GRADE Evidence Profile – Q1: NIV vs usual care in stable patients with COPD

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Long-term NIV	usual care	Relative (95% CI)	Absolute (95% CI)		
Mortality (follow up: range 3 months to 12 months)												
13	randomised trials	serious [a]	not serious	not serious	serious [b]	none	78/422 (18.5%)	101/415 (24.3%)	**RR 0.86** (0.58 to 1.27)	**34 fewer per 1,000** (from 66 more to 102 fewer)	⨁⨁◯◯ LOW	CRITICAL
Number of Hospitalizations (follow up: range 3 months to 12 months)												
3	randomised trials	serious [a]	not serious [c]	not serious	serious [b]	none	154	154	-	MD **1.26 fewer** (2.59 fewer to 0.08 more)	⨁⨁◯◯ LOW	CRITICAL
Quality of Life (higher is better) (follow up: range 3 months to 12 months; assessed with: Multiple validated scales)												
7	randomised trials	serious [a]	serious [d]	not serious	serious [b]	none	244	244	-	SMD **0.49 SD higher** (0.01 lower to 0.98 higher)	⨁◯◯◯ VERY LOW	CRITICAL
Change in Dyspnea Score (assessed with: lower is better)												
5	randomised trials	serious [a]	not serious	not serious	not serious	none	92	96	-	SMD **0.51 SD lower** (0.95 lower to 0.06 lower)	⨁⨁⨁◯ MODERATE	CRITICAL
Change in PaCO2 (follow up: range 3 months to 12 months; assessed with: mmHg)												
12	randomised trials	serious [a]	not serious [c]	not serious	not serious	none	374	373	-	MD **3.37 mmHg lower** (5.75 lower to 0.99 lower)	⨁⨁⨁◯ MODERATE	IMPORTANT
Change in PaO2 (follow up: range 3 months to 12 months; assessed with: mmHg)												
9	randomised trials	serious [a]	not serious	not serious	not serious	none	278	277	-	MD **3.09 mmHg higher** (1.45 higher to 4.74 higher)	⨁⨁⨁◯ MODERATE	IMPORTANT
Change in FEV1 (follow up: range 3 months to 12 months; assessed with: % or L)												
10	randomised trials	serious [a]	not serious	not serious	serious [e]	none	366	374	-	SMD **0.07 SD higher** (0.14 lower to 0.27 higher)	⨁⨁◯◯ LOW	IMPORTANT
Change in FVC (follow up: range 3 months to 12 months; assessed with: & or L)												
8	randomised trials	serious [a]	not serious	not serious	serious [e]	none	287	296	-	SMD **0.1 SD higher** (0.06 lower to 0.26 higher)	⨁⨁◯◯ LOW	IMPORTANT
Change in 6 minute walk distance (follow up: range 3 months to 12 months; assessed with: metres)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Long-term NIV	usual care	Relative (95% CI)	Absolute (95% CI)		
10	randomised trials	serious [a]	not serious	not serious	not serious	none	256	260	-	MD **32.03 metres higher** (10.79 higher to 53.26 higher)	⨁⨁⨁◯ MODERATE	IMPORTANT
Change in Sleep Efficiency (follow up: range 3 months to 12 months)												
3	randomised trials	serious [a]	not serious	not serious	serious [f]	none	61	65	-	SMD **0.55 SD lower** (1.13 lower to 0.03 higher)	⨁⨁◯◯ LOW	IMPORTANT
Minor Adverse Events (follow up: range 3 months to 12 months; assessed with: discomfort, skin break or rash)												
3	randomised trials	serious [a]	not serious	serious [g]	not serious	none	27/189 (14.3%)	0/175 (0.0%)	**RR 10.35** (2.45 to 43.71)	**0 fewer per 1,000** (from 0 fewer to 0 fewer)	⨁⨁◯◯ LOW	IMPORTANT

CI: Confidence interval; **RR:** Risk ratio; **MD:** Mean difference; **SMD:** Standardised mean difference

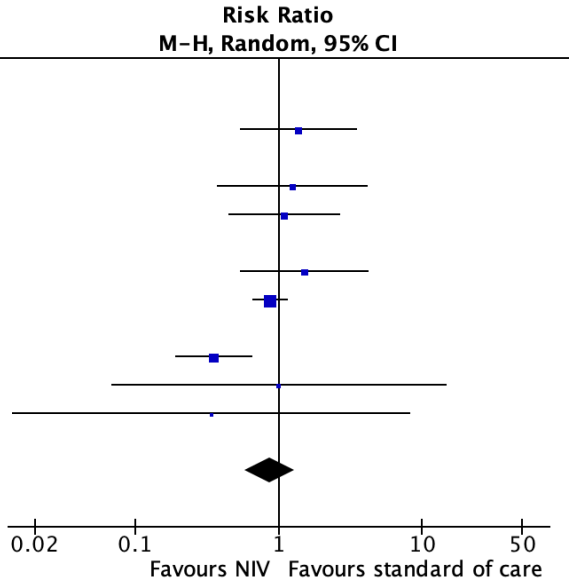
Explanations

a. Most trials unblinded, variable lost-to-follow up which was significant in some included trials.
b. Wide confidence intervals that do not exclude significant harm.
c. High Isqaured (>70%) however in most included studies point estimate is on the side of benefit.
d. High Isquared (>70%) and variable effects between studies.
e. No effect, however confidence intervals don't exclude significant benefit or harm.
f. Wide confidence intervals that do not exclude benefit.
g. Varying importance and severity of these minor adverse effects.

Forest plot 1: Mortality

Study or Subgroup	NIV Events	NIV Total	Standard Care Events	Standard Care Total	Weight	Risk Ratio M-H, Random, 95% CI	Year
Strumpf 1991	0	7	0	7		Not estimable	1991
Gay 1996	0	7	0	6		Not estimable	1996
Kaminski 1999	4	7	5	12	12.3%	1.37 [0.54, 3.47]	1999
Garrod 2000	0	23	0	22		Not estimable	2000
Casanova 2000	5	26	4	26	8.5%	1.25 [0.38, 4.14]	2000
Clini 2002	8	43	8	47	13.0%	1.09 [0.45, 2.66]	2002
Sin 2007	0	11	0	10		Not estimable	2007
Duiverman 2008	8	37	5	35	10.8%	1.51 [0.55, 4.19]	2008
McEvoy 2009	40	72	46	72	31.8%	0.87 [0.66, 1.14]	2009
Bhatt 2013	0	15	0	12		Not estimable	2013
Kohnlein 2014	12	102	31	93	20.0%	0.35 [0.19, 0.65]	2014
Eman Shebi 2015	1	15	1	15	2.1%	1.00 [0.07, 14.55]	2015
Zhou 2017	0	57	1	58	1.5%	0.34 [0.01, 8.15]	2017
Total (95% CI)		**422**		**415**	**100.0%**	**0.86 [0.58, 1.27]**	
Total events	78		101				

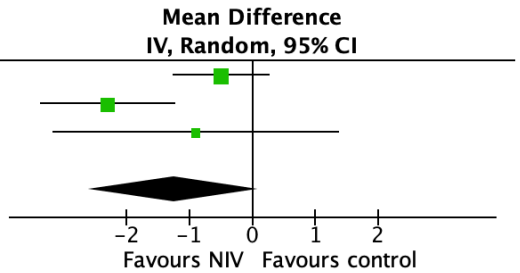
Heterogeneity: Tau^2 = 0.11; Chi^2 = 11.73, df = 7 (P = 0.11); I^2 = 40%
Test for overall effect: Z = 0.76 (P = 0.45)



Forest plot 2: Hospitalizations

Study or Subgroup	NIV Mean	NIV SD	NIV Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI	Year
Clini 2002	0.9	1.2	39	1.4	2.3	47	42.2%	-0.50 [-1.26, 0.26]	2002
Chiang 2004	0.8	1.4	13	3.1	1.4	14	37.4%	-2.30 [-3.36, -1.24]	2004
Kohnlein 2014	2.2	10.2	102	3.1	5.4	93	20.4%	-0.90 [-3.16, 1.36]	2014
Total (95% CI)			**154**			**154**	**100.0%**	**-1.26 [-2.59, 0.08]**	

Heterogeneity: Tau^2 = 0.96; Chi^2 = 7.39, df = 2 (P = 0.02); I^2 = 73%
Test for overall effect: Z = 1.84 (P = 0.07)

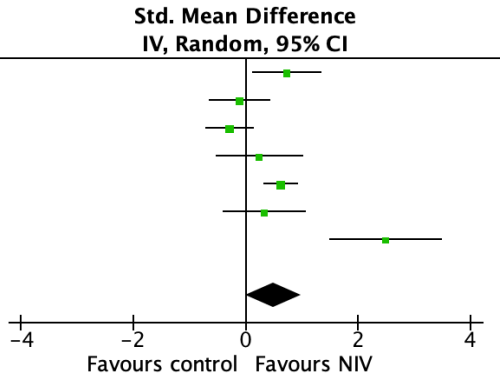


Forest plot 3: Quality of Life

Study or Subgroup	NIV Mean	NIV SD	NIV Total	Standard Care Mean	Standard Care SD	Standard Care Total	Weight	Std. Mean Difference IV, Random, 95% CI	Year
Garrod 2000	24.1	17.4	23	11.8	15.8	22	14.4%	0.73 [0.12, 1.33]	2000
Duiverman 2008	-3.6	12.3	24	-2.3	10.4	32	15.2%	-0.11 [-0.64, 0.42]	2008
McEvoy 2009	5	13	50	10	21.3	40	16.3%	-0.29 [-0.71, 0.13]	2009
Bhatt 2013	0.6	2.58	15	-0.1	3.1	12	12.8%	0.24 [-0.52, 1.00]	2013
Kohnlein 2014	7.5	21.1	102	-6	22.2	93	17.3%	0.62 [0.33, 0.91]	2014
Marquez-Martin 2014	1.1	0.85	15	0.83	0.75	15	13.2%	0.33 [-0.39, 1.05]	2014
Eman Shebi 2015	12	4.9	15	1	3.6	15	10.7%	2.49 [1.50, 3.47]	2015
Total (95% CI)			**244**			**229**	**100.0%**	**0.49 [-0.01, 0.98]**	

Heterogeneity: $Tau^2 = 0.35$; $Chi^2 = 34.92$, df = 6 ($P < 0.00001$); $I^2 = 83\%$

Test for overall effect: Z = 1.93 (P = 0.05)

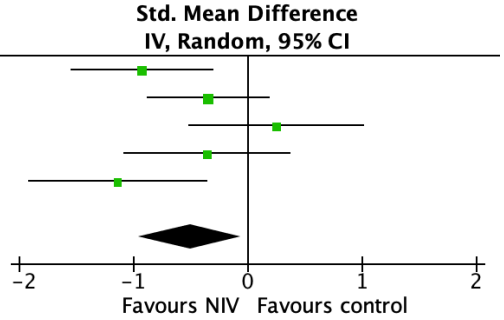


Forest plot 4: Dyspnea

Study or Subgroup	NIV Mean	NIV SD	NIV Total	Standard Care Mean	Standard Care SD	Standard Care Total	Weight	Std. Mean Difference IV, Random, 95% CI	Year
Garrod 2000	-4.94	8.37	23	1.65	5.11	22	21.7%	-0.93 [-1.55, -0.31]	2000
Duiverman 2008	-1.5	4.5	24	0	4.1	32	24.3%	-0.35 [-0.88, 0.19]	2008
Bhatt 2013	-3.4	4.7	15	-4.7	5.4	12	17.8%	0.25 [-0.51, 1.01]	2013
Marquez-Martin 2014	-2	1.5	15	-1.5	1.2	15	18.8%	-0.36 [-1.08, 0.36]	2014
Eman Shebi 2015	-1.1	1.1	15	-0.1	0.5	15	17.4%	-1.14 [-1.92, -0.36]	2015
Total (95% CI)			**92**			**96**	**100.0%**	**-0.51 [-0.95, -0.06]**	

Heterogeneity: $Tau^2 = 0.13$; $Chi^2 = 8.62$, df = 4 ($P = 0.07$); $I^2 = 54\%$

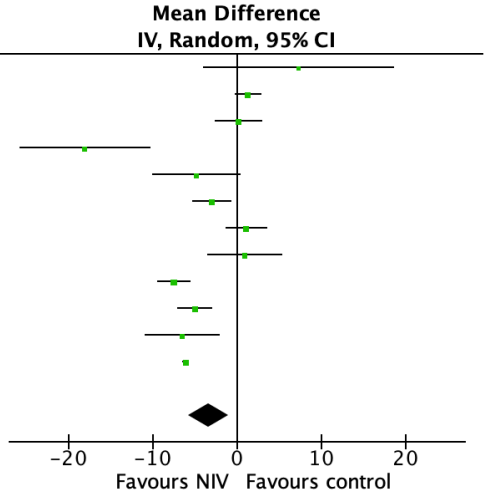
Test for overall effect: Z = 2.24 (P = 0.02)



Forest plot 5: PCO2

Study or Subgroup	NIV Mean	NIV SD	NIV Total	Standard Care Mean	Standard Care SD	Standard Care Total	Weight	Mean Difference IV, Random, 95% CI	Year
Gay 1996	57.5	14.4	7	50.2	4.3	6	3.2%	7.30 [-3.91, 18.51]	1996
Casanova 2000	0.4	2.7	20	-0.9	2.2	24	10.2%	1.30 [-0.17, 2.77]	2000
Garrod 2000	1.3	4.95	23	1.1	4.35	22	9.3%	0.20 [-2.52, 2.92]	2000
Chiang 2004	-12.6	6.5	13	5.5	13	14	5.0%	-18.10 [-25.77, -10.43]	2004
Sin 2007	40.4	6	11	45.2	6	10	7.1%	-4.80 [-9.94, 0.34]	2007
Duiverman 2008	-1.5	2.8	24	1.5	5.6	32	9.7%	-3.00 [-5.24, -0.76]	2008
McEvoy 2009	-0.9	8	72	-2	6.6	72	9.6%	1.10 [-1.30, 3.50]	2009
Bhatt 2013	1.9	6.7	15	1	4.8	12	7.8%	0.90 [-3.44, 5.24]	2013
Kohnlein 2014	-11.25	6	102	-3.75	7.5	93	9.9%	-7.50 [-9.42, -5.58]	2014
Marquez-Martin 2014	-5	3.7	15	0	1.3	15	9.9%	-5.00 [-6.98, -3.02]	2014
Eman Shebi 2015	-8.1	5.3	15	-1.6	6.8	15	7.8%	-6.50 [-10.86, -2.14]	2015
Zhou 2017	-10.41	0.97	57	-4.32	0.68	58	10.6%	-6.09 [-6.40, -5.78]	2017
Total (95% CI)			**374**			**373**	**100.0%**	**-3.37 [-5.75, -0.99]**	

