Appendix E: **GRADE Profiles**

E.1 Information needs of people with Parkinson's disease and their families and carers

E.1.1 Impulse control behaviours

Quality of life impact of having ICD

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	ICD	No ICD	Mean difference / Odds ratio: (95% CI)	Quality
Effect of ICD on quality of life (PDQ39)									
Phu (2014)	Cohort	Not serious	N/A[1]	Not serious[2]	Not serious[3]	15	85	MD = 18 (2.24 to 33.76)	HIGH
Patient experience: major depressive disorder in ICD									
Phu (2014)	Cohort	Not serious	N/A[1]	Not serious[2]	Serious[4]	15	85	OR = 3.07 (0.80 to 11.69)	MODERATE

[1] N/A: not applicable as only one study contributed to this analysis
[2] No serious indirectness: population matches review protocol
[3] CI do not cross MID of 1.6 points (Peto et al., 2001)
[4] Serious imprecision: Non-significant results

Reluctance to start medication for Parkinson's disease

Quality assessment								
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Number of patients	Number of physicians	Quality
A mutual misunderstanding by patients and physicians								
Mestre 2014	Cross-sectional	Serious[1]	N/A[2]	Not serious[3]	Not serious	62/201	268	MODERATE

[1] Serious risk of bias: Methodology not clear, not clear whether all survey/questionnaire materials were standardised or validated as assessed by the reviewer (no well-validated methodology quality checklist available for cross-sectional studies)
[2] N/A: not applicable as only one study contributed to this analysis
[3] No serious indirectness: population matches review protocol

E.1.2 Women of childbearing age

Birth complications in women with PD

Quality assessment								
Studies	Design	Risk of bias	Inconsistency	Indirectness	imprecision	Successful pregnancies	Spontaneous miscarriages in the first 4 months of pregnancy	Quality
Number of spontaneous miscarriages in the first 4 months of pregnancy								
Golbe 1987	Qualitative	Very serious[a]	N/A[b]	Not serious[c]	Serious[d]	N= 17	N= 3/17 (15%)	VERY LOW
Number of total elective abortions								
Golbe 1987	Qualitative	Very serious[a]	N/A[b]	Not serious[c]	Serious[d]	N= 17	N= 4/17 (24%)	VERY LOW
Mean PD disease duration								
Golbe 1987	Qualitative	Very serious[a]	N/A[b]	Not serious[c]	Serious[d]	4.2 (4.5) years	3 (2.6) years	VERY LOW

[a] Very serious risk of bias as assessed by CASP qualitative quality checklist
[b] N/A: not applicable as only one study contributed to this analysis
[c] No serious indirectness: population matches review protocol
[d] Serious imprecision: Number of participants small

Pregnancy complications and related drug therapy in women with PD

Quality assessment								
Studies	Design	Risk of bias	Inconsistency	Indirectness	imprecision	Treatment	No treatment	Quality
Rate of complications associated with amantadine								
Golbe 1987	Case series	Very serious[a]	N/A[b]	Not serious[c]	Serious[d]	4/4 (100%) (2 miscarriage, 1 preeclampsia, 1 1st tri bleeding)	4/16 (25%) (vaginal bleeding or severe nausea)	VERY LOW
Rate of complications associated with levodopa/carbidopa								

Quality assessment								
Studies	Design	Risk of bias	Inconsistency	Indirectness	imprecision	Treatment	No treatment	Quality
Golbe 1987	Case series	Very serious[a]	N/A[b]	Not serious[c]	Serious[d]	4/6 (66%) (worsening of PD symptoms)	NA	VERY LOW

[a] Very serious risk of bias as assessed by CASP qualitative quality checklist
[b] N/A: not applicable as only one study contributed to this analysis
[c] No serious indirectness: population matches review protocol
[d] Serious imprecision: Number of participants small

