.
5. Surrogate outcome of lung functional results, not patient important outcome.
6. Only 52 events and the confidence interval not excluding an appreciable benefit or no difference.

Question: Should patients with IPF be treated with sildenafil, a phosphodiesterase-5 inhibitors?

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Sildenafil	Placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality (follow up 6-9 months)												
2	Randomized trials	Not serious	Not serious	Not serious	Very serious [1]	None	2/103 (1.9%)	4/106 (3.8%)	**RR 0.51** (0.1 to 2.72)	18 fewer per 1000 (from 34 fewer to 65 more)	⊕⊕◯◯ LOW	CRITICAL
Exacerbations												
1	Randomized trial	Not serious	Not serious	Not serious	Very serious [2]	None	1/89 (1.1%)	3/91 (3.3%)	**RR 0.34** (0.04 to 3.22)	22 fewer per 1000 (from 32 fewer to 73 more)	⊕⊕◯◯ LOW	CRITICAL
Quality of Life (SGRQ) (higher numbers are worse)												
1	Randomized trial	Not serious	Not serious	Not serious	Serious [3]	None	89	91	-	MD **4.09 lower** (7.31 lower to 0.87 lower)	⊕⊕⊕◯ MODERATE	CRITICAL
Disease progression (assessed with: FVC in Litres (higher numbers are better))												
2	Randomized trials	Not serious	Not serious	Not serious	Serious [4]	None	103	106	-	MD **0.07 higher** (0.2 lower to 0.34 higher)	⊕⊕⊕◯ MODERATE	CRITICAL
Disease Progression (assessed with: DLCO (higher numbers are better))												
2	Randomized trials	Not serious	Not serious	Serious [5]	Serious [4]	None	103	106	-	MD **0.01 lower** (0.33 lower to 0.31 higher)	⊕⊕◯◯ LOW	CRITICAL
Dyspnea (assessed with a change in Borg Dyspnea Score ; higher numbers are worse)												
2	Randomized trials	Not serious	Not serious	Not serious	Serious [4]	None	103	106	-	MD **0.18 lower** (0.61 lower to 0.25 higher)	⊕⊕⊕◯ MODERATE	IMPORTANT
SOBQ Dyspnea Score Change (higher numbers are worse)												
1	Randomized trial	Not serious	Not serious	Not serious	Very serious [6]	None	89	91	-	MD **6.59 lower** (0 higher to 0 higher)	⊕⊕◯◯ LOW	IMPORTANT
Oxygen Saturation (higher numbers are better)												
2	Randomized trials	Not serious	Not serious	Serious [5]	Serious [4]	None	103	106	-	MD **0.04 lower** (0.82 lower to 0.74 higher)	⊕⊕◯◯ LOW	IMPORTANT
Function (measured with Change in 6-minite walk distance (higher numbers are better)												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Sildenafil	Placebo	Relative (95% CI)	Absolute (95% CI)		
2	Randomized trials	Not serious	Not serious	Serious [5]	Serious [4]	None	103	106	-	MD **2.75 higher** (50.99 lower to 45.5 higher)	⨁⨁◯◯ LOW	IMPORTANT

MD – mean difference, RR – relative risk

1. Only 6 events in one study and no events in the other study. confidence interval does not exclude an appreciable benefit or an apreciable harm
2. Only 4 events and confidence interval does not eclude an appreciable benefit or an apreciable harm .
3. Only 180 patients; confidence interval does not exclude an appreciable benefit or no diference
4. Wide confidence intervals with a failure to exclude appreciable benefit or harm with intervention.
5. Unsure clinical significance in terms of patient importance related to this outcome
6. No Standard Deviation supplied - unsure re: imprecision values – difficult to make any conclusions based on this outcome reported in a single trial.