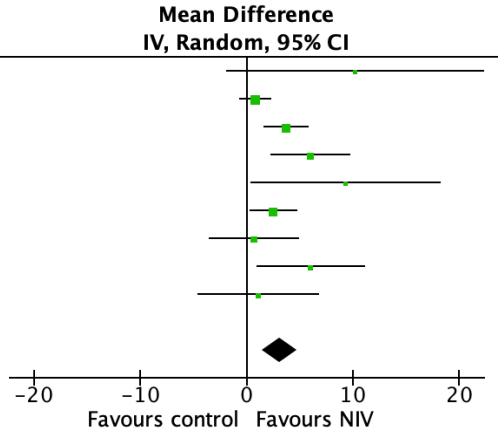
Heterogeneity: $Tau^2 = 13.92$; $Chi^2 = 174.83$, df = 11 ($P < 0.00001$); $I^2 = 94\%$
Test for overall effect: Z = 2.78 (P = 0.005)



Forest plot 6: PO2

Study or Subgroup	NIV Mean	NIV SD	NIV Total	Standard Care Mean	Standard Care SD	Standard Care Total	Weight	Mean Difference IV, Random, 95% CI	Year
Gay 1996	70.5	4.7	7	60.3	14.4	6	1.8%	10.20 [-1.84, 22.24]	1996
Casanova 2000	0.6	2.7	20	-0.2	2.1	24	22.2%	0.80 [-0.65, 2.25]	2000
Garrod 2000	2.3	2.1	23	-1.4	4.35	22	19.2%	3.70 [1.69, 5.71]	2000
Duiverman 2008	4.5	7.8	24	-1.5	5.7	32	11.4%	6.00 [2.31, 9.69]	2008
Bhatt 2013	2.1	13.2	15	-7.2	10.3	12	3.1%	9.30 [0.44, 18.16]	2013
Marquez-Martin 2014	3.5	3.9	15	1	1.8	15	18.3%	2.50 [0.33, 4.67]	2014
Kohnlein 2014	1.43	15	102	0.75	14.5	93	10.0%	0.68 [-3.46, 4.82]	2014
Eman Shebi 2015	7.1	7.4	15	1.1	6.6	15	7.7%	6.00 [0.98, 11.02]	2015
Zhou 2017	72.8	17.7	57	71.7	12.8	58	6.5%	1.10 [-4.55, 6.75]	2017
Total (95% CI)			**278**			**277**	**100.0%**	**3.09 [1.45, 4.74]**	

Heterogeneity: $Tau^2 = 2.64$; $Chi^2 = 16.72$, df = 8 ($P = 0.03$); $I^2 = 52\%$
Test for overall effect: Z = 3.68 (P = 0.0002)

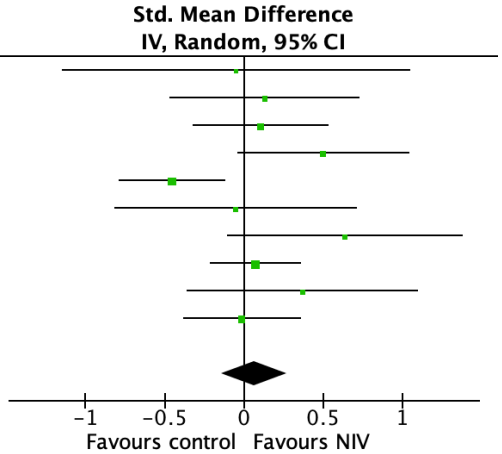


Forest plot 7: FEV1

Study or Subgroup	NIV Mean	NIV SD	NIV Total	Standard Care Mean	Standard Care SD	Standard Care Total	Weight	Std. Mean Difference IV, Random, 95% CI	Year
Gay 1996	-0.02	0.23	7	-0.01	0.1	6	3.1%	-0.05 [-1.14, 1.04]	1996
Casanova 2000	1	8.65	20	0	6.43	24	8.3%	0.13 [-0.46, 0.72]	2000
Clini 2002	0.9	10.2	39	-0.2	10.6	47	12.6%	0.10 [-0.32, 0.53]	2002
Duiverman 2008	-0.03	0.14	24	-0.14	0.26	32	9.5%	0.50 [-0.04, 1.04]	2008
McEvoy 2009	-1.2	9	72	2.4	6.6	72	16.0%	-0.45 [-0.78, -0.12]	2009
Bhatt 2013	0.9	10.1	15	1.4	8	12	5.7%	-0.05 [-0.81, 0.71]	2013
Marquez-Martin 2014	3	7.4	15	-0.5	1.6	15	6.0%	0.64 [-0.10, 1.37]	2014
Kohnlein 2014	0.5	10.3	102	-0.2	9	93	18.1%	0.07 [-0.21, 0.35]	2014
Eman Shebi 2015	2.6	10.7	15	-1.3	9.9	15	6.2%	0.37 [-0.35, 1.09]	2015
Zhou 2017	27.4	19.3	57	27.6	4.8	58	14.6%	-0.01 [-0.38, 0.35]	2017
Total (95% CI)			**366**			**374**	**100.0%**	**0.07 [-0.14, 0.27]**	

Heterogeneity: $Tau^2 = 0.04$; $Chi^2 = 15.01$, df = 9 (P = 0.09); $I^2 = 40\%$

Test for overall effect: Z = 0.63 (P = 0.53)

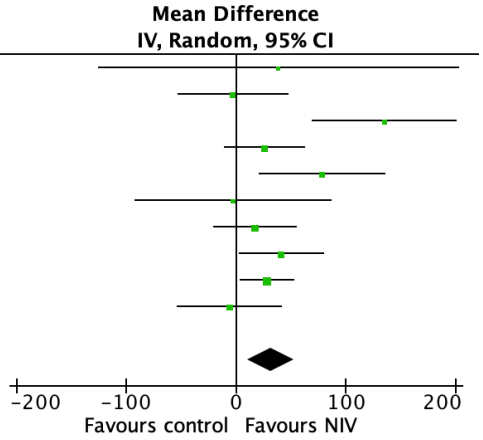


Forest plot 8: Six minute walk distance

Study or Subgroup	NIV Mean	NIV SD	NIV Total	Standard Care Mean	Standard Care SD	Standard Care Total	Weight	Mean Difference IV, Random, 95% CI	Year
Gay 1996	46.39	190.5	7	8	103.6	6	1.6%	38.39 [-125.28, 202.06]	1996
Clini 2002	-18	126.7	39	-15	105.1	47	10.0%	-3.00 [-52.84, 46.84]	2002
Chiang 2004	101.2	81.8	13	-33.8	90.9	14	7.2%	135.00 [69.85, 200.15]	2004
Sin 2007	30	32.2	11	4	49.7	10	13.6%	26.00 [-10.21, 62.21]	2007
Duiverman 2008	-4	44.4	24	-82	155.9	32	8.6%	78.00 [21.14, 134.86]	2008
Bhatt 2013	14	116.3	15	17	117.8	12	4.6%	-3.00 [-91.92, 85.92]	2013
Kohnlein 2014	17.2	117.2	102	0	144.8	93	13.3%	17.20 [-19.99, 54.39]	2014
Marquez-Martin 2014	83	61.9	15	42	42.8	15	13.1%	41.00 [2.92, 79.08]	2014
Eman Shebi 2015	30	35.4	15	2	32.7	15	17.4%	28.00 [3.61, 52.39]	2015
Schneeberger 2017	34	63	15	40	71	16	10.7%	-6.00 [-53.19, 41.19]	2017
Total (95% CI)			**256**			**260**	**100.0%**	**32.03 [10.79, 53.26]**	

Heterogeneity: $Tau^2 = 521.24$; $Chi^2 = 18.02$, df = 9 (P = 0.03); $I^2 = 50\%$

Test for overall effect: Z = 2.96 (P = 0.003)

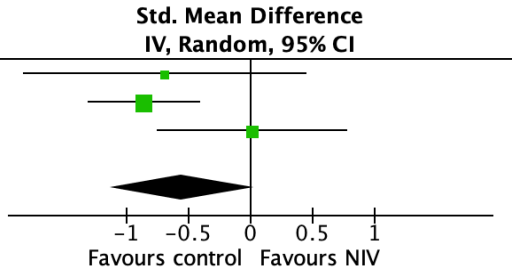


Forest plot 9: Sleep efficiency

Study or Subgroup	NIV Mean	NIV SD	NIV Total	Standard Care Mean	Standard Care SD	Standard Care Total	Weight	Std. Mean Difference IV, Random, 95% CI	Year
Gay 1996	-3	4.1	7	2.8	10.7	6	18.9%	-0.69 [-1.83, 0.45]	1996
Clini 2002	-0.8	0.8	39	0.1	1.2	47	49.4%	-0.86 [-1.30, -0.41]	2002
Bhatt 2013	-0.3	6.5	15	-0.4	8.3	12	31.7%	0.01 [-0.75, 0.77]	2013
Total (95% CI)			**61**			**65**	**100.0%**	**-0.55 [-1.13, 0.03]**	

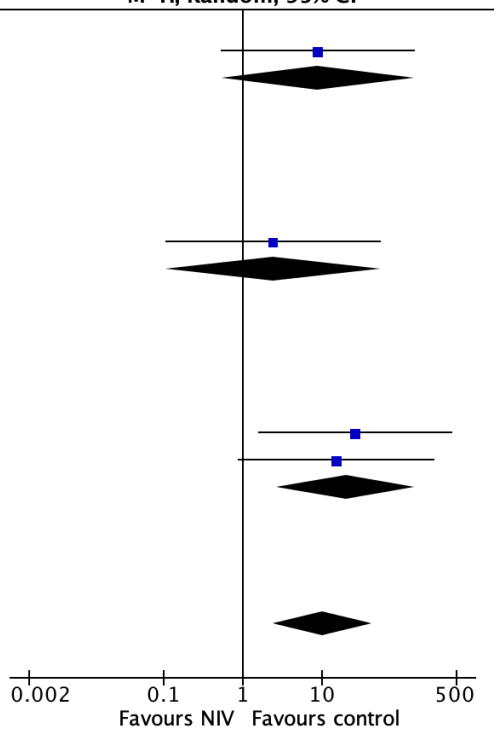
Heterogeneity: $Tau^2 = 0.13$; $Chi^2 = 3.78$, df = 2 (P = 0.15); $I^2 = 47\%$

Test for overall effect: Z = 1.86 (P = 0.06)



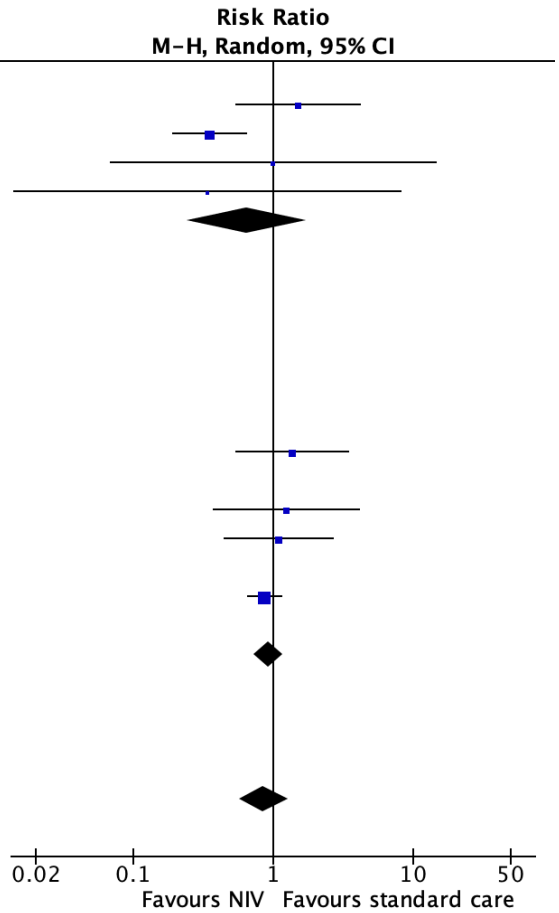
Forest plot 10: Minor adverse events

Study or Subgroup	NIV Events	NIV Total	Standard Care Events	Standard Care Total	Weight	Risk Ratio M-H, Random, 95% CI
1.14.1 Discomfort						
Bhatt 2013	5	15	0	12	26.5%	8.94 [0.54, 147.13]
Subtotal (95% CI)		**15**		**12**	**26.5%**	**8.94 [0.54, 147.13]**
Total events	5		0			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.53 (P = 0.13)						
1.14.2 Skin Break						
Bhatt 2013	1	15	0	12	21.4%	2.44 [0.11, 54.97]
Subtotal (95% CI)		**15**		**12**	**21.4%**	**2.44 [0.11, 54.97]**
Total events	1		0			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.56 (P = 0.58)						
1.14.3 Skin Rash						
Kohnlein 2014	14	102	0	93	26.4%	26.47 [1.60, 437.54]
Zhou 2017	7	57	0	58	25.8%	15.26 [0.89, 261.08]
Subtotal (95% CI)		**159**		**151**	**52.1%**	**20.16 [2.74, 148.35]**
Total events	21		0			
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 0.07$, df = 1 (P = 0.78); $I^2 = 0\%$						
Test for overall effect: Z = 2.95 (P = 0.003)						
Total (95% CI)		**189**		**175**	**100.0%**	**10.35 [2.45, 43.71]**
Total events	27		0			
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 1.44$, df = 3 (P = 0.70); $I^2 = 0\%$						
Test for overall effect: Z = 3.18 (P = 0.001)						
Test for subgroup differences: $Chi^2 = 1.27$, df = 2 (P = 0.53), $I^2 = 0\%$						



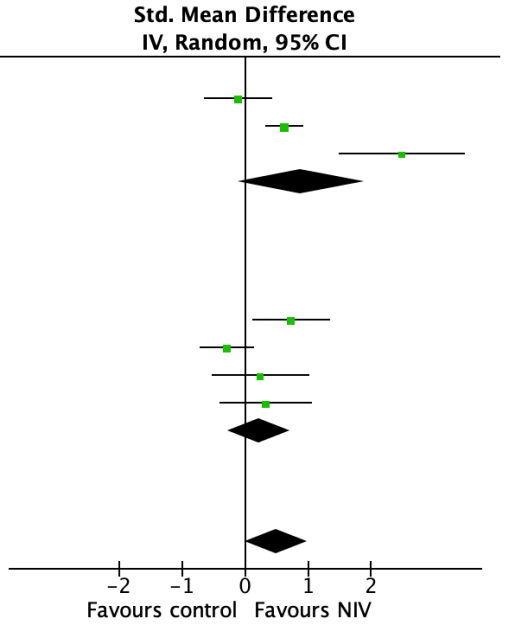
Forest plot 11: Mortality subgroup analysis - targeted PCO2 vs. non-targeted PCO2