Neurological complications of pregnancy in women with PD

Quality assessment Number of patients								
Example Studies	Design	Risk of bias	Inconsistency	Indirectness	imprecision	Total number of pregnancies	Events	Quality
Exacerbation of PD symptoms (worsening or development of new symptoms)								
Golbe 1987	Case series	Very serious[a]	N/A[b]	Not serious[c]	Serious[d]	17	11/17 (64.7%)	VERY LOW
Improvement of PD symptoms post-delivery (in population who experienced worsening during pregnancy)								
Golbe 1987	Case series	Very serious[a]	N/A[b]	Not serious[c]	Serious[d]	11	1/11 (9.09%)	VERY LOW
Development of serious post-partum depression requiring medication								
Golbe 1987	Case series	Very serious[a]	N/A[b]	Not serious[c]	Serious[d]	4	0/4 (0%)	VERY LOW

[a] Very serious risk of bias as assessed by CASP qualitative quality checklist
[b] N/A: not applicable as only one study contributed to this analysis
[c] No serious indirectness: population matches review protocol
[d] Serious imprecision: Number of participants small

Post-partum depression/anxiety

Quality assessment Number of patients								
Example Studies	Design	Risk of bias	Inconsistency	Indirectness	imprecision	Total number of pregnancies	Events	Quality
Development of serious post-partum depression requiring medication								
Golbe 1987	Case series	Very serious[a]	N/A[b]	Not serious[c]	Serious[d]	4	0/4 (0%)	VERY LOW

[a] Very serious risk of bias as assessed by CASP qualitative quality checklist
[b] N/A: not applicable as only one study contributed to this analysis
[c] No serious indirectness: population matches review protocol
[d] Serious imprecision: Number of participants small

E.2 Pharmacological management of motor symptoms

E.2.1 First-line treatment of motor symptoms

E.2.1.1 Treatment-naïve population

UPDRS Total – MAOB (Rasagiline, Selegiline) vs. placebo

Change in UPDRS Total from baseline to 36 weeks/12 months – MAOB vs. placebo

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
Change in UPDRS total score									
2 studies: Olanow et al., 2009; Palhågen et al., 1998	RCT	Serious[1]	Not serious	Serious[5]	Not serious[6]	613	612	-3.07 (-3.78, -2.37)	LOW

[1] Downgraded 1 level: Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A: Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness: population was as specified in review protocol
[4] Downgraded 1 level: Non-significant results
[5] Downgraded 1 level: Some patients from the placebo group had early transfer from phase 1 to phase 2, where active treatment was given
[6] CI do not cross MID of 7.3 points (Schrag et al., 2006)

Beck Depression Inventory - Pramipexole vs. placebo

BDI from baseline to 9 months – Pramipexole vs. placebo

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
Change in BDI score									
1 study: Schapira et al., 2013	RCT	Serious[1]	N/A[2]	Serious[5]	Not serious	211	200	-1.4 (-2.23, -0.57)	LOW

[1] Downgraded 1 level: Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A: Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness: population was as specified in review protocol
[4] Downgraded 1 level: Non-significant results

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
[5] Downgraded 1 level: Some patients from the placebo group had early transfer from phase 1 to phase 2, where active treatment was given									
[6] CI do not cross MID of 7.3 points (Schrag et al., 2006)									

Adverse events – Ropinirole vs. Pramipexole (dopamine agonists)

Any AE leading to trial discontinuation – Ropinirole vs. pramipexole

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	RR (95% CI)	
Adverse event									
1 study: Thomas et al., 2006	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[4]	30	30	1.67 (0.44, 6.36)	LOW
[1] Downgraded 1 level: Serious risk of bias as assessed by NICE RCT quality checklist [2] N/A: Not applicable, only 1 study contributed to this analysis [3] No serious indirectness: population was as specified in review protocol [4] Downgraded 1 level: Non-significant results [5] Downgraded 1 level: Some patients from the placebo group had early transfer from phase 1 to phase 2, where active treatment was given [6] CI do not cross MID of 7.3 points (Schrag et al., 2006)									