Question: **Should patients with IPF be treated with N-acetylcysteine monotherapy?**

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Acetylcysteine monotherapy**	**other treatments**	**Relative (95% CI)**	**Absolute (95% CI)**		
Mortality (follow up: median 12 months)												
2	randomized trials	not serious	not serious	not serious	very serious [1]	none	6/177 (3.4%)	3/131 (2.3%)	**RR 1.97** (0.5 to 7.71)	22 more per 1000 (from 11 fewer to 154 more)	⊕⊕◯◯ LOW	CRITICAL
Adverse Effects (follow up: median 12 months)												
2	randomized trials	serious [2]	not serious	not serious	serious [23]	none	25/143 (17.5%)	20/143 (14.0%)	**RR 1.23** (0.72 to 2.1)	32 more per 1000 (from 39 fewer to 154 more)	⊕⊕◯◯ LOW	CRITICAL
Quality of Life (follow up: median 12 months; assessed with: change in St George's Respiratory Questionnaire; higher scores are better)												
1	randomized trial	not serious	not serious	not serious	serious [4]	none	133	131	–	MD 1.2 points lower (4.9 lower to 2.4 higher)	⊕⊕⊕◯ MODERATE	CRITICAL
Disease progression (assessed with change in FVC in liters; higher scores are better) (follow up: median 12 months)												
2	randomized trials	not serious	not serious	not serious	Not serious	none	171	169	–	MD **0.02 L higher** (0.04 lower to 0.08 higher)	⊕⊕⊕⊕ HIGH	CRITICAL
Function (assessed with 6-minute walk test in meters; higher scores are better) (follow up: median 12 months)												
2	randomized trials	not serious	serious [6]	serious [7]	serious [8]	none	143	143	–	MD **44.33 meters higher** (2.92 higher to 85.75 higher)	⊕◯◯◯ VERY LOW	CRITICAL

MD – mean difference, RR – relative risk

1. Only 9 events in one study and no events in another study; confidence interval does not exclude an appreciable benefit or an appreciable harm
2. There is a high risk regarding of selective report of outcomes: It is unclear if adverse events were specifically measured, however, authors report that none were observed in Tomika 2005. Martinez 2014 reported adverse events in details. Homma 2012 only reported no difference between two groups but no numbers mentioned, thus this is not included in the meta-analysis..
3. . Only 45 events; confidence interval does not exclude an appreciable benefit or an appreciable harm
4. Martinez 2014 was the only study included in the analysis. The confidence interval does not exclude an appreciable benefit or harm.
5. The confidence interval was wide to exclude an appreciable benefit or harm.
6. .Two RCTs were included in the analysis. Martinez 2014 is a RCT with 264 patients, while Tomika 2005 is a study including 22 patients. The point estimate of mean difference between treatment and control group for Martinez 2014 is 24.1 meters and 66.4 meters for Tomika 2005, indicating point estimates vary across studies.
7. Surrogate outcome.
8. The confidence interval was wide, does not exclude an appreciable benefit, or harm and contain “no difference”.

Question: Should patients with IPF be treated with bosentan or macitentan, dual endothelin-receptor antagonists (ER-A & ER-B)?

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Dual ERA-A	no dual ERA-A	Relative (95% CI)	Absolute (95% CI)		
Mortality (follow up: range 34 to 86 weeks)												
3	randomized trials	not serious	not serious	not serious	very serious[1]	none	23/600 (3.8%)	12/352 (3.4%)	**RR 1.13** (0.57 to 2.27)	4 more per 1000 (from 15 fewer to 43 more)	⊕⊕◯◯ LOW	CRITICAL
Mortality or disease progression [2] (follow up: range 34 to 86 weeks; assessed with: composite of death or disease progression)												
3	randomized trials	not serious	serious [3]	not serious	serious [1]	none	209/597 (35.0%)	141/351 (40.2%)	**RR 0.85** (0.71 to 1)	60 fewer per 1000 (from 0 fewer to 116 fewer)	⊕⊕◯◯ LOW	CRITICAL
Disease progression (follow up: range 52 to 80 weeks; assessed with: change in FVC in liters; better indicated by higher values)												
2	randomized trials	not serious	not serious	not serious	serious [1]	none	522	267	-	MD **0.02 higher** (0.09 lower to 0.13 higher)	⊕⊕⊕◯ MODERATE	CRITICAL
Adverse events (follow up: range 34 to 80 weeks)												
3	randomized trials	not serious	not serious	not serious	not serious	none	702/932 (75.3%)	353/477 (74.0%)	**RR 1.02** (0.96 to 1.07)	15 more per 1000 (from 30 fewer to 52 more)	⊕⊕⊕⊕ HIGH	CRITICAL
Serious Adverse Events (follow up: range 34 to 80 weeks)												
3	randomized trials	not serious	not serious	not serious	not serious	none	188/599 (31.4%)	123/352 (34.9%)	**RR 0.89** (0.74 to 1.08)	38 fewer per 1000 (from 28 more to 91 fewer)	⊕⊕⊕⊕ HIGH	CRITICAL