Study or Subgroup	NIV Events	NIV Total	Standard Care Events	Standard Care Total	Weight	Risk Ratio M-H, Random, 95% CI	Year
1.1.1 Target N CO2							
Duiverman 2008	8	37	5	35	10.8%	1.51 [0.55, 4.19]	2008
Kohnlein 2014	12	102	31	93	20.0%	0.35 [0.19, 0.65]	2014
Eman Shebi 2015	1	15	1	15	2.1%	1.00 [0.07, 14.55]	2015
Zhou 2017	0	57	1	58	1.5%	0.34 [0.01, 8.15]	2017
Subtotal (95% CI)		**211**		**201**	**34.4%**	**0.65 [0.24, 1.75]**	
Total events	21		38				
Heterogeneity: Tau^2 = 0.45; Chi^2 = 6.12, df = 3 (P = 0.11); I^2 = 51%							
Test for overall effect: Z = 0.84 (P = 0.40)							
1.1.2 Not target CO2							
Strumpf 1991	0	7	0	7		Not estimable	1991
Gay 1996	0	7	0	6		Not estimable	1996
Kaminski 1999	4	7	5	12	12.3%	1.37 [0.54, 3.47]	1999
Garrod 2000	0	23	0	22		Not estimable	2000
Casanova 2000	5	26	4	26	8.5%	1.25 [0.38, 4.14]	2000
Clini 2002	8	43	8	47	13.0%	1.09 [0.45, 2.66]	2002
Sin 2007	0	11	0	10		Not estimable	2007
McEvoy 2009	40	72	46	72	31.8%	0.87 [0.66, 1.14]	2009
Bhatt 2013	0	15	0	12		Not estimable	2013
Subtotal (95% CI)		**211**		**214**	**65.6%**	**0.93 [0.73, 1.18]**	
Total events	57		63				
Heterogeneity: Tau^2 = 0.00; Chi^2 = 1.32, df = 3 (P = 0.72); I^2 = 0%							
Test for overall effect: Z = 0.61 (P = 0.54)							
Total (95% CI)		**422**		**415**	**100.0%**	**0.86 [0.58, 1.27]**	
Total events	78		101				
Heterogeneity: Tau^2 = 0.11; Chi^2 = 11.73, df = 7 (P = 0.11); I^2 = 40%							
Test for overall effect: Z = 0.76 (P = 0.45)							
Test for subgroup differences: Chi^2 = 0.45, df = 1 (P = 0.50), I^2 = 0%							



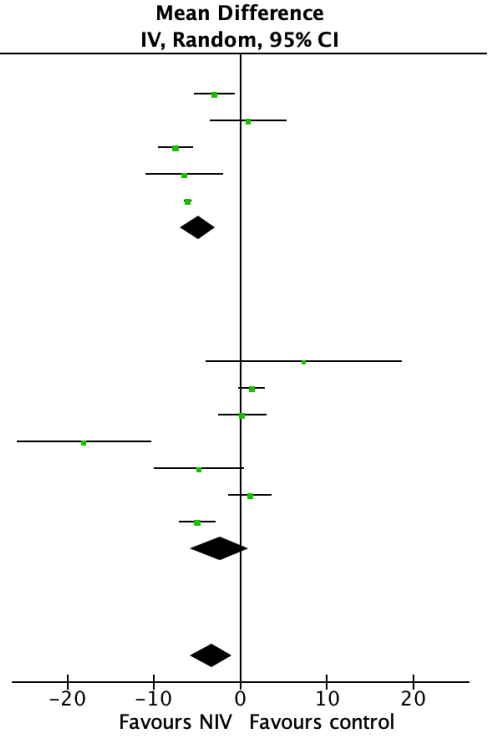
Forest plot 12: Quality of Life subgroup analysis - targeted PCO2 vs. non-targeted PCO2

Study or Subgroup	NIV Mean	NIV SD	NIV Total	Standard Care Mean	Standard Care SD	Standard Care Total	Weight	Std. Mean Difference IV, Random, 95% CI	Year
1.6.1 Target N CO2									
Duiverman 2008	-3.6	12.3	24	-2.3	10.4	32	15.2%	-0.11 [-0.64, 0.42]	2008
Kohnlein 2014	7.5	21.1	102	-6	22.2	93	17.3%	0.62 [0.33, 0.91]	2014
Eman Shebi 2015	12	4.9	15	1	3.6	15	10.7%	2.49 [1.50, 3.47]	2015
Subtotal (95% CI)			**141**			**140**	**43.2%**	**0.89 [-0.12, 1.91]**	
Heterogeneity: Tau^2 = 0.70; Chi^2 = 21.13, df = 2 (P < 0.0001); I^2 = 91%									
Test for overall effect: Z = 1.72 (P = 0.09)									
1.6.2 Not target CO2									
Garrod 2000	24.1	17.4	23	11.8	15.8	22	14.4%	0.73 [0.12, 1.33]	2000
McEvoy 2009	5	13	50	10	21.3	40	16.3%	-0.29 [-0.71, 0.13]	2009
Bhatt 2013	0.6	2.58	15	-0.1	3.1	12	12.8%	0.24 [-0.52, 1.00]	2013
Marquez-Martin 2014	1.1	0.85	15	0.83	0.75	15	13.2%	0.33 [-0.39, 1.05]	2014
Subtotal (95% CI)			**103**			**89**	**56.8%**	**0.22 [-0.28, 0.71]**	
Heterogeneity: Tau^2 = 0.16; Chi^2 = 7.94, df = 3 (P = 0.05); I^2 = 62%									
Test for overall effect: Z = 0.85 (P = 0.39)									
Total (95% CI)			**244**			**229**	**100.0%**	**0.49 [-0.01, 0.98]**	
Heterogeneity: Tau^2 = 0.35; Chi^2 = 34.92, df = 6 (P < 0.00001); I^2 = 83%									
Test for overall effect: Z = 1.93 (P = 0.05)									
Test for subgroup differences: Chi^2 = 1.38, df = 1 (P = 0.24), I^2 = 27.5%									



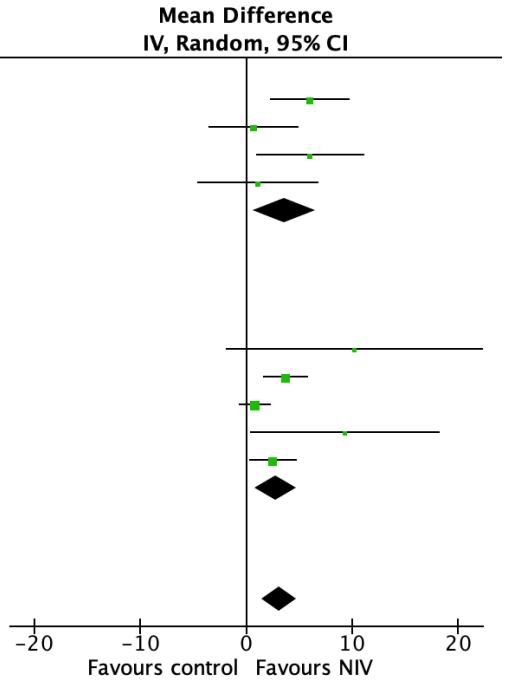
Forest plot 13: PCO2 subgroup analysis - targeted PCO2 vs. non-targeted PCO2

Study or Subgroup	NIV Mean	NIV SD	NIV Total	Standard Care Mean	Standard Care SD	Standard Care Total	Weight	Mean Difference IV, Random, 95% CI	Year
1.7.1 Target N CO2									
Duiverman 2008	-1.5	2.8	24	1.5	5.6	32	9.7%	-3.00 [-5.24, -0.76]	2008
Bhatt 2013	1.9	6.7	15	1	4.8	12	7.8%	0.90 [-3.44, 5.24]	2013
Kohnlein 2014	-11.25	6	102	-3.75	7.5	93	9.9%	-7.50 [-9.42, -5.58]	2014
Eman Shebi 2015	-8.1	5.3	15	-1.6	6.8	15	7.8%	-6.50 [-10.86, -2.14]	2015
Zhou 2017	-10.41	0.97	57	-4.32	0.68	58	10.6%	-6.09 [-6.40, -5.78]	2017
Subtotal (95% CI)			**213**			**210**	**45.8%**	**-4.92 [-6.94, -2.90]**	
Heterogeneity: $Tau^2 = 3.57$; $Chi^2 = 19.25$, df = 4 (P = 0.0007); $I^2 = 79\%$									
Test for overall effect: Z = 4.77 (P < 0.00001)									
1.7.2 Not target CO2									
Gay 1996	57.5	14.4	7	50.2	4.3	6	3.2%	7.30 [-3.91, 18.51]	1996
Casanova 2000	0.4	2.7	20	-0.9	2.2	24	10.2%	1.30 [-0.17, 2.77]	2000
Garrod 2000	1.3	4.95	23	1.1	4.35	22	9.3%	0.20 [-2.52, 2.92]	2000
Chiang 2004	-12.6	6.5	13	5.5	13	14	5.0%	-18.10 [-25.77, -10.43]	2004
Sin 2007	40.4	6	11	45.2	6	10	7.1%	-4.80 [-9.94, 0.34]	2007
McEvoy 2009	-0.9	8	72	-2	6.6	72	9.6%	1.10 [-1.30, 3.50]	2009
Marquez-Martin 2014	-5	3.7	15	0	1.3	15	9.9%	-5.00 [-6.98, -3.02]	2014
Subtotal (95% CI)			**161**			**163**	**54.2%**	**-2.35 [-5.70, 0.99]**	
Heterogeneity: $Tau^2 = 15.18$; $Chi^2 = 51.80$, df = 6 (P < 0.00001); $I^2 = 88\%$									
Test for overall effect: Z = 1.38 (P = 0.17)									
Total (95% CI)			**374**			**373**	**100.0%**	**-3.37 [-5.75, -0.99]**	
Heterogeneity: $Tau^2 = 13.92$; $Chi^2 = 174.83$, df = 11 (P < 0.00001); $I^2 = 94\%$									
Test for overall effect: Z = 2.78 (P = 0.005)									
Test for subgroup differences: $Chi^2 = 1.65$, df = 1 (P = 0.20), $I^2 = 39.5\%$									



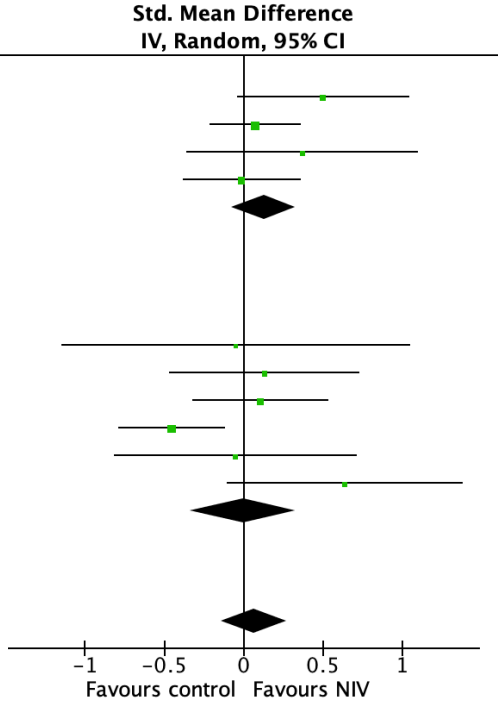
Forest plot 14: PO2 subgroup analysis - targeted PCO2 vs. non-targeted PCO2

Study or Subgroup	NIV Mean	NIV SD	NIV Total	Standard Care Mean	Standard Care SD	Standard Care Total	Weight	Mean Difference IV, Random, 95% CI	Year
1.8.1 Target N CO2									
Duiverman 2008	4.5	7.8	24	-1.5	5.7	32	11.4%	6.00 [2.31, 9.69]	2008
Kohnlein 2014	1.43	15	102	0.75	14.5	93	10.0%	0.68 [-3.46, 4.82]	2014
Eman Shebi 2015	7.1	7.4	15	1.1	6.6	15	7.7%	6.00 [0.98, 11.02]	2015
Zhou 2017	72.8	17.7	57	71.7	12.8	58	6.5%	1.10 [-4.55, 6.75]	2017
Subtotal (95% CI)			**198**			**198**	**35.5%**	**3.60 [0.63, 6.57]**	
Heterogeneity: Tau^2 = 3.82; Chi^2 = 5.15, df = 3 (P = 0.16); I^2 = 42%									
Test for overall effect: Z = 2.37 (P = 0.02)									
1.8.2 Not target CO2									
Gay 1996	70.5	4.7	7	60.3	14.4	6	1.8%	10.20 [-1.84, 22.24]	1996
Garrod 2000	2.3	2.1	23	-1.4	4.35	22	19.2%	3.70 [1.69, 5.71]	2000
Casanova 2000	0.6	2.7	20	-0.2	2.1	24	22.2%	0.80 [-0.65, 2.25]	2000
Bhatt 2013	2.1	13.2	15	-7.2	10.3	12	3.1%	9.30 [0.44, 18.16]	2013
Marquez-Martin 2014	3.5	3.9	15	1	1.8	15	18.3%	2.50 [0.33, 4.67]	2014
Subtotal (95% CI)			**80**			**79**	**64.5%**	**2.79 [0.76, 4.82]**	
Heterogeneity: Tau^2 = 2.54; Chi^2 = 9.92, df = 4 (P = 0.04); I^2 = 60%									
Test for overall effect: Z = 2.70 (P = 0.007)									
Total (95% CI)			**278**			**277**	**100.0%**	**3.09 [1.45, 4.74]**	
Heterogeneity: Tau^2 = 2.64; Chi^2 = 16.72, df = 8 (P = 0.03); I^2 = 52%									
Test for overall effect: Z = 3.68 (P = 0.0002)									
Test for subgroup differences: Chi^2 = 0.19, df = 1 (P = 0.66), I^2 = 0%									



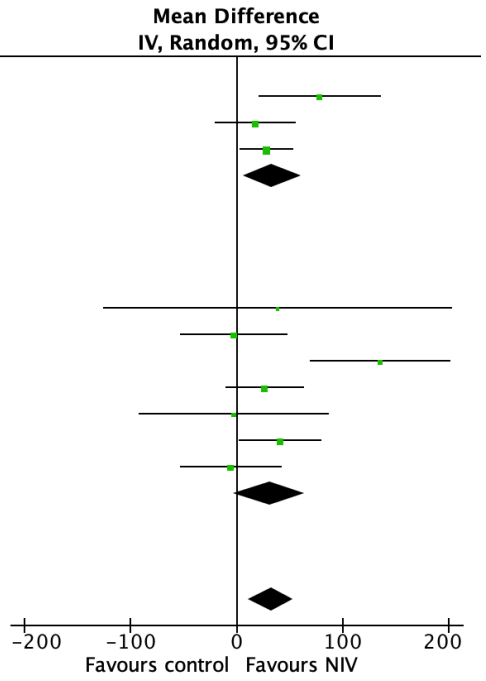
Forest plot 15: FEV1 subgroup analysis - targeted PCO2 vs. non-targeted PCO2