Adverse events - Rasagiline vs. placebo

Adverse event rate (any AE) – Rasagiline vs. placebo

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	IRR (95% CI)	
Adverse event rate									
1 study: Olanow et al., 2009	RCT	Serious[1]	N/A[2]	Serious[5]	Not serious	576	588	0.80 (0.65, 0.99)	LOW
[1] Downgraded 1 level: Serious risk of bias as assessed by NICE RCT quality checklist [2] N/A: Not applicable, only 1 study contributed to this analysis [3] No serious indirectness: population was as specified in review protocol [4] Downgraded 1 level: Non-significant results [5] Downgraded 1 level: Some patients from the placebo group had early transfer from phase 1 to phase 2, where active treatment was given [6] CI do not cross MID of 7.3 points (Schrag et al., 2006)									

Adverse event rate (AE related to dopaminergic therapy) – Rasagiline vs. placebo

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	IRR (95% CI)	
Adverse event rate									
1 study: Olanow et al., 2009	RCT	Serious[1]	N/A[2]	Serious[5]	Serious[4]	576	588	0.72 (0.49, 1.07)	VERY LOW

[1] Downgraded 1 level: Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A: Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness: population was as specified in review protocol
[4] Downgraded 1 level: Non-significant results
[5] Downgraded 1 level: Some patients from the placebo group had early transfer from phase 1 to phase 2, where active treatment was given
[6] CI do not cross MID of 7.3 points (Schrag et al., 2006)

Adverse events – Levodopa/carbidopa (150/37.5 mg/day and 300/75 mg/day) vs. placebo

Adverse event rate (any AE) – Levodopa/carbidopa (150/37.5 mg/d and 300/75 mg/day) vs. placebo

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	IRR (95% CI)	
Adverse event rate									
1 study: Fahn et al., 2005	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[4]	180	90	1.00 (0.84, 1.20)	LOW

[1] Downgraded 1 level: Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A: Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness: population was as specified in review protocol
[4] Downgraded 1 level: Non-significant results
[5] Downgraded 1 level: Some patients from the placebo group had early transfer from phase 1 to phase 2, where active treatment was given
[6] CI do not cross MID of 7.3 points (Schrag et al., 2006)

Adverse event rate (AE related to dopaminergic therapy) – Levodopa/carbidopa (150/37.5 mg/d and 300/75 mg/day) vs. placebo

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	IRR (95% CI)	
Adverse event rate									
1 study: Fahn et al., 2005	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[4]	180	90	0.85 (0.60, 1.21)	LOW

[1] Downgraded 1 level: Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A: Not applicable, only 1 study contributed to this analysis

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	IRR (95% CI)	

[3] No serious indirectness: population was as specified in review protocol
[4] Downgraded 1 level: Non-significant results
[5] Downgraded 1 level: Some patients from the placebo group had early transfer from phase 1 to phase 2, where active treatment was given
[6] CI do not cross MID of 7.3 points (Schrag et al., 2006)

Serious adverse event rate – Levodopa/carbidopa (150/37.5 mg/d and 300/75 mg/day) vs. placebo

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	IRR (95% CI)	
Adverse event rate									
1 study: Fahn et al., 2005	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[4]	180	90	1.50 (0.41, 5.54)	LOW

[1] Downgraded 1 level: Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A: Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness: population was as specified in review protocol
[4] Downgraded 1 level: Non-significant results
[5] Downgraded 1 level: Some patients from the placebo group had early transfer from phase 1 to phase 2, where active treatment was given
[6] CI do not cross MID of 7.3 points (Schrag et al., 2006)

Adverse events – Levodopa/cabidopa (600/150 mg/day) vs. placebo

Adverse event rate (any AE) – Levodopa/carbidopa (600/150 mg/day) vs. placebo

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	IRR (95% CI)	
Adverse event rate									
1 study: Fahn et al., 2005	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[4]	91	90	1.18 (0.97, 1.43)	LOW

[1] Downgraded 1 level: Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A: Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness: population was as specified in review protocol
[4] Downgraded 1 level: Non-significant results
[5] Downgraded 1 level: Some patients from the placebo group had early transfer from phase 1 to phase 2, where active treatment was given
[6] CI do not cross MID of 7.3 points (Schrag et al., 2006)