MD – mean difference, RR – relative risk

1. Only 35 events; confidence interval does not exclude an appreciable harm or an appreciable benefit
2. Disease progression was defined as worsening pulmonary function tests or acute decompensation (unexplained rapid deterioration over 4 wk with increased dyspnea requiring hospitalization and oxygen supplementation > 5 L/min to maintain a resting oxygen saturation [arterial blood gas;SaO2]>90% or PaO2 >55 mmHg [sea level] or 50mmHg [above 1,400 m])
3. Although definition of disease progression did not vary between studies, it was the only significant outcome. Mortality was not different between groups

Question: **Should patients with IPF be treated with bilateral lung transplantation versus single lung transplantation?**

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Bilateral Lung transplantation**	**Single lung transplantation**	**Relative (95% CI)**	**Absolute (95% CI)**		
Mortality (follow up: range 0-17 years)												
3	observational studies	serious[1]	serious [2]	not serious	not serious [3]	none	2527	1491	**HR 0.47** (0.19 to 1.17)[3]	not estimable	⨁◯◯◯ VERY LOW	CRITICAL
Survival time (Mean; Higher numbers indicate better outcome)												
1	observational studies [5]	serious [5]	not serious	not serious	not serious	none	2431	1429	not estimable	8.34 vs 7.37 years	⨁◯◯◯ VERY LOW	CRITICAL

MD – mean difference, RR – relative risk

1. The indication of whether the patient should accept single or bilateral lung transplantation was based on individual surgeon's decision, which was a potential source of risk of bias. Besides, the operative trauma, age, comorbidities, and center experience may make bilateral and single transplantation patients not comparable, which can also induce potential risk of bias.
2. There was inconsistency in the point estimate of the hazard ratios. In addition ,the I^2=83%.
3. The confidence interval cross 1 and the threshold of decision.
4. The pooled result would be **HR = 0.59** (0.30 to 1.14) if the analysis inclued De Oliveira 2012. De Oliveira 2012 only reported the Kaplan-Meire Survival Curve rather than the HR or numbers of events, and this study compared 65 single lung transplantation with 14 bilateral lung transplantation. We reconstructed data from the figure and included it in the meta analysis.
5. Force 2011 reported the mean survival time (2431 single lung transplantation and 1429 bilateral lung transplantation). The mean survival time were 7.37 years for single vs 8.34 years for bilateral lung transplantation in Force 2011. This analysis was not adjusted for potential confounders.

<table>
<tr><th></th><th>Criteria</th><th>Judgements</th><th>Research evidence</th><th>Additional considerations</th></tr>
<tr><td>Problem</td><td>Is there a problem priority?</td><td>○ No
○ Probably no
○ Uncertain
○ Probably yes
○ Yes
○ Varies</td><td></td><td></td></tr>
<tr><td rowspan="2">Benefits & harms of the options</td><td>What is the overall certainty of this evidence?</td><td>○ No included studies
○ Very low
○ Low
○ Moderate
○ High</td><td rowspan="2">The relative importance or values of the main outcomes of interest:
<table><tr><th>Outcome</th><th>Relative importance</th><th>Certainty of the evidence (GRADE)</th></tr><tr><td></td><td></td><td></td></tr></table>Summary of findings: comparison
<table><tr><th>Outcome</th><th>Without intervention</th><th>With intervention</th><th>Difference (95% CI)</th><th>Relative effect (RR) (95% CI)</th></tr><tr><td></td><td></td><td></td><td></td><td></td></tr></table></td><td rowspan="2"></td></tr>
<tr><td>Is there important uncertainty about how much people value the</td><td>○ Important uncertainty or</td></tr>
</table>