Study or Subgroup	NIV Mean	NIV SD	NIV Total	Standard Care Mean	Standard Care SD	Standard Care Total	Weight	Std. Mean Difference IV, Random, 95% CI	Year
1.9.1 Target N CO2									
Duiverman 2008	-0.03	0.14	24	-0.14	0.26	32	9.5%	0.50 [-0.04, 1.04]	2008
Kohnlein 2014	0.5	10.3	102	-0.2	9	93	18.1%	0.07 [-0.21, 0.35]	2014
Eman Shebi 2015	2.6	10.7	15	-1.3	9.9	15	6.2%	0.37 [-0.35, 1.09]	2015
Zhou 2017	27.4	19.3	57	27.6	4.8	58	14.6%	-0.01 [-0.38, 0.35]	2017
Subtotal (95% CI)			**198**			**198**	**48.4%**	**0.13 [-0.07, 0.32]**	
Heterogeneity: Tau^2 = 0.00; Chi^2 = 2.99, df = 3 (P = 0.39); I^2 = 0%									
Test for overall effect: Z = 1.26 (P = 0.21)									
1.9.2 Not target CO2									
Gay 1996	-0.02	0.23	7	-0.01	0.1	6	3.1%	-0.05 [-1.14, 1.04]	1996
Casanova 2000	1	8.65	20	0	6.43	24	8.3%	0.13 [-0.46, 0.72]	2000
Clini 2002	0.9	10.2	39	-0.2	10.6	47	12.6%	0.10 [-0.32, 0.53]	2002
McEvoy 2009	-1.2	9	72	2.4	6.6	72	16.0%	-0.45 [-0.78, -0.12]	2009
Bhatt 2013	0.9	10.1	15	1.4	8	12	5.7%	-0.05 [-0.81, 0.71]	2013
Marquez-Martin 2014	3	7.4	15	-0.5	1.6	15	6.0%	0.64 [-0.10, 1.37]	2014
Subtotal (95% CI)			**168**			**176**	**51.6%**	**-0.00 [-0.33, 0.33]**	
Heterogeneity: Tau^2 = 0.08; Chi^2 = 9.72, df = 5 (P = 0.08); I^2 = 49%									
Test for overall effect: Z = 0.02 (P = 0.98)									
Total (95% CI)			**366**			**374**	**100.0%**	**0.07 [-0.14, 0.27]**	
Heterogeneity: Tau^2 = 0.04; Chi^2 = 15.01, df = 9 (P = 0.09); I^2 = 40%									
Test for overall effect: Z = 0.63 (P = 0.53)									
Test for subgroup differences: Chi^2 = 0.44, df = 1 (P = 0.51), I^2 = 0%									



Forest plot 16: Six minute walk distance subgroup analysis - targeted PCO2 vs. non-targeted PCO2

Study or Subgroup	NIV Mean	NIV SD	NIV Total	Standard Care Mean	Standard Care SD	Standard Care Total	Weight	Mean Difference IV, Random, 95% CI	Year
1.11.1 Target N CO2									
Duiverman 2008	-4	44.4	24	-82	155.9	32	8.6%	78.00 [21.14, 134.86]	2008
Kohnlein 2014	17.2	117.2	102	0	144.8	93	13.3%	17.20 [-19.99, 54.39]	2014
Eman Shebi 2015	30	35.4	15	2	32.7	15	17.4%	28.00 [3.61, 52.39]	2015
Subtotal (95% CI)			**141**			**140**	**39.3%**	**33.33 [6.53, 60.12]**	
Heterogeneity: Tau^2 = 218.37; Chi^2 = 3.21, df = 2 (P = 0.20); I^2 = 38%									
Test for overall effect: Z = 2.44 (P = 0.01)									
1.11.2 Not target CO2									
Gay 1996	46.39	190.5	7	8	103.6	6	1.6%	38.39 [-125.28, 202.06]	1996
Clini 2002	-18	126.7	39	-15	105.1	47	10.0%	-3.00 [-52.84, 46.84]	2002
Chiang 2004	101.2	81.8	13	-33.8	90.9	14	7.2%	135.00 [69.85, 200.15]	2004
Sin 2007	30	32.2	11	4	49.7	10	13.6%	26.00 [-10.21, 62.21]	2007
Bhatt 2013	14	116.3	15	17	117.8	12	4.6%	-3.00 [-91.92, 85.92]	2013
Marquez-Martin 2014	83	61.9	15	42	42.8	15	13.1%	41.00 [2.92, 79.08]	2014
Schneeberger 2017	34	63	15	40	71	16	10.7%	-6.00 [-53.19, 41.19]	2017
Subtotal (95% CI)			**115**			**120**	**60.7%**	**30.71 [-2.39, 63.80]**	
Heterogeneity: Tau^2 = 1062.87; Chi^2 = 14.78, df = 6 (P = 0.02); I^2 = 59%									
Test for overall effect: Z = 1.82 (P = 0.07)									
Total (95% CI)			**256**			**260**	**100.0%**	**32.03 [10.79, 53.26]**	
Heterogeneity: Tau^2 = 521.24; Chi^2 = 18.02, df = 9 (P = 0.03); I^2 = 50%									
Test for overall effect: Z = 2.96 (P = 0.003)									
Test for subgroup differences: Chi^2 = 0.01, df = 1 (P = 0.90), I^2 = 0%									



GRADE Evidence Profile – Q2: NIV vs usual care after an exacerbation of COPD

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other consid-erations	Long-term NIV	Usual care	Relative (95% CI)	Absolute (95% CI)		
Mortality (follow up: range 1 years to 2 years)												
4	RCTs	serious [a]	not serious	not serious	serious [b]	none	55/201 (27.4%)	62/205 (30.2%)	**RR 0.92** (0.67 to 1.25)	**24 fewer per 1,000** (from 76 more to 100 fewer)	⨁⨁◯◯ LOW	CRITICAL
Exacerbations per year (follow up: range 1 years to 2 years)												
3	RCTs	serious [c]	not serious	not serious	serious [b]	none	181	185	-	SMD **0.19 SD lower** (0.40 lower to 0.01 higher)	⨁⨁◯◯ LOW	CRITICAL
Hospitalizations (follow up: range 1 years to 2 years)												
3	RCTs	serious [a]	serious [d]	not serious	serious [b]	none	71/181 (39.2%)	93/185 (50.3%)	**RR 0.61** (0.30 to 1.24)	**196 fewer per 1,000** (from 121 more to 352 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Dyspnea score (follow up: range 1 years to 2 years; assessed with: Medical Research Council Dyspnea (MRC) Score)												
2	RCTs	serious [c]	not serious [e]	not serious	serious [b]	none	69	71	-	MD **0.8 lower** (2.17 lower to 0.58 higher)	⨁⨁◯◯ LOW	CRITICAL
Quality of Life (follow up: range 1 years to 2 years; assessed with: Severe Respiratory Insufficiency Questionnaire)												
2	RCTs	serious [c]	not serious [e]	not serious	serious [b]	none	85	77	-	MD **2.89 points higher** (6.8 higher to 1.03 lower)	⨁⨁◯◯ LOW	CRITICAL
PaO2 (follow up: range 6 months to 2 years)												
4	RCTs	serious [c]	serious [d]	not serious	serious [b]	none	107	99	-	MD **1.53 mmHg lower** (4.24 lower to 1.17 higher)	⨁◯◯◯ VERY LOW	IMPORTANT
PaCO2 (follow up: range 6 months to 2 years)												
5	RCTs	serious [c]	not serious	not serious	not serious	none	134	126	-	MD **3.41 mmHg lower** (4.09 lower to 2.73 lower)	⨁⨁⨁◯ MODERATE	IMPORTANT
Exercise tolerance (follow up: range 6 months to 2 years; assessed with: 6 minute walk test)												

2	RCTs	serious [c]	very serious [f]	not serious	serious [b]	none	30	25	-	MD **8.64 m lower** (209 lower to 192 higher)	⊕○○○ VERY LOW	IMPORTANT
FEV1 (follow up: range 6 months to 1 years)												
2	RCTs	serious [c]	not serious	not serious	serious [b]	none	58	51	-	SMD **0.36 SD lower** (0.74 lower to 0.03 higher)	⊕⊕○○ LOW	IMPORTANT

CI: Confidence interval; **RCT:** Randomized controlled trial; **RR:** Risk ratio; **MD:** Mean difference; **SMD:** Standardised mean difference

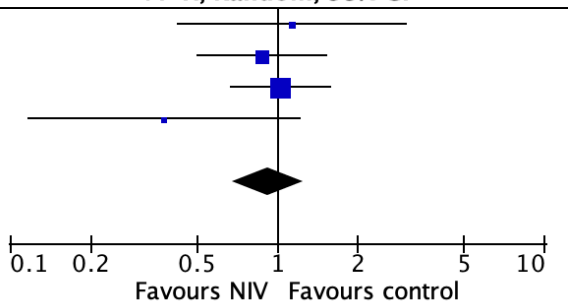
Explanations

a. Lack of blinding of patients and clinicians which could result in cointervention.
b. Wide 95% confidence intervals which do not exclude significant benefit nor significant harm.
c. Lack of blinding of patients and clinicians in most studies which could result in cointervention and/or biased assessment of subjective outcomes, as well as significant loss to follow-up for end-of-study measurements.
d. I-squared (I^2) values high, with individual studies on different sides of the line of no effect.
e. Though high I-squared (I^2) values, all point estimates are on the side of benefit.
f. Very high I-squared (I^2) values with studies on each side of the line of no effect.

Forest plot 1: Mortality

Study or Subgroup	NPPV Events	NPPV Total	Control Events	Control Total	Weight	Risk Ratio M-H, Random, 95% CI
Cheung 2010	6	23	6	26	9.9%	1.13 [0.42, 3.02]
Murphy 2017	16	57	19	59	31.0%	0.87 [0.50, 1.52]
Struik 2014	30	101	29	100	52.1%	1.02 [0.67, 1.57]
Xiang 2007	3	20	8	20	7.0%	0.38 [0.12, 1.21]
Total (95% CI)		**201**		**205**	**100.0%**	**0.92 [0.67, 1.25]**
Total events	55		62			

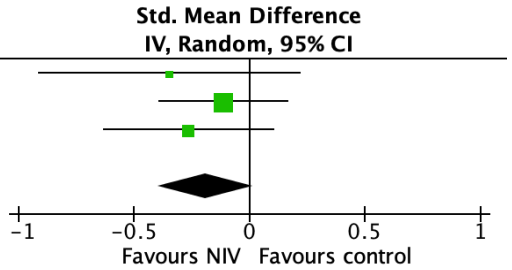
Heterogeneity: Tau^2 = 0.00; Chi^2 = 2.70, df = 3 (P = 0.44); I^2 = 0%
Test for overall effect: Z = 0.55 (P = 0.59)



Forest plot 2: Exacerbations

Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI	Year
Cheung 2010	0.521	0.5	23	0.69	0.46	26	13.2%	-0.35 [-0.91, 0.22]	2010
Struik 2014	1	6.7	101	2	10.4	100	55.2%	-0.11 [-0.39, 0.16]	2014
Murphy 2017	3.8	3.2	57	5.1	6.1	59	31.6%	-0.26 [-0.63, 0.10]	2017
Total (95% CI)			**181**			**185**	**100.0%**	**-0.19 [-0.40, 0.01]**	

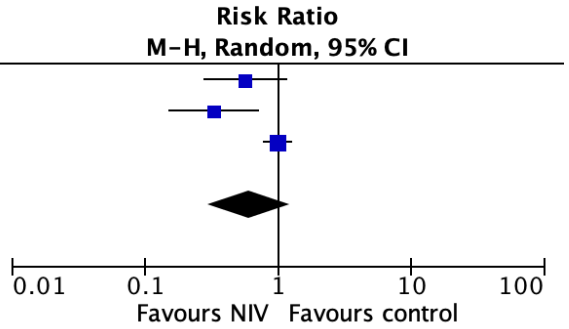
Heterogeneity: Tau^2 = 0.00; Chi^2 = 0.74, df = 2 (P = 0.69); I^2 = 0%
Test for overall effect: Z = 1.83 (P = 0.07)



Forest plot 3: Hospitalizations

Study or Subgroup	Experimental Events	Experimental Total	Control Events	Control Total	Weight	Risk Ratio M-H, Random, 95% CI
Cheung 2010	7	23	14	26	30.1%	0.57 [0.28, 1.15]
Murphy 2017	7	57	22	59	28.7%	0.33 [0.15, 0.71]
Struik 2014	57	101	57	100	41.1%	0.99 [0.78, 1.26]
Total (95% CI)		**181**		**185**	**100.0%**	**0.61 [0.30, 1.24]**
Total events	71		93			

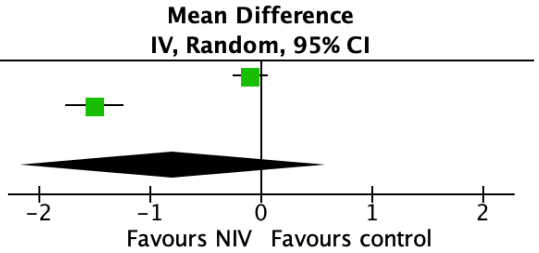
Heterogeneity: Tau^2 = 0.31; Chi^2 = 9.48, df = 2 (P = 0.009); I^2 = 79%
Test for overall effect: Z = 1.36 (P = 0.17)



Forest plot 4: Dyspnea

Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
Struik 2014	-0.4	0.35	49	-0.3	0.41	51	50.3%	-0.10 [-0.25, 0.05]
Xiang 2007	2.4	0.5	20	3.9	0.3	20	49.7%	-1.50 [-1.76, -1.24]
Total (95% CI)			**69**			**71**	**100.0%**	**-0.80 [-2.17, 0.58]**

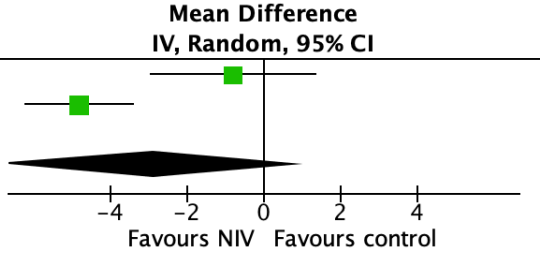
Heterogeneity: $Tau^2 = 0.97$; $Chi^2 = 85.98$, df = 1 ($P < 0.00001$); $I^2 = 99\%$
Test for overall effect: Z = 1.14 (P = 0.26)



Forest plot 6: Quality of life

Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
Murphy 2017	-1.5	3.5	35	-0.7	4.7	26	47.9%	-0.80 [-2.95, 1.35]
Struik 2014	-7	3.7	50	-2.2	3.5	51	52.1%	-4.80 [-6.21, -3.39]
Total (95% CI)			**85**			**77**	**100.0%**	**-2.89 [-6.80, 1.03]**