Adverse event rate (AE related to dopaminergic therapy) – Levodopa/carbidopa (600/150 mg/day) vs. placebo

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	IRR (95% CI)	
Adverse event rate									
1 study: Fahn et al., 2005	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[4]	91	90	1.23 (0.84, 1.78)	LOW

[1] Downgraded 1 level: Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A: Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness: population was as specified in review protocol
[4] Downgraded 1 level: Non-significant results
[5] Downgraded 1 level: Some patients from the placebo group had early transfer from phase 1 to phase 2, where active treatment was given
[6] CI do not cross MID of 7.3 points (Schrag et al., 2006)

Serious adverse event rate – Levodopa/carbidopa (600/150 mg/day) vs. placebo

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	IRR (95% CI)	
Adverse event rate									
1 study: Fahn et al., 2005	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[4]	91	90	0.66 (0.11, 3.95)	LOW

[1] Downgraded 1 level: Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A: Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness: population was as specified in review protocol
[4] Downgraded 1 level: Non-significant results
[5] Downgraded 1 level: Some patients from the placebo group had early transfer from phase 1 to phase 2, where active treatment was given
[6] CI do not cross MID of 7.3 points (Schrag et al., 2006)

Adverse events - Pramipexole vs. placebo

Any adverse event - Pramipexole vs. placebo

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	RR (95% CI)	
Adverse event									
1 study: Schapira et al., 2013	RCT	Not serious	N/A[2]	Serious[5]	Serious[4]	261	274	1.04 (0.94, 1.15)	LOW

[1] Downgraded 1 level: Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A: Not applicable, only 1 study contributed to this analysis

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	RR (95% CI)	

[3] No serious indirectness: population was as specified in review protocol
[4] Downgraded 1 level: Non-significant results
[5] Downgraded 1 level: Some patients from the placebo group had early transfer from phase 1 to phase 2, where active treatment was given
[6] CI do not cross MID of 7.3 points (Schrag et al., 2006)

Any serious adverse event - Pramipexole vs. placebo

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	RR (95% CI)	
Adverse event									
1 study: Schapira et al., 2013	RCT	Not serious	N/A[2]	Serious[5]	Serious[4]	261	274	0.99 (0.52, 1.88)	LOW

[1] Downgraded 1 level: Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A: Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness: population was as specified in review protocol
[4] Downgraded 1 level: Non-significant results
[5] Downgraded 1 level: Some patients from the placebo group had early transfer from phase 1 to phase 2, where active treatment was given
[6] CI do not cross MID of 7.3 points (Schrag et al., 2006)

Any serious adverse event leading to discontinuation - Pramipexole vs. placebo

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	RR (95% CI)	
Adverse event									
1 study: Schapira et al., 2013	RCT	Not serious	N/A[2]	Serious[5]	Serious[4]	261	274	1.01 (0.60, 1.70)	LOW

[1] Downgraded 1 level: Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A: Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness: population was as specified in review protocol
[4] Downgraded 1 level: Non-significant results
[5] Downgraded 1 level: Some patients from the placebo group had early transfer from phase 1 to phase 2, where active treatment was given
[6] CI do not cross MID of 7.3 points (Schrag et al., 2006)

Network meta-analyses

UPDRS Total

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in UPDRS Total score					
5 MAOB: Mally et al., 1995; Palhågen et al., 1998; Olanow et al., 2009. DA: Schapira et al., 2013. Levodopa: Fahn et al., 2005.	Serious[1]	Not serious	Not serious[3]	Not serious	MODERATE

[1]Downgrade 1 level: Limitations in the design or execution of the study
[2]No heterogeneity (i^2 =0%)
[3]Considered not serious as population, interventions, comparator and outcomes are as defined in protocol

UPDRS II (ADL)

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in UPDRS ADL score					
4 MAOB: Mally et al., 1995; Palhågen et al., 1998. DA: Schapira et al., 2013. Levodopa: Fahn et al., 2005.	Serious[1]	Serious[2]	Not serious[3]	Not serious	LOW

[1]Downgrade 1 level: Limitations in the design or execution of the study
[2]Considerable between study heterogeneity (i^2 >40%)
[3]Considered not serious as population, interventions, comparator and outcomes are as defined in protocol

UPDRS III (Motor)

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in UPDRS Motor score					
4	Serious[1]	Serious[2]	Not serious[3]	Not serious	LOW

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
MAOB: Mally et al., 1995; Palhågen et al., 1998. DA: Schapira et al., 2013. Levodopa: Fahn et al., 2005.					