	Criteria	Judgements	Research evidence	Additional considerations
	main outcomes?	variability ○ Possibly important uncertainty or variability ○ Probably no important uncertainty of variability ○ No important uncertainty of variability ○ No known undesirable		
	Are the desirable anticipated effects large?	○ No ○ Probably no ○ Uncertain ○ Probably yes ○ Yes ○ Varies		

	Criteria	Judgements	Research evidence	Additional considerations
	Are the undesirable anticipated effects small?	○ No ○ Probably no ○ Uncertain ○ Probably yes ○ Yes ○ Varies		
	Are the desirable effects large relative to undesirable effects?	○ No ○ Probably no ○ Uncertain ○ Probably yes ○ Yes ○ Varies		
Resource use	**Are the resources required small?**	○ No		

	Criteria	Judgements	Research evidence	Additional considerations
		○ Probably no ○ Uncertain ○ Probably yes ○ Yes ○ Varies		
	Is the incremental cost small relative to the net benefits?	○ No ○ Probably no ○ Uncertain ○ Probably yes ○ Yes ○ Varies		
Equity	**What would be the impact on health inequities?**	○ Increased ○ Probably increased		

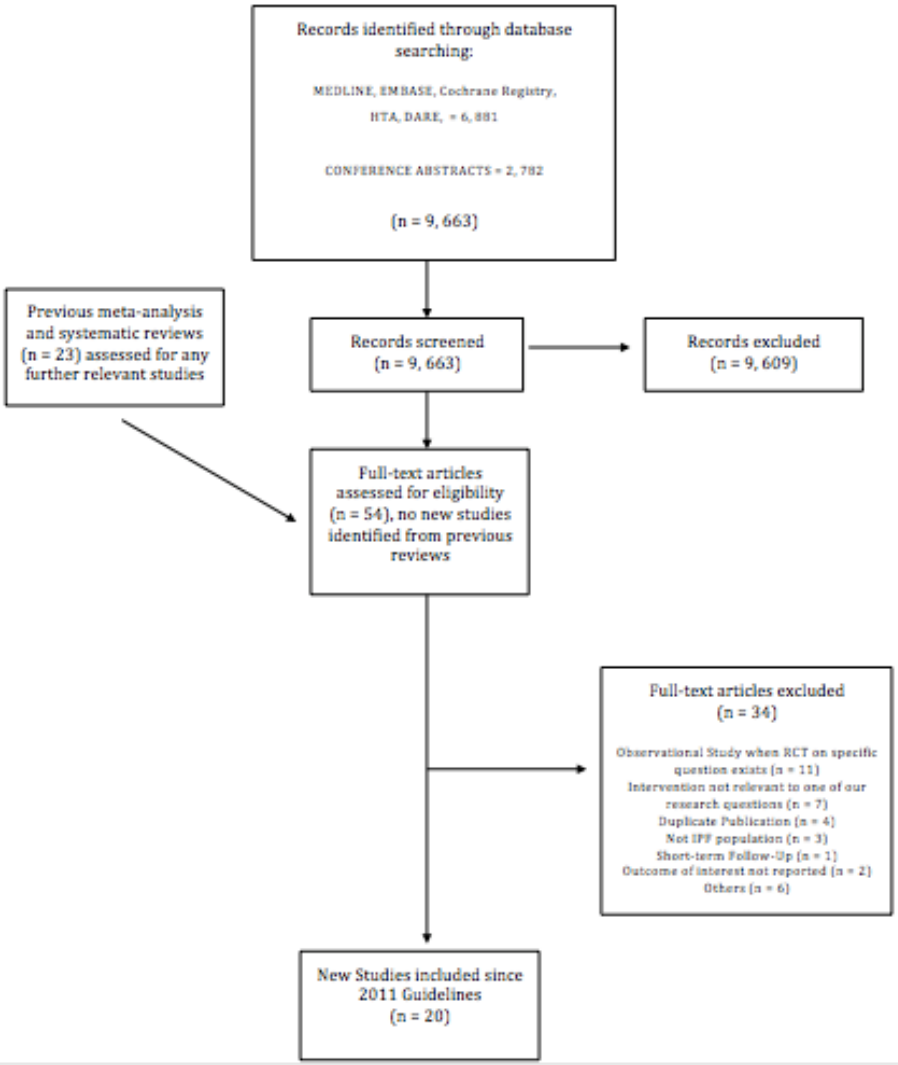
	Criteria	Judgements	Research evidence	Additional considerations
		○ Uncertain ○ Probably reduced ○ Reduced ○ Varies		
Acceptability	**Is the option acceptable to key stakeholders?**	○ No ○ Probably no ○ Uncertain ○ Probably yes ○ Yes ○ Varies		
Feasibility	**Is the option feasible to implement?**	○ No ○ Probably no ○ Uncertain		