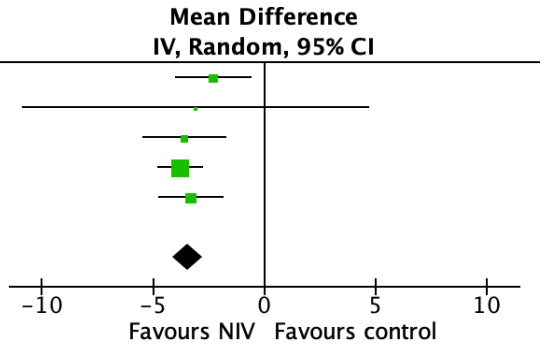
Heterogeneity: $Tau^2 = 7.14$; $Chi^2 = 9.34$, df = 1 ($P = 0.002$); $I^2 = 89\%$
Test for overall effect: Z = 1.44 (P = 0.15)



Forest plot 6: PCO2

Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
Cheung 2010	45.4	3	23	47.7	3	26	16.2%	-2.30 [-3.98, -0.62]
De Baker 2011	44.5	4.7	10	47.6	8.2	5	0.8%	-3.10 [-10.86, 4.66]
Murphy 2017	-5.9	3.1	31	-2.3	3.9	27	13.7%	-3.60 [-5.43, -1.77]
Struik 2014	-9.8	2.7	50	-6	2.3	48	46.7%	-3.80 [-4.79, -2.81]
Xiang 2007	49.5	2.2	20	52.8	2.4	20	22.6%	-3.30 [-4.73, -1.87]
Total (95% CI)			**134**			**126**	**100.0%**	**-3.41 [-4.09, -2.73]**

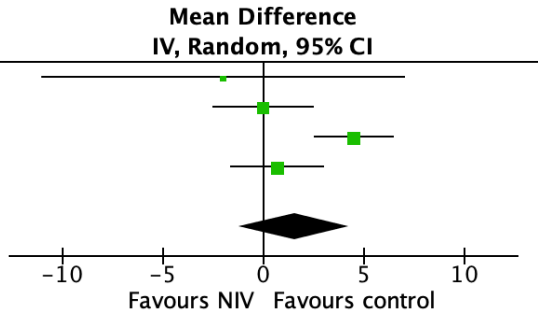
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 2.34$, df = 4 ($P = 0.67$); $I^2 = 0\%$
Test for overall effect: Z = 9.86 ($P < 0.00001$)



Forest plot 7: PO2

Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
De Baker 2011	-68	7	10	-66	9	5	7.3%	-2.00 [-11.00, 7.00]
Murphy 2017	7	4.1	32	7	5.3	26	29.5%	0.00 [-2.48, 2.48]
Struik 2014	0	4.6	45	-4.5	5	48	32.6%	4.50 [2.55, 6.45]
Xiang 2007	-60.8	4.7	20	-61.5	2.3	20	30.6%	0.70 [-1.59, 2.99]
Total (95% CI)			**107**			**99**	**100.0%**	**1.53 [-1.17, 4.24]**

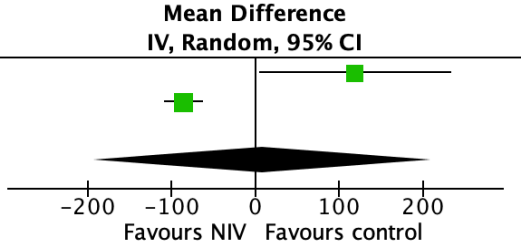
Heterogeneity: $Tau^2 = 4.85$; $Chi^2 = 10.78$, df = 3 (P = 0.01); $I^2 = 72\%$
Test for overall effect: Z = 1.11 (P = 0.27)



Forest plot 8: Exercise tolerance

Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
De Baker 2011	-282	146	10	-401	78	5	46.2%	119.00 [5.59, 232.41]
Xiang 2007	-213	45	20	-127	23	20	53.8%	-86.00 [-108.15, -63.85]
Total (95% CI)			**30**			**25**	**100.0%**	**8.64 [-191.66, 208.95]**

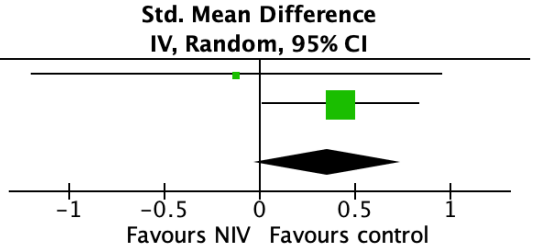
Heterogeneity: $Tau^2 = 19274.45$; $Chi^2 = 12.09$, df = 1 (P = 0.0005); $I^2 = 92\%$
Test for overall effect: Z = 0.08 (P = 0.93)



Forest plot 9: Pulmonary function- FEV1

Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
De Baker 2011	-38.5	14.6	10	-36.8	8.7	5	12.7%	-0.12 [-1.20, 0.95]
Struik 2014	-0.01	0.07	48	-0.04	0.07	46	87.3%	0.43 [0.02, 0.83]
Total (95% CI)			**58**			**51**	**100.0%**	**0.36 [-0.03, 0.74]**

Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 0.87$, df = 1 (P = 0.35); $I^2 = 0\%$
Test for overall effect: Z = 1.82 (P = 0.07)



GRADE Evidence Profile – Q3: NIV with targeted normalization of PaCO2 levels compared to NIV without targeting normal PaCO2 level for long-term NIV in COPD patients

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	NIV with targeted normalization of PaCO2 levels	NIV without targeting normal PaCO2 level	Relative (95% CI)	Absolute (95% CI)		
Quality of Life (follow up: mean 6 weeks; assessed with: SRI-SS score (higher is better))												
3	randomised trials	serious [a]	not serious	not serious	serious [b,c]	none	34	36	-	MD **0.95 points lower** (8.33 lower to 6.42 higher)	⨁⨁◯◯ LOW	CRITICAL
Dyspnea (follow up: mean 6 weeks; assessed with: Borg score)												
1	randomised trials	serious [a]	not serious	not serious	serious [c]	none	15	15	-	MD **1.54 points higher** (0.56 higher to 2.52 higher)	⨁⨁◯◯ LOW	CRITICAL
FEV1 (follow up: mean 6 weeks; assessed with: L)												
1	randomised trials	serious [a]	not serious	not serious	serious [b,c]	none	11	14	-	MD **0.04 L higher** (0.34 lower to 0.42 higher)	⨁⨁◯◯ LOW	IMPORTANT
PaCO2 (follow up: mean 6 weeks; assessed with: mmHg)												
5	randomised trials	serious [a]	not serious	not serious	serious [c]	none	65	68	-	MD **4.93 mmHg lower** (7.43 lower to 2.42 lower)	⨁⨁◯◯ LOW	IMPORTANT
PaO2 (follow up: mean 6 weeks; assessed with: mmHg)												
3	randomised trials	serious [a]	not serious	not serious	serious [b,c]	none	39	39	-	MD **3.4 mmHg higher** (2.39 lower to 9.19 higher)	⨁⨁◯◯ LOW	IMPORTANT
6MWD (follow up: mean 6 weeks; assessed with: metres)												
1	randomised trials	serious [a]	not serious	not serious	serious [b,c]	none	15	15	-	MD **14 metres higher** (70.42 lower to 98.42 higher)	⨁⨁◯◯ LOW	IMPORTANT
Sleep Comfort (follow up: mean 6 weeks; assessed with: VAS scale)												
1	randomised trials	serious [a]	not serious	not serious	very serious [b,c]	none	8	7	-	MD **1 cm higher** (28.42 lower to 30.42 higher)	⨁◯◯◯ VERY LOW	IMPORTANT

CI: Confidence interval; **MD:** Mean difference

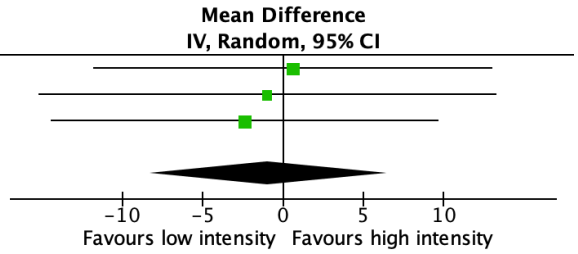
Explanations

a. Unblinded intervention may affect co-intervention use. Crossover study with potential for carryover effect.
b. Wide confidence intervals don't exclude significant harm or significant benefit.
c. Small number of patients limit precision.

Forest plot 1: Quality of Life

Study or Subgroup	High Intensity Mean	SD	Total	Low Intensity Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI	Year
Dreher 2010	63.22	16.9	15	62.6	17.7	15	35.4%	0.62 [-11.76, 13.00]	2010
Murphy 2012	66	16	8	67	12	7	26.9%	-1.00 [-15.21, 13.21]	2012
Duiverman 2017	51.3	15	11	53.7	15.5	14	37.6%	-2.40 [-14.42, 9.62]	2017
Total (95% CI)			**34**			**36**	**100.0%**	**-0.95 [-8.33, 6.42]**	

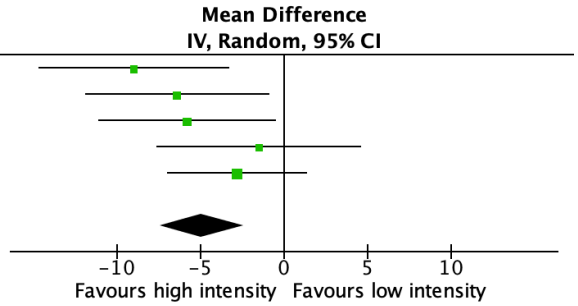
Heterogeneity: Tau^2 = 0.00; Chi^2 = 0.12, df = 2 (P = 0.94); I^2 = 0%
Test for overall effect: Z = 0.25 (P = 0.80)



Forest plot 2: PCO2

Study or Subgroup	High Intensity Mean	SD	Total	Low Intensity Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI	Year
Dreher 2010	47.8	7.2	15	56.8	8.6	15	17.3%	-9.00 [-14.68, -3.32]	2010
Dreher 2011	50.2	6.2	16	56.6	9.6	17	18.4%	-6.40 [-11.88, -0.92]	2011
Lukacsovits 2012	49.4	7.8	15	55.2	6.9	15	19.7%	-5.80 [-11.07, -0.53]	2012
Murphy 2012	52.5	6	8	54	6	7	15.3%	-1.50 [-7.59, 4.59]	2012
Duiverman 2017	45.3	5.33	11	48.1	5.18	14	29.4%	-2.80 [-6.96, 1.36]	2017
Total (95% CI)			**65**			**68**	**100.0%**	**-4.93 [-7.43, -2.42]**	

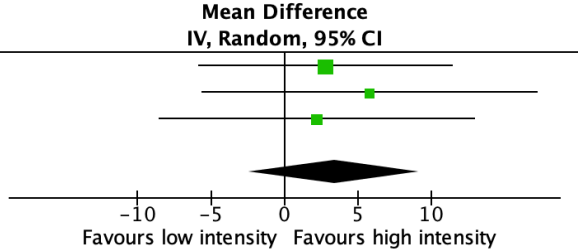
Heterogeneity: Tau^2 = 1.05; Chi^2 = 4.58, df = 4 (P = 0.33); I^2 = 13%
Test for overall effect: Z = 3.86 (P = 0.0001)



Forest plot 3: PO2

Study or Subgroup	High Intensity Mean	SD	Total	Low Intensity Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI	Year
Dreher 2011	71.7	10.8	16	68.9	14.3	17	45.1%	2.80 [-5.81, 11.41]	2011
Lukacsovits 2012	71.6	19.1	15	65.8	11.9	15	25.8%	5.80 [-5.59, 17.19]	2012
Murphy 2012	61.5	9.75	8	59.3	11.25	7	29.1%	2.20 [-8.53, 12.93]	2012
Total (95% CI)			**39**			**39**	**100.0%**	**3.40 [-2.39, 9.19]**	

Heterogeneity: Tau^2 = 0.00; Chi^2 = 0.24, df = 2 (P = 0.89); I^2 = 0%
Test for overall effect: Z = 1.15 (P = 0.25)



Question 4: Adaptive ventilatory modes in long-term NIV for COPD

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other consid-erations	Adaptive NIV	Conventional NIV	Relative (95% CI)	Absolute (95% CI)		
Quality of Life (follow up: range 2 months to 3 months; assessed with: Severe Respiratory Insufficiency Questionnaire (1 study); St. George's Respiratory Questionnaire (2 studies))												
3	RCTs	serious [a]	not serious	not serious	serious [b]	none	54	54	-	SMD **0.28 SD higher** (0.66 higher to 0.1 lower)	⨁⨁◯◯ LOW	CRITICAL
Sleep quality (follow up: range 1 days to 3 months; assessed with: Visual analogue scale (2 studies); unvalidated questionnaire (2 studies); Epworth Sleepiness Scale (1 study))												
5	RCTs	serious [a]	not serious	serious [c]	serious [b]	none	80	80	-	SMD **0.14 lower** (0.53 lower to 0.26 higher)	⨁◯◯◯ VERY LOW	IMPORTANT
Exercise tolerance (follow up: range 2 months to 3 months; assessed with: Shuttle walk test (2 studies); 6 minute walk test (1 study))												
3	RCTs	serious [a]	not serious	not serious	serious [b]	none	47	47	-	SMD **0.1 lower** (0.51 lower to 0.3 higher)	⨁⨁◯◯ LOW	IMPORTANT
PaCO2 (follow up: range 1 days to 3 months)												
6	RCTs	serious [a]	not serious	not serious	serious [b]	none	91	91	-	MD **1.95 mmHg lower** (4.29 lower to 0.4 higher)	⨁⨁◯◯ LOW	IMPORTANT
Oxygenation (follow up: range 1 days to 3 months; assessed with: PaO2, or SaO2 oximetry)												
6	RCTs	serious [a]	not serious	not serious	serious [b]	none	91	91	-	SMD **0.04 lower** (0.33 lower to 0.26 higher)	⨁⨁◯◯ LOW	IMPORTANT

CI: Confidence interval; **SMD:** Standardised mean difference; **MD:** Mean difference

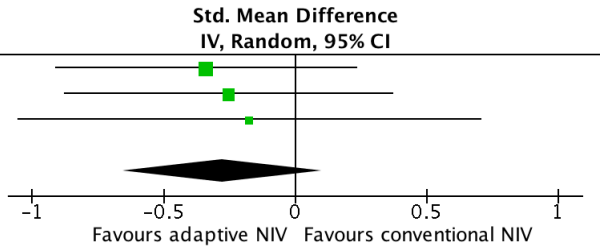
Explanations

a. Lack of blinding which could result in cointervention, or affect judgement of subjective outcomes.
b. Wide 95% confidence interval which fails to exclude significant benefit or harm.
c. Most studies did not use a validated instrument to assess sleep quality.