[1]Downgrade 1 level: Limitations in the design or execution of the study
[2]Considerable between study heterogeneity (i^2 >40%)
[3]Considered not serious as population, interventions, comparator and outcomes are as defined in protocol

E.2.1.2 Full population

Low–dose levodopa versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
UPDRS (ADL)	1 (Fahn)	Serious[1]	N/A	Not serious	Not serious[5]	MD -1.60 (-2.64, -0.56)	Moderate
UPDRS (motor)	1 (Fahn)	Serious[1]	N/A	Not serious	Serious[3]	MD -2.90 (-4.94, -0.86)	Low
UPDRS (total)	1 (Fahn)	Serious[1]	N/A	Not serious	Serious[4]	MD -5.00 (-7.76, -2.24)	Low
Any AE	1 (Fahn)	Serious[1]	N/A	Not serious	Serious[2]	IRR 1.01 (0.84, 1.20)	Low
SAE	1 (Fahn)	Serious[1]	N/A	Not serious	Serious[2]	IRR 1.50 (0.41, 5.54)	Low
Dopaminergic AE	1 (Fahn)	Serious[1]	N/A	Not serious	Serious[2]	IRR 0.85 (0.60, 1.21)	Low

[1]Study at high risk of bias; [2]Non-significant result; [3]CI cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006); [4]CI cross MID of 7.3 points (Schrag et al., 2006); [5]CI do not cross MID of 3 points (Schrag et al., 2006)

High-dose levodopa versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
UPDRS (ADL)	1 (Fahn)	Serious[1]	N/A	Not serious	Serious[3]	MD -2.20 (-3.41, -0.99)	Low
UPDRS (motor)	1 (Fahn)	Serious[1]	N/A	Not serious	Serious[4]	MD -5.40 (-7.85, -2.95)	Low
UPDRS (total)	1 (Fahn)	Serious[1]	N/A	Not serious	Serious[5]	MD -8.00 (-11.25, -4.75)	Low
Any AE	1 (Fahn)	Serious[1]	N/A	Not serious	Serious[2]	IRR 1.18	Low

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
						(0.97, 1.43)	
SAE	1 (Fahn)	Serious[1]	N/A	Not serious	Serious[2]	IRR 0.66 (0.11, 3.95)	Low
Dopaminergic AE	1 (Fahn)	Serious[1]	N/A	Not serious	Serious[2]	IRR 1.23 (0.85, 1.78)	Low

[1]Study at high risk of bias; [2]Non-significant result; [3]CI cross MID of 3 points (Schrag et al., 2006); [4]CI cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006); [5]CI cross MID of 7.3 points (Schrag et al., 2006)

Extended-release levodopa versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
UPDRS (ADL)	1 (Pahwa)	Not serious	N/A	Serious[1]	Not serious[4]	MD -9.23 (-11.61, -6.85)	Moderate
UPDRS (motor)	1 (Pahwa)	Not serious	N/A	Serious[1]	Not serious[5]	MD -9.23 (-11.61, -6.85)	Moderate
PDQ-39	1 (Pahwa)	Not serious	N/A	Serious[1]	Not serious[6]	MD -5.31 (-8.90, -1.73)	Moderate
Any AE	1 (Pahwa)	Serious[2]	N/A	Serious[1]	Serious[3]	RR 0.92 (0.79, 1.06)	Very low
AE discontinuation	1 (Pahwa)	Serious[2]	N/A	Serious[1]	Serious[3]	RR 2.74 (1.00, 7.52)	Very low