	Criteria	Judgements	Research evidence	Additional considerations
		○ Probably yes ○ Yes ○ Varies		

MEDLINE Search Strategy

1	exp pulmonary fibrosis/
2	exp Fibrosis/
3	exp Respiratory Tract Diseases/
4	exp Respiratory System/
5	3 or 4
6	2 and 5
7	1 or 6 [PF Subject Headings]
8	((lung$ or respir$ or pulmonary or alveol$) adj6 fibros$ or fibrotic or fibrous)).tw.
9	(Cystic adj fibro$).mp.
10	8 not 9 [Cystic Fibrosis articles eliminated]
11	(idiopa$ adj2 pulmonary adj2 fibro$).tw.
12	idiopathic pulmonary fibrosis/
13	ipf.mp. and idiopa$ or pulmonary).tw.
14	11 or 12 or 13 [Add IPF back in]
15	7 or 10 or 14 [PF Subject Headings -- Lung/Fibrosis Textword terms]
16	(interstitial$ adj3 fibros$ or fibrotic or fibrous)).tw.
17	5 and 16 [Respiratory Subject Headings -- interstitial fibrosis Textword terms]
18	15 or 17 [PF Subject headings -- PF Textwords]
19	(((usual or ordinary) adj3 interstitial) and pneumo$).tw.
20	(cryptog$ and fibros$ or fibrotic or fibrous) and alveol$).tw.
21	idiopa$ interstitial pneumoni$.tw.
22	19 or 20 or 21 [Other textwords used to describe IPF]
23	Lung Diseases, Interstitial/ or interstitial adj lung adj disease$).tw. or parenchymal.tw. and lung$ or pulmonary)).tw.
24	(((unknow$ or uncertain$) adj4 origin$ or cause$ or aetiol$ or etiol$)) or idiopa$).tw.
25	23 and 24 [Concept to describe idiopathic disease concept of ILD]
26	18 or 22 or 25 [PF Subject headings -- PF Textwords -- IPF Textwords]
27	limit 26 to human [must be indexed with human animal could be included]
28	limit 26 to animal [must be indexed with animal human could be included)]
29	26 not 27 not 28 [citations not indexed with human or animal]
30	27 or 29 [limited to human may include animals) or citations without animal/human limits]
31	201404$.dp,ed,ep.
32	201405$.dp,ed,ep.
33	31 or 32
34	30 and 33
35	limit 34 to english language
36	limit 34 to abstracts
37	35 or 36

Flow chart of search results.



Quality of the Evidence and Implications

Quality of the Evidence (GRADE)	The **quality of the evidence** is a judgment about the extent to which we can be confident that the estimates of effect are correct. These judgments are made using the GRADE system, and are provided for each outcome. The judgments are based on the type of study design (randomized trials versus observational studies), the risk of bias, the consistency of the results across studies, and the precision of the overall estimate across studies. For each outcome, the quality of the evidence is rated as high, moderate, low, or very low using the following definitions:
High (⊕⊕⊕⊕)	Further research is very unlikely to change our confidence in the estimate of effect.
Moderate (⊕⊕⊕◯)	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low (⊕⊕◯◯)	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low (⊕◯◯◯)	We are very uncertain about the estimate. (For more information about the GRADE system, see: www.gradeworkinggroup.org)