Forest plot 1: Quality of Life

Study or Subgroup	Hybrid NIV Mean	SD	Total	Pressure NIV Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI	Year
Oscroft 2010	55	13	24	59	10	24	44.2%	-0.34 [-0.91, 0.23]	2010
Oscroft 2014	59.7	20.8	20	64.4	15.1	20	37.1%	-0.25 [-0.88, 0.37]	2014
Storre 2014	-62.4	18.9	10	-59.3	14.8	10	18.6%	-0.17 [-1.05, 0.70]	2014
Total (95% CI)			**54**			**54**	**100.0%**	**-0.28 [-0.66, 0.10]**	

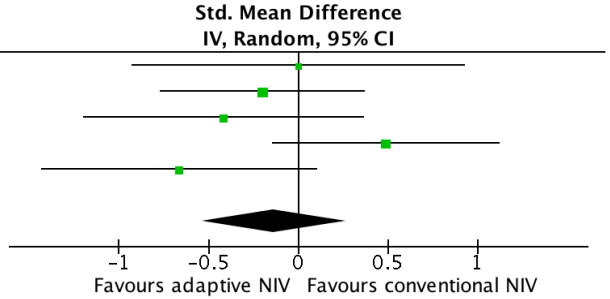
Heterogeneity: Tau² = 0.00; Chi² = 0.10, df = 2 (P = 0.95); I² = 0%
Test for overall effect: Z = 1.43 (P = 0.15)



Forest plot 2: Sleep quality

Study or Subgroup	Hybrid NIV Mean	SD	Total	Pressure NIV Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI	Year
Crisafulli 2009	5.1	2	9	5.1	1.7	9	14.0%	0.00 [-0.92, 0.92]	2009
Oscroft 2010	-78	22	24	-73	27	24	26.3%	-0.20 [-0.77, 0.37]	2010
Ekkenkamp 2014	-81	7.4	13	-77	10.8	13	17.9%	-0.42 [-1.20, 0.36]	2014
Oscroft 2014	-6.2	4.1	20	-8.9	6.5	20	23.4%	0.49 [-0.14, 1.12]	2014
Nilius 2017	7	3.6	14	9.3	3.1	14	18.4%	-0.66 [-1.43, 0.10]	2017
Total (95% CI)			**80**			**80**	**100.0%**	**-0.14 [-0.53, 0.26]**	

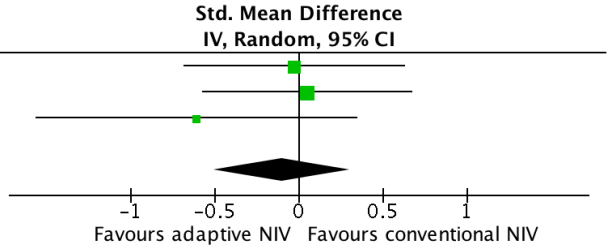
Heterogeneity: Tau² = 0.07; Chi² = 6.22, df = 4 (P = 0.18); I² = 36%
Test for overall effect: Z = 0.67 (P = 0.51)



Forest plot 3: Exercise tolerance

Study or Subgroup	Hybrid NIV Mean	SD	Total	Pressure NIV Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI	Year
Oscroft 2010	-173	97	18	-170	107	18	38.7%	-0.03 [-0.68, 0.62]	2010
Oscroft 2014	-129	90	20	-133	71	20	43.0%	0.05 [-0.57, 0.67]	2014
Storre 2014	-311	115	9	-238	113	9	18.3%	-0.61 [-1.56, 0.34]	2014
Total (95% CI)			**47**			**47**	**100.0%**	**-0.10 [-0.51, 0.30]**	

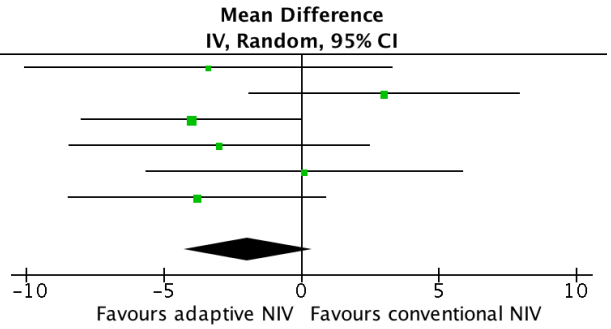
Heterogeneity: Tau² = 0.00; Chi² = 1.37, df = 2 (P = 0.50); I² = 0%
Test for overall effect: Z = 0.49 (P = 0.62)



Forest plot 4: PCO2

Study or Subgroup	Hybrid NIV Mean	SD	Total	Pressure NIV Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI	Year
Crisafulli 2009	59.8	7.7	9	63.2	6.7	9	10.7%	-3.40 [-10.07, 3.27]	2009
Oscroft 2010	50.3	8.3	24	47.3	9	24	17.7%	3.00 [-1.90, 7.90]	2010
Oscroft 2014	50.4	3.8	20	54.4	8.3	20	23.9%	-4.00 [-8.00, 0.00]	2014
Ekkenkamp 2014	40.6	7.6	14	43.6	7.1	14	15.0%	-3.00 [-8.45, 2.45]	2014
Storre 2014	47.2	3.9	10	47.1	8.4	10	13.7%	0.10 [-5.64, 5.84]	2014
Nilius 2017	49.1	6.4	14	52.9	6.2	14	19.1%	-3.80 [-8.47, 0.87]	2017
Total (95% CI)			**91**			**91**	**100.0%**	**-1.95 [-4.29, 0.40]**	

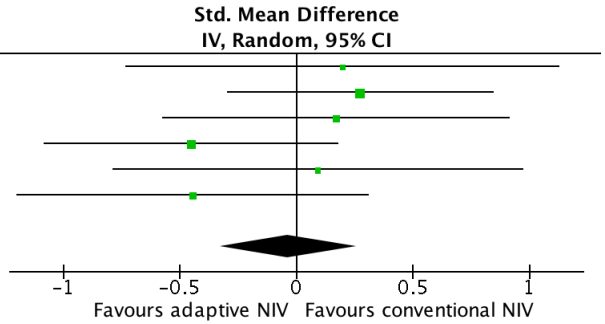
Heterogeneity: Tau² = 1.82; Chi² = 6.35, df = 5 (P = 0.27); I² = 21%
Test for overall effect: Z = 1.63 (P = 0.10)



Forest plot 5: PO2

Study or Subgroup	Hybrid NIV Mean	SD	Total	Pressure NIV Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI	Year
Crisafulli 2009	-70.2	9.2	9	-72.7	14.4	9	10.0%	0.20 [-0.73, 1.12]	2009
Oscroft 2010	-58.5	9	24	-60.8	7.5	24	26.5%	0.27 [-0.30, 0.84]	2010
Ekkenkamp 2014	-73.4	15.6	14	-76	14	14	15.5%	0.17 [-0.57, 0.91]	2014
Oscroft 2014	-65	12.8	20	-59.1	12.8	20	21.7%	-0.45 [-1.08, 0.18]	2014
Storre 2014	-77.6	18.2	10	-79.1	12.5	10	11.1%	0.09 [-0.79, 0.97]	2014
Nilius 2017	-92.6	3.5	14	-90.9	3.9	14	15.2%	-0.45 [-1.20, 0.31]	2017
Total (95% CI)			**91**			**91**	**100.0%**	**-0.04 [-0.33, 0.26]**	

Heterogeneity: Tau² = 0.00; Chi² = 4.58, df = 5 (P = 0.47); I² = 0%
Test for overall effect: Z = 0.25 (P = 0.81)



QUESTION

Should Long-term NIV vs. usual care be used for stable patients with COPD?	
POPULATION:	stable patients with COPD
INTERVENTION:	Long-term NIV
COMPARISON:	usual care
MAIN OUTCOMES:	Mortality; Number of Hospitalizations; Quality of Life (higher is better); Change in Dyspnea Score ; Change in PaCO2; Change in PaO2; Change in FEV1; Change in FVC; Change in 6 minute walk distance; Change in Sleep Efficiency; Minor Adverse Events;
SETTING:	
PERSPECTIVE:	
BACKGROUND:	
CONFLICT OF INTERESTS:	

ASSESSMENT

Problem

Is the problem a priority?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know		The panel decided on these PICO questions in advance of the guideline meeting on the basis of their importance to clinical practice.

Desirable Effects

How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Trivial ○ Small ○ Moderate ● Large ○ Varies ○ Don't know	**Desirable** Quality of life increased (0.49 SD higher) Reduced dyspnea score (0.51 SD lower). Really key-- this is a critical outcome for patients. Small improvement in 6 minute walk test (~30 m longer) **Undesirable** Slightly lower sleep efficency (0.55 SD lower)	Significant concerns raised about the heterogeneity of ventilatory settings-- do studies with higher CO2 clearance or settings demonstrate a greater effect? Further analysis requested on settings and whether or not there are other effects. Sensitivity analysis done looking at high-vs-low CO2 targeting.

	Increase in minor adverse events eg skin breakdown etc. (10 fold increase) **Neutral/little effect** Reduced mortality with NIV (14%) -not explicitly respiratory causes; and potentially high rate of dropouts, but imprecise with wide 95% CI. Reduced hopsitalizations (mean 1.26 fewer), but imprecise with wide 95% CI No significant effect upon FEV1 or FVC.	Studies targeting normal CO2 had stronger signal for benefit of NIV.

Undesirable Effects

How substantial are the undesirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know	**Desirable** Quality of life increased (0.49 SD higher) Reduced dyspnea score (0.51 SD lower). Really key-- this is a critical outcome for patients. Small improvement in 6 minute walk test (~30 m longer) **Undesirable** Slightly lower sleep efficency (0.55 SD lower) Increase in minor adverse events eg skin breakdown etc. (10 fold increase) **Neutral/little effect** Reduced mortality with NIV (14%) -not explicitly respiratory causes; and potentially high rate of dropouts, but imprecise with wide 95% CI. Reduced hopsitalizations (mean 1.26 fewer), but imprecise with wide 95% CI No significant effect upon FEV1 or FVC.	Dropout is also a concern here. Those that did not tolerate NIV fell out and therefore not exposed to benefit. (getting at lack of adherence as well) Meecham-Jones study data missing for sleep quality, though unlikely to change effects.

Certainty of evidence

What is the overall certainty of the evidence of effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Very low ● Low ○ Moderate ○ High ○ No included studies	Moderate certainty for dyspnea scores, changes in PaO2, PCO2. Low or very low certainty evidence for all other outcomes	

Values

Is there important uncertainty about or variability in how much people value the main outcomes?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Important uncertainty or variability ● Possibly important uncertainty or variability ○ Probably no important uncertainty or variability ○ No important uncertainty or variability	Exacerbations, dyspnea, and quality of life are among the most important outcomes in patients with COPD. Symptom relief was generally found to be more important than adverse events. PMID: 30002103	

Balance of effects

Does the balance between desirable and undesirable effects favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ● Probably favors the intervention ○ Favors the intervention ○ Varies ○ Don't know		The onyl negative effects see were sleep efficiency and minor adverse events; most other outcomes were positive (QOL, dyspnea, exercise tolerance) or neutral. Of note, mortality and hospitalizations the signal is towards benefit, which is reassuring.

Resources required

How large are the resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Large costs ● Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ○ Don't know	The estimated costs of providing a domiciliary NIV service are reported in Table 34 and for NIV were £2373 in the first year and £1536 in subsequent years. This estimate was in between cost estimates reported in the two studies identified in the clinical review. Tuggey et al. (2003)40 estimate domiciliary NIV to cost £1060 per year in 2003 prices, which converts to £1344 in 2012 prices (assuming a 3% inflation rate), and Clini et al. (2009)134 estimated NIV to cost €1920 in 2008 prices, which converts to £2727 (converting to GBP at the mid-year conversion rate of 1.263168 and inflating to 2012 prices at a rate of 3%).	

Certainty of evidence of required resources

What is the certainty of the evidence of resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Very low ● Low ○ Moderate ○ High ○ No included studies		Low certainty as the cost and avialability of resources for NIV may vary greatly across settings.

Cost effectiveness

Does the cost-effectiveness of the intervention favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ● Probably favors the intervention ○ Favors the intervention ○ Varies ○ No included studies	Economic modelling suggested that NIV may be cost-effective in a stable population at a threshold of £30,000 per quality-adjusted life-year (QALY) gained (incremental cost-effectiveness ratio £28,162), but this is associated with uncertainty. In the case of the post-hospital population, results for three separate base cases ranged from usual care dominating to NIV being cost-effective, with an incremental cost-effectiveness ratio of less than £10,000 per QALY gained. All estimates were sensitive suggested that reductions in the rate of hospital admissions per patient per year of 24% and 15% in the stable and post-hospital populations, respectively, are required for NIV to be cost-effective.	The group vascillated between "Favours the intervention" vs. "Probably favours the intervention." All agreed towards the side of cost effectiveness of NIV vs. comparison.

Equity

What would be the impact on health equity?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Reduced ○ Probably reduced ○ Probably no impact ○ Probably increased ○ Increased ○ Varies ○ Don't know		COPD patients disproportionately come from disadvantaged populations, so treating COPD may improve equity. In low or middle income countries/populations, home NIV may not be feasible and a recommendation for NIV may exacerbate health equity vs. more financially adventageous regions.

Acceptability
Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know		Many clinicians would find NIV acceptable due to its use in acute exacerbations-- it is a familiar therapy to those who treat COPD. Patients may vary with regard to acceptability of NIV in the long-term settings, however if it improves dyspnea and quality of life it may be acceptable.

Feasibility
Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ○ Yes ● Varies ○ Don't know		Some regions may not have infrastructure to support this; however there is widespread use of NIV in other countries which can provide practice models to guide practice. This will vary depending on the health care system, resources, and patient location.

SUMMARY OF JUDGEMENTS

	JUDGEMENT						
PROBLEM	No	Probably no	Probably yes	**Yes**		Varies	Don't know
DESIRABLE EFFECTS	Trivial	Small	Moderate	**Large**		Varies	Don't know
UNDESIRABLE EFFECTS	Large	Moderate	**Small**	Trivial		Varies	Don't know
CERTAINTY OF EVIDENCE	Very low	**Low**	Moderate	High			No included studies
VALUES	Important uncertainty or variability	**Possibly important uncertainty or variability**	Probably no important uncertainty or variability	No important uncertainty or variability			
BALANCE OF EFFECTS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	**Probably favors the intervention**	Favors the intervention	Varies	Don't know

RESOURCES REQUIRED	Large costs	**Moderate costs**	Negligible costs and savings	Moderate savings	Large savings	Varies	Don't know
CERTAINTY OF EVIDENCE OF REQUIRED RESOURCES	Very low	**Low**	Moderate	High			No included studies
COST EFFECTIVENESS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	**Probably favors the intervention**	Favors the intervention	Varies	No included studies
EQUITY	Reduced	Probably reduced	Probably no impact	Probably increased	Increased	Varies	Don't know
ACCEPTABILITY	No	Probably no	Probably yes	**Yes**		Varies	Don't know
FEASIBILITY	No	Probably no	Probably yes	Yes		**Varies**	Don't know

TYPE OF RECOMMENDATION

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	●	○

CONCLUSIONS

Recommendation

The ERS TF suggests long-term NIV be used for patients with chronic stable hypercapnic COPD (conditional recommendation, low certainty evidence).