[1]Population not treatment-naïve; [2]Selection of adverse events to report unclear; [3]Non-significant result; [4]CI do not cross MID of 3 points (Schrag et al., 2006); [5]CI do not cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006); [6]CI do not cross MID of 1.6 points (Peto et al., 2001)

Dopamine agonists versus placebo

Short-term follow-up (≤6 months)

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
UPDRS (ADL)	6 (Hauser, Hubble, Jankovic, Mizuno, PSG 2003, Zhang)	Not serious[1]	Not serious	Serious[2]	Not serious[6]	MD -1.22 (-1.62, -0.81)	Moderate
UPDRS (motor)	6 (Hauser, Hubble, Jankovic, Mizuno, PSG 2003, Zhang)	Not serious[1]	Not serious	Serious[2]	Serious[5]	MD -3.20 (-4.08, -2.31)	Low
UPDRS (total)	2 (Adler, PSG 1997)	Not serious[1]	Not serious	Serious[2]	Not serious[7]	MD -4.85 (-6.65, -3.06)	Moderate
Epworth sleep scale	2 (Hauser, Jankovic)	Not serious	Serious[3]	Serious[2]	Not serious	MD 1.40 (0.59, 2.22)	Low
PDQ-39	1 (Hauser)	Not serious	N/A	Serious[2]	Not serious[8]	MD -6.81 (-11.42, -2.20)	Moderate
EQ-VAS	1 (Hauser)	Not serious	N/A	Serious[2]	Serious[4]	MD 4.86 (-1.11, 10.84)	Low

[1]Individual studies at risk of bias, but overall risk of bias rated low due to consistency of effect between studies at high and low risk of bias; [2]Population not treatment-naïve; [3]Considerable between study heterogeneity (i^2 >40%); [4]Non-significant result; [5]CI cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006); [6]CI do not cross MID of 3 points (Schrag et al., 2006); [7]CI do not cross MID of 7.3 points (Schrag et al., 2006); [8]CI do not cross MID of 1.6 points (Peto et al., 2001)

Medium term follow-up (6 months – 2.5 years)

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
UPDRS (ADL)	2 (Poewe, Schapira)	Not serious	Serious[1]	Serious[2]	Not serious[6]	MD -1.54 (-2.47, -0.62)	Low
UPDRS (motor)	2 (Poewe, Schapira)	Not serious	Serious[1]	Serious[2]	Serious[4]	MD -4.19 (-6.00, --2.38)	Very low
UPDRS (total)	1 (Schapira)	Not serious	N/A	Not serious	Not serious[7]	MD -4.80 (-6.46, -3.14)	High
BDI	1 (Schapira)	Not serious	N/A	Not serious	Not serious	MD -1.40 (-2.23, -0.57)	High
PDQ-39	1 (Poewe)	Not serious	N/A	Serious[2]	Serious[5]	MD -3.63 (-7.01, -0.25)	Low
EQ-VAS	1 (Poewe)	Not serious	N/A	Serious[2]	Serious[3]	MD 2.94 (-1.46, 7.34)	Low

[1]Considerable between study heterogeneity (i^2 >40%); [2]Population not treatment-naïve; [3]Non-significant result; [4]CI cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006); [5]CI cross MID of 1.6 points (Peto et al., 2001); [6]CI do not cross MID of 3 points (Schrag et al., 2006); [7]CI do not cross MID of 7.3 points (Schrag et al., 2006)

Adverse events

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Any AE (Pramipexole)	5 (Hauser, Hubble, Poewe, PSG 1997, Schapira)	Serious[1]	Not serious	Serious[2]	Serious[3]	RR 1.05 (1.00,1.14)	Very low
Any AE (Rotigotine)	5 (Giladi, Jankovic, PSG 2003, Watts, Zhang)	Serious[1]	Serious[4]	Serious[2]	Not serious	IRR 1.44 (1.09, 1.90)	Very low