Justification

Most outcomes favour NIV, including patient-important outcomes of dyspnea, QOL, exercise tolerance, with reassuring signal for mortality and exacerbations (towards benefit of NIV), and few harms (minor reduction in sleep efficacy and minor adverse events). Factors such as cost, acceptability, feasibility probably in favour though could vary between patients and settings. Overall balance of effects favour NIV though certainty of evidence is low, hence the panel chose a conditional/weak recommendation for NIV only; this allows a tailored approach to patient and setting-specific conditions as well.

Subgroup considerations

We examined the subgroup of studies which targeted normal CO2 (generally newer studies). The signal if anything was for more benefit of NIV if lower CO2 targeted, again suggesting hypercapnic patients derive the most benefit from NIV.

Implementation considerations

The panel recognized that the acceptability, feasibility, and costs of NIV vary greatly. For some patients and clinicans, the potential benefits (dyspnea, QOL, exercise tolerance; possible reduction in hospitalizations, though imprecise evidence) may not be worth it. This is consistent with a conditional recommendation in GRADE.

Monitoring and evaluation

See research priorities, below.

Research priorities

1) Strategies for initiating NIV. It is obvious that ventilator setting and acclimatization to NIV are crucial for effectiveness, including better adherence. NIV may be initiated in the hospital or at home. In-hospital initiation can be easily performed in some centers; however, it is more expensive and complex.

2) The benefits of NIV in subgroups of patients with COPD. The variability of both adherence and treatment response may vary according to different clinical phenotypes. Indeed, it seems that the response is better in those patients with PaCO2 > 50mmHg and PaCO2 reduction to normal following NIV. A phenogrouping strategy of hypercapnic COPD subgroups is needed for better defined the populations to be prioritized in further studies.

3) The impact of comorbid conditions in this population e.g. the effect of obesity, OSA-overlap, cardiovascular diseases, and clinical frailty upon clinical outcome.

4) Assessment of other underestimated factors, such as lack of social support and patient-ventilator asynchrony, which may impact the effectiveness of long-term NIV.

5) Cost effectiveness studies reporting the health economic value of long-term NIV in chronic stable COPD.

QUESTION

Should Long-term NIV vs. usual care be used for patients with COPD after an acute hypercapnic respiratory failure episode?	
POPULATION:	patients with COPD after an acute hypercapnic respiratory failure episode
INTERVENTION:	Long-term NIV
COMPARISON:	usual care
MAIN OUTCOMES:	Mortality; Exacerbations; Hospitalizations; Dyspnea score; Quality of Life; PaO2; PaCO2; Exercise tolerance; FEV1;
SETTING:	
PERSPECTIVE:	
BACKGROUND:	
CONFLICT OF INTERESTS:	

ASSESSMENT

Problem

Is the problem a priority?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know		The panel decided on these PICO questions in advance of the guideline meeting on the basis of their importance to clinical practice.

Desirable Effects

How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Trivial ○ Small ● Moderate ○ Large ○ Varies ○ Don't know	**Desirable effect** Exacerbations-- small reduction, but imprecise (SMD -0.19 exacerbations, -0.4 to 0.01) Hospitalizations-- small reduction, but imprecise (RR 0.61, 95%CI 0.30 to 1.24) Reduction in PCO2 (-3.41 mmHg) Dyspnea (MD -0.80)	Limited evidence based for this discussion. Small sample size and number of trials limit precision. May be a reduction in exacerbations and hospitalizations but limited by imprecision; other patient important outcomes likely also favour NIV but again issues with imprecision.

	Quality of life (MD -2.89 measured using SRI) **Undesirable effect** PO2 MD 1.53 mmHg **Little to no effect** Mortality RR 0.92 Exercise tolerance MD 8.64	Inclusion of Chung study which is considered to be at high risk of bias may limit interpretation of some outcome (eg. exacerbations). Sensitivity analysis including and excluding Chung has minimal impact upon point estimates; including increases precision slightly. TIming of initiation is also an important consideration-- HOT-HMV demonstrated reduction in exacerbations in select population of patients who remain hypercapenic ~2-4 weeks after their exacerbation; this is the subgroup most likely to benefit and possibly why smaller benefits seen with Struik. Sensitivity analysis excluding RESCUE trial demonstrates statistical significance for reduction in events; unfortunately no subgroup of persistent hypercapnia from Struik et all to compare.

Undesirable Effects

How substantial are the undesirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Large ○ Moderate ● Small ○ Trivial ○ Varies ○ Don't know	**Desirable effect** Exacerbations-- small reduction, but imprecise (SMD -0.19 exacerbations, -0.4 to 0.01) Hospitalizations-- small reduction, but imprecise (RR 0.61, 95%CI 0.30 to 1.24) Reduction in PCO2 (-3.41 mmHg) Dyspnea (MD -0.80) Quality of life (MD -2.89 measured using SRI) **Undesirable effect** PO2 MD 1.53 mmHg **Little to no effect** Mortality RR 0.92 Exercise tolerance MD 8.64	Even though not reflected in this evidence base, adverse events from Q1 also apply here as indirect evidence as presumably the mask and interfaces have the same effects whether used early after exacerbation or later in stable phase. Its possible that newer techniques and newer interfaces have less side effects.

Certainty of evidence

What is the overall certainty of the evidence of effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Very low ● Low	Low for almost all outcomes due to imprecision.	

○ Moderate ○ High ○ No included studies		

Values

Is there important uncertainty about or variability in how much people value the main outcomes?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Important uncertainty or variability ● Possibly important uncertainty or variability ○ Probably no important uncertainty or variability ○ No important uncertainty or variability	Exacerbations, dyspnea, and quality of life are among the most important outcomes in patients with COPD. Symptom relief was generally found to be more important than adverse events. PMID: 30002103	

Balance of effects

Does the balance between desirable and undesirable effects favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ● Probably favors the intervention ○ Favors the intervention ○ Varies ○ Don't know		Probably favors the intervention, especially in subgroup of patients with persistent hypercapnia 2-4 weeks post exacerbation as seen in HOT-HMV trial. Reduction of events is of great importance to patients. Possibly improvements in QOL and dyspnea as well. All outcomes limited by imprecision. ALso, little evidence of harm from NIV-- suspect that still occurs-- will assume similar rates of minor adverse reactions as in PICO 1, as mask fit etc. likely similar in post-exacerbation as chronic stable population.

Resources required

How large are the resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Large costs ● Moderate costs ○ Negligible costs and savings	The estimated costs of providing a domiciliary NIV service are reported in Table 34 and for NIV were £2373 in the first year and £1536 in subsequent years. This estimate was in between cost estimates reported in the two	

○ Moderate savings ○ Large savings ○ Varies ○ Don't know	studies identified in the clinical review. Tuggey et al. (2003)40 estimate domiciliary NIV to cost £1060 per year in 2003 prices, which converts to £1344 in 2012 prices (assuming a 3% inflation rate), and Clini et al. (2009)134 estimated NIV to cost €1920 in 2008 prices, which converts to £2727 (converting to GBP at the mid-year conversion rate of 1.263168 and inflating to 2012 prices at a rate of 3%).	

Certainty of evidence of required resources

What is the certainty of the evidence of resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Very low ● Low ○ Moderate ○ High ○ No included studies		Low certainty as the cost and avialability of resources for NIV may vary greatly across settings .

Cost effectiveness

Does the cost-effectiveness of the intervention favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ● Probably favors the intervention ○ Favors the intervention ○ Varies ○ No included studies	Economic modelling suggested that NIV may be cost-effective in a stable population at a threshold of £30,000 per quality-adjusted life-year (QALY) gained (incremental cost-effectiveness ratio £28,162), but this is associated with uncertainty. In the case of the post-hospital population, results for three separate base cases ranged from usual care dominating to NIV being cost-effective, with an incremental cost-effectiveness ratio of less than £10,000 per QALY gained. All estimates were sensitive to effectiveness estimates, length of benefit from NIV (currently unknown) and some costs. Modelling suggested that reductions in the rate of hospital admissions per patient per year of 24% and 15% in the stable and post-hospital populations, respectively, are required for NIV to be	Probably favours NIV in hypercapnic, "frequent flyer" population with recurrent hospitalizations and exacerbations.

	cost-effective	

Equity
What would be the impact on health equity?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Reduced ○ Probably reduced ○ Probably no impact ○ Probably increased ○ Increased ● Varies ○ Don't know		COPD patients disproportionately come from disadvantaged populations, so treating COPD may improve equity. In low or middle income countries/populations, home NIV may not be feasible and a recommendation for NIV may exacerbate health equity vs. more financially adventageous regions.

Acceptability
Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know		Many clinicians would find NIV acceptable due to its use in acute exacerbations-- it is a familiar therapy to those who treat COPD. This may be more acceptable in this post-acute exacerbation setting than the chronic COPD setting as the transition from acute to long-term NIV. Patients may vary with regard to acceptability of NIV in the long-term settings, however if it improves dyspnea and quality of life and exacerbations it may be acceptable.

Feasibility
Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know		May be more feasible than Q1 as clinical pathways exist pre- and post-discharge to facilitate initiation of NIV (eg. inpatient respirology consultation, arrange for equipment while still in hospital, etc).

SUMMARY OF JUDGEMENTS

	JUDGEMENT						
PROBLEM	No	Probably no	Probably yes	**Yes**		Varies	Don't know

DESIRABLE EFFECTS	Trivial	Small	**Moderate**	Large		Varies	Don't know
UNDESIRABLE EFFECTS	Large	Moderate	**Small**	Trivial		Varies	Don't know
CERTAINTY OF EVIDENCE	Very low	**Low**	Moderate	High			No included studies
VALUES	Important uncertainty or variability	**Possibly important uncertainty or variability**	Probably no important uncertainty or variability	No important uncertainty or variability			
BALANCE OF EFFECTS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	**Probably favors the intervention**	Favors the intervention	Varies	Don't know
RESOURCES REQUIRED	Large costs	**Moderate costs**	Negligible costs and savings	Moderate savings	Large savings	Varies	Don't know
CERTAINTY OF EVIDENCE OF REQUIRED RESOURCES	Very low	**Low**	Moderate	High			No included studies
COST EFFECTIVENESS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	**Probably favors the intervention**	Favors the intervention	Varies	No included studies
EQUITY	Reduced	Probably reduced	Probably no impact	Probably increased	Increased	**Varies**	Don't know
ACCEPTABILITY	No	Probably no	**Probably yes**	Yes		Varies	Don't know
FEASIBILITY	No	Probably no	**Probably yes**	Yes		Varies	Don't know

TYPE OF RECOMMENDATION

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	●	○

CONCLUSIONS

Recommendation

The ERS TF suggests long-term NIV be used in patients with COPD following a life-threatening episode of acute hypercapnic respiratory failure requiring acute NIV, if hypercapnia persists beyond 2-4 weeks following the episode (conditional recommendation, low certainty evidence).

Justification

Generally low certainty of evidence, but most important outcomes are neutral (mortality, exercise tolerance) or favour NIV (exacerbations, hospitalizations, HRQOL measured using SRI) without any major harms seen. Reduction in events seen in HOT-HMV trial likely because that trial included patients with persistent hypercapnia some time (2-4 weeks) after event; unfortunately subgroup data from Struik/RESCUE study does not have equivalent subgroup data; NIV appears to result in statistically significant reduction in exacerbations in the persistent hypercapnic subgroup. Subgroup analysis excluding Cheung et all study (thought to be high risk of bias) does not significantly affect estimates. Overall less certainty of effects but desirable effects likely outweigh undesirable effects; given low certainty of evidence only conditional/week recommendation could be made, and this after considering patient values & preferences; acceptability, feasibility and cost of NIV in local setting. Future evidence could change this recommendation in the future.

Subgroup considerations

Major issue is that subgroup of persistently hypercapnic patients from HOT-HMV appear to be most likely to benefit; this subgroup has statistically signficant reduction in hospitalizations compared to Struik/RESCUE data.

Implementation considerations

Patients who are started early may not remain hypercapnic; suggest that the group most likely to benefit from NIV is the group who remains hypercapnic 2-4 weeks after the episode, as seen in HOT-HMV. The panel recognized that the acceptability, feasibility, and costs of NIV vary greatly. For some patients and clinicans, the potential benefits (dyspnea, QOL, exercise tolerance; possible reduction in hospitalizations, though imprecise evidence) may not be worth it. This is consistent with a conditional recommendation in GRADE.

Monitoring and evaluation

Research priorities

1. Developing more accurate criteria for identifying patients who are likely to benefit from long-term NIV, such as severity of illness (hypothesis that treatment of higher PaCO2 at initiation will drive greater clinical benefits), trajectory of hypercapnia recovery after exacerbation (as some patients return to eucapnia more rapidly than others) and treatment response (e.g. early reduction in PaCO2 level after starting home NIV, with the hypothesis that greater reduction in PaCO2 will drive greater clinical benefit).

2.Physiological and biological mechanisms of action of long-term NIV: physiological mechanisms determining reduction in PaCO2; the biological effects of PaCO2 reduction in chronic hypercapnia upon immune, pulmonary vasculature, and skeletal muscle; biological mechanisms determining reduction in exacerbation; and physiological mechanisms determining enhanced sleep quality.

3. The effects of NIV upon mental health and cognition upon patients, including effects upon HRQL post ARF, cognitive function post-AHRF, the relationship between HRQL and cognitive function upon adherence and acceptability of home NIV.

4. Health service delivery research to promote the delivery of post-acute NIV to the right patient at the right time and prevent the ‘overuse’ or ‘underuse’ of the treatment.

5. Assessment of novel home treatments, e.g. high flow humidified nasal oxygen, that are capable of reducing PaCO2 in stable hypercapnic COPD patients.

QUESTION

Should NIV with targeted normalization of PaCO2 levels vs. NIV without targeting normal PaCO2 level be used for long-term NIV in COPD patients?	
POPULATION:	long-term NIV in COPD patients
INTERVENTION:	NIV with targeted normalization of PaCO2 levels
COMPARISON:	NIV without targeting normal PaCO2 level
MAIN OUTCOMES:	Quality of Life; Dyspnea; FEV1; PaCO2; PaO2; 6MWD; Sleep Comfort;
SETTING:	
PERSPECTIVE:	
BACKGROUND:	
CONFLICT OF INTERESTS:	

ASSESSMENT

Problem

Is the problem a priority?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know		The panel chose the PICO questions based upon their apparent relevance to clinical practice.

Desirable Effects

How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Trivial ● Small ○ Moderate ○ Large ○ Varies ○ Don't know	**Desirable effects** Reduction in PaCO2, generally at rest (reduction 5 mmHg) **Undesirable effects** Slightly higher dyspnea scores in only 1 RCT	Consider sensitivity analysis excluding Lukacsovits 2012 as measurements for PaO2 and PaCO2 were measured during NIV. Murphy looked at high-intensity vs. high-pressure. Thus substantial questions therefore raised about the directness of the evidence.

	Minimal effect Quality of life measured with SRI Sleep quality Exercise tolerance	The panel examined the studies from PICO 1 and the subgroup of studies which targeted normalization of CO2 vs. studies which did not target normalization. The subgroup which targeted normal CO2 demonstrated more benefit for NIV (Ie effect sizes were larger in studies which targeted normal CO2). The panel judged overal there may be a small benefit to targeting normal CO2.

Undesirable Effects

How substantial are the undesirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Large ○ Moderate ○ Small ● Trivial ○ Varies ○ Don't know	**Desirable effects** Reduction in PaCO2, generally at rest (reduction 5 mmHg) **Undesirable effects** Slightly higher dyspnea scores in only 1 RCT **Minimal effect** Quality of life measured with SRI Sleep quality Exercise tolerance	Trival undesirable effects-- slightly higher dypnea scores in a single study.

Certainty of evidence

What is the overall certainty of the evidence of effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
● Very low ○ Low ○ Moderate ○ High ○ No included studies	Very low certainty evidence. Less certainty as well because studies were generally short-term, without measuring effects upon mortality, hospitalizations, exacerbations.	

Values

Is there important uncertainty about or variability in how much people value the main outcomes?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Important uncertainty or variability ● Possibly important uncertainty or variability ○ Probably no important uncertainty or variability ○ No important uncertainty or variability	Exacerbations, dyspnea, and quality of life are among the most important outcomes in patients with COPD. Symptom relief was generally found to be more important than adverse events. PMID: 30002103	

Balance of effects

Does the balance between desirable and undesirable effects favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ● Probably favors the intervention ○ Favors the intervention ○ Varies ○ Don't know		While very low certainty of evidence of a small benefit, the anticipated harms are trivial, meaning the evidence probably favours the intervention.

Resources required

How large are the resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Large costs ○ Moderate costs ● Negligible costs and savings ○ Moderate savings ○ Large savings ○ Varies ○ Don't know		Implimenting higher intensity CO2 reduction required more time in hospital in one study, though admittedly limited evidence. No formal analysis of cost in any studies.

Certainty of evidence of required resources

What is the certainty of the evidence of resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS

● Very low ○ Low ○ Moderate ○ High ○ No included studies		

Cost effectiveness

Does the cost-effectiveness of the intervention favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ○ Favors the intervention ○ Varies ● No included studies		Given lack of certainty of benefits, lack of evidence for costs, the task force could not provide a judgement of the cost-effectiveness of targeting normalization of CO2 levels. If targeting normal CO2 was technically challenging in a given setting, it may not be cost-effective.

Equity

What would be the impact on health equity?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Reduced ○ Probably reduced ● Probably no impact ○ Probably increased ○ Increased ○ Varies ○ Don't know		Probably no impact upon equity when implementing in a patient popualtion already receiving NIV.

Acceptability

Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know		This approach is probably acceptable to clinicians, who apprecaite having a clear "target" for CO2. It makes physiologic sense to clinicians as well. Patients probably have no issues with acceptability unless settings need to be very high to achieve normal CO2; in such cases high-intensity NIV may not be acceptable.

Feasibility

Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know		In situations where NIV is planned, more targeting a significant reduction of CO2 is feasible, though actual normalization of CO2 levels is unlikely to be achieved for many patients, based upon the results of the included studies, which did not demonstrate complete normalization of CO2.

SUMMARY OF JUDGEMENTS

	JUDGEMENT						
PROBLEM	No	Probably no	Probably yes	**Yes**		Varies	Don't know
DESIRABLE EFFECTS	Trivial	**Small**	Moderate	Large		Varies	Don't know
UNDESIRABLE EFFECTS	Large	Moderate	Small	**Trivial**		Varies	Don't know
CERTAINTY OF EVIDENCE	**Very low**	Low	Moderate	High			No included studies
VALUES	Important uncertainty or variability	**Possibly important uncertainty or variability**	Probably no important uncertainty or variability	No important uncertainty or variability			
BALANCE OF EFFECTS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	**Probably favors the intervention**	Favors the intervention	Varies	Don't know
RESOURCES REQUIRED	Large costs	Moderate costs	**Negligible costs and savings**	Moderate savings	Large savings	Varies	Don't know

CERTAINTY OF EVIDENCE OF REQUIRED RESOURCES	**Very low**	Low	Moderate	High			No included studies
COST EFFECTIVENESS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	Varies	**No included studies**
EQUITY	Reduced	Probably reduced	**Probably no impact**	Probably increased	Increased	Varies	Don't know
ACCEPTABILITY	No	Probably no	Probably yes	**Yes**		Varies	Don't know
FEASIBILITY	No	Probably no	Probably yes	**Yes**		Varies	Don't know

TYPE OF RECOMMENDATION

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	●	○

CONCLUSIONS

Recommendation

The ERS TF suggests titrating long-term NIV to normalize or reduce PaCO2 levels in patients with COPD (conditional recommendation, very low certainty evidence).

Justification

We make a conditional recommendation due to the minimal potential harms of targeted normalization of CO2 and it is recognized that this is unlikely to be achieved in many patients. While high-intensity NIV may or may not have benefits, this is the approach most commonly used in many centres, and thus this is probably the most acceptable approach for many clinicians. Setting NIV to target a reduction in PaCO2 may require more time spent in hospital and therefore possibly increase costs and decrease feasibility of NIV. Recognizing the lack of compelling evidence, the panel made a conditional recommendation for high-intensity NIV, but low-intensity approaches may also be acceptable and useful in many patients.

Subgroup considerations

Implementation considerations

It is likely reasonable for initially aim for a normal CO2 in most patients, given the lack of harms and the possible, though very small, benefits with such an approach. If achieving normal CO2 was very difficult or the settings very high and uncomfortable for the patient, tartgeting normal CO2 may not worth lots of effort to achieve.

Monitoring and evaluation

Research priorities

1) The impact of NIV ventilator strategy targeted to maximise PaCO2 reduction compared to conventional ventilator modes on long-term clinical outcomes (i.e. hyperinflation, exacerbations, cardiovascular complications, hospitalisations, survival, costs, patient's adherence).

QUESTION

Should Adaptive volume-targeted NIV vs. conventional NIV be used for long-term NIV in patients with COPD?	
POPULATION:	long-term NIV in patients with COPD
INTERVENTION:	Adaptive volume-targeted NIV
COMPARISON:	conventional NIV
MAIN OUTCOMES:	Quality of Life; Sleep quality; Exercise tolerance; PaCO2; Oxygenation;
SETTING:	
PERSPECTIVE:	
BACKGROUND:	
CONFLICT OF INTERESTS:	

ASSESSMENT

Problem

Is the problem a priority?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ○ Yes ○ Varies ● Don't know		Question about variable modes is clearly important, however most modes are not studied, All studies in the systematic reviewer compared volume-targeted ventilator modes. The number of studies looking at these modes may reflect the desire of industry to find evidence to support these modes, rather than patient or clinician needs. Safety is the priority question given that these modes could result in hypoventilaion or result in large leaks. The studies included in PICOs 1-3 generally used fixed-pressure modes, making the applicability of evidence for these questions uncertain for auto-titrating modes: there is no long-term evidence of the impact of these modes upon mortality, hospitalizations etc.

Desirable Effects

How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
● Trivial ○ Small ○ Moderate ○ Large ○ Varies ○ Don't know	**Desirable effects** Small improvement in short-term (~2-3 months) QoL PCO2 **Undesirable effects** **Little to no effect** Sleep quality Exercise tolerance **Unknown** Mortality Hospitalizations Exacerbations	

Undesirable Effects

How substantial are the undesirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Large ○ Moderate ○ Small ● Trivial ○ Varies ○ Don't know	**Desirable effects** Small improvement in short-term (~2-3 months) QoL PCO2 **Undesirable effects** **Little to no effect** Sleep quality Exercise tolerance **Unknown** Mortality Hospitalizations Exacerbations	Potential harms not included in short-term studies included in meta-analysis. Theoretical risk of large leaks and hypoventilation with auto-titrating modes which may not be evident in short-term studies. Risks of impacting mortality, hospitalizations, and exacerbations.

Certainty of evidence

What is the overall certainty of the evidence of effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Very low ● Low ○ Moderate ○ High ○ No included studies	Generally low certainty of pooled evidence.	Substantial uncertainty about the effectiveness across algorithms, machines, makes and models.

Values

Is there important uncertainty about or variability in how much people value the main outcomes?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Important uncertainty or variability ● Possibly important uncertainty or variability ○ Probably no important uncertainty or variability ○ No important uncertainty or variability	Exacerbations, dyspnea, and quality of life are among the most important outcomes in patients with COPD. Symptom relief was generally found to be more important than adverse events. PMID: 30002103	

Balance of effects

Does the balance between desirable and undesirable effects favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Favors the comparison ○ Probably favors the comparison ● Does not favor either the intervention or the comparison ○ Probably favors the intervention ○ Favors the intervention ○ Varies ○ Don't know		Auto-titrating modes may be more beneficial in those in the setting of exacerbation or in the time of titration when trying to decide on levels of support. Despite possible improvement in short-term QOL, there is no clear benefit to use overall, and ongoing safety concerns about leaks and hypoventilation in the long-term.

Resources required

How large are the resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Large costs ○ Moderate costs ○ Negligible costs and savings ○ Moderate savings ○ Large savings ● Varies ○ Don't know		It will cost more to upgrade from an existing machine. For a new machine, the price of a ventilator with these capabilities will vary, but could be anything from no extra cost to expensive depending on make, model, funding strategy, etc.

Certainty of evidence of required resources

What is the certainty of the evidence of resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Very low ○ Low ○ Moderate ○ High ● No included studies		No real evidence to address cost.

Cost effectiveness

Does the cost-effectiveness of the intervention favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ○ Favors the intervention ○ Varies ● No included studies		Given lack of evidence for benefits and harms, lack of certianty around costs it is difficult to assess cost-effectiveness.

Equity

What would be the impact on health equity?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Reduced ○ Probably reduced ○ Probably no impact		Cost is higher so implementing these modes may increase inequity, hightening disparities between those with financial resources and those without. May reduce inequity as patients who do not have

○ Probably increased ○ Increased ● Varies ○ Don't know		access to a sleep laboratory to titrate NIV may benefit from some titration of NIV settings, though this question has not been studied in the existing clincal trials.

Acceptability

Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Compliance between adaptive VTPCV and conventional NIV were not significantly different in 5 studies (Crisafulli 2009, Oscroft 2010, Oscroft 2014, Ekkernkamp 2014, Storre 2014) which used data directly from the NIV devices. Patient self-reported toleance was not significantly different in 2 studies (Oscroft 2010, Storre 2014). Self-reported comfort was not different in 2 studies (Crisafulli 2009, Oscroft 2010). One study (Nilius 2017) used a questionnaire to assess acceptability, again finding no differences.	Overal doesn't seem to be any differences in acceptability of these modes vs. conventional fixed pressure modes.

Feasibility

Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know		Probably feasible, if ventilators already have these modes included. If not, would be less feasible.

SUMMARY OF JUDGEMENTS

	JUDGEMENT						
PROBLEM	No	Probably no	Probably yes	Yes		Varies	**Don't know**
DESIRABLE EFFECTS	**Trivial**	Small	Moderate	Large		Varies	Don't know
UNDESIRABLE EFFECTS	Large	Moderate	Small	**Trivial**		Varies	Don't know
CERTAINTY OF EVIDENCE	Very low	**Low**	Moderate	High			No included studies
VALUES	Important uncertainty or variability	**Possibly important uncertainty or variability**	Probably no important uncertainty or variability	No important uncertainty or variability			
			Does not favor				

BALANCE OF EFFECTS	Favors the comparison	Probably favors the comparison	**either the intervention or the comparison**	Probably favors the intervention	Favors the intervention	Varies	Don't know
RESOURCES REQUIRED	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	**Varies**	Don't know
CERTAINTY OF EVIDENCE OF REQUIRED RESOURCES	Very low	Low	Moderate	High			**No included studies**
COST EFFECTIVENESS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	Varies	**No included studies**
EQUITY	Reduced	Probably reduced	Probably no impact	Probably increased	Increased	**Varies**	Don't know
ACCEPTABILITY	No	Probably no	Probably yes	**Yes**		Varies	Don't know
FEASIBILITY	No	Probably no	Probably yes	**Yes**		Varies	Don't know

TYPE OF RECOMMENDATION

Strong recommendation against the intervention	**Conditional recommendation against the intervention**	Conditional recommendation for either the intervention or the comparison	Conditional recommendation for the intervention	Strong recommendation for the intervention
○	●	○	○	○

CONCLUSIONS

Recommendation

The ERS TF suggests using fixed pressure support mode as first-choice ventilator mode in patients with COPD using long-term NIV (conditional recommendation, very low certainty evidence).

Justification

While there may be benefits based upon the included short-term studies, further research is needed to demonstrate safety, especially given potential safety concerns

about dynamic auto-changes of the ventilator resulting in leaks or hypoventilation. Given that virtually all studies in other questions (PICO 1, 2, 3) used fixed-pressure modes, and that the PICO recomendations are thus based upon studies using this mode, it is unclear if the existing evidence would also apply to ventilators using these auto-titrating modes. Lastly, changing to these auto-modes may require purchase or exchange of new ventilators, reducing feasibility. While these modes may theoretically be useful to titrate without monitoring, until the safety of such approach is demonstrated the risks of unmonitored titration are unclear and this is not recommended.

Overall, given the uncertainty of long-term benefits, difficulty applying the evidence and recommendations from PICOs 1-3, the panel made a conditional recommendation against using auto-titrating modes as the initial mode of ventilation. However, conditional recommendation also means there may be some circumstances where these modes are considered for use, though the panel did not identify any such situations.

Subgroup considerations

Implementation considerations

Monitoring and evaluation

Research priorities

1) The role of adaptive/auto-titrating modes to improve the long-term outcome of COPD, acute exacerbation vs chronic stable hypercapnic COPD and optimization of overnight ventilation, especially in specific subgroups in which ventilatory requirements may vary substantially overnight.

2) The assessment of auto-EPAP modes (in addition to adaptive/auto-titrating modes) in the sub-group of patients with COPD-OSA overlap syndrome

3) The clinical efficacy and cost effectiveness of of auto-titrating modes in the inpatient vs outpatient settings avoiding the need for hospitalization to initiate NIV, thereby increasing access to NIV.