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Any AE (Ropinirole)	1 (Adler)	Serious[1]	N/A	Serious[2]	Serious[3]	RR 1.06 (0.99, 1.13)	Very low
SAE (Pramipexole)	3 (Hauser, Poewe, Schapira)	Serious[1]	Not serious	Serious[2]	Serious[3]	RR 1.24 (0.74, 2.06)	Very low
SAE (Rotigotine)	2 (Giladi, PSG 2007)	Serious[1]	Not serious	Serious[2]	Serious[3]	IRR 1.41 (0.68, 2.92)	Very low
SAE (Ropinirole)	1 (Giladi)	Serious[1]	N/A	Serious[2]	Serious[3]	IRR 1.73 (0.82, 3.63)	Very low
Dopaminergic AE (Pramipexole)	1 (Olanow)	Serious[1]	N/A	Not serious	Serious[3]	IRR 1.20 (0.86, 1.67)	Very low
AE discontinuation (Pramipexole)	3 (Hauser, Poewe, Schapira)	Serious[1]	Serious[4]	Serious[2]	Serious[3]	RR 0.36 (0.02, 5.97)	Very low
AE discontinuation (Rotigotine)	3 (Giladi, Watts, Zhang)	Serious[1]	Serious[4]	Serious[2]	Not serious	RR 2.07 (1.23, 3.48)	Very low
AE discontinuation (Ropinirole)	2 (Adler, Giladi)	Serious[1]	Not serious	Serious[2]	Not serious	RR 2.35 (1.43, 3.86)	Low

[1]Selection of adverse events to report unclear; [2]Population not treatment-naïve; [3]Non-significant result; [4]Considerable between study heterogeneity (i^2 >40%)

Monoamine oxidase inhibitors versus placebo

Short-term follow-up

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
UPDRS (ADL)	3 (Mally, Palhågen, PSG 2002)	Not serious[1]	Not serious	Serious[2]	Not serious[7]	MD -1.14 (-1.57, -0.71)	Moderate

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
UPDRS (motor)	3 (Mally, Palhågen, PSG 2002)	Not serious[1]	Serious[3]	Serious[2]	Serious[5]	MD -4.37 (-7.52, -1.23)	Very low
UPDRS (total)	3 (Hubble, Mally, Palhågen)	Not serious[1]	Serious[3]	Serious[2]	Serious[6]	MD -6.38 (-12.33, -0.43)	Very low
BDI	1 (PSG 2002)	Not serious	N/A	Serious[2]	Serious[4]	MD -0.28 (-0.72, 0.16)	Low
PDQUALIF	1 (PSG 2002)	Not serious	N/A	Serious[2]	Not serious	MD -2.83 (-3.06, -2.59)	Moderate

[1]Individual studies at risk of bias, but overall risk of bias rated low due to consistency of effect between studies at high and low risk of bias; [2]Population not treatment-naïve; [3]Considerable between study heterogeneity (i^2 >40%); [4]Non-significant result; [5]CI cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006); [6]CI cross MID of 7.3 points (Schrag et al., 2006); [7]CI do not cross MID of 3 points (Schrag et al., 2006)

Medium term follow-up (6 months – 2.5 years)

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
UPDRS (ADL)	1 (Palhågen)	Serious[1]	N/A	Not serious	Not serious[3]	MD -0.30 (-1.50, 0.90)	Moderate
UPDRS (motor)	1 (Palhågen)	Serious[1]	N/A	Not serious	Serious[2]	MD -1.90 (-5.26, 1.46)	Low
UPDRS (total)	2 (Olanow, Palhågen)	Serious[1]	Not serious	Not serious	Not serious[4]	MD -3.07 (-3.78, -2.37)	Moderate

[1]Included studies at high risk of bias; [2]CI cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006); [3]CI do not cross MID of 3 points (Schrag et al., 2006); [4]CI do not cross MID of 7.3 points (Schrag et al., 2006)

Adverse events

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Any AE (Rasagiline)	2 (Olanow, Stern)	Serious[1]	Not serious	Serious[2]	Not serious	IRR 0.82 (0.68, 1.00)	Low
SAE (Rasagiline)	1 (PSG 2002)	Serious[1]	N/A	Serious[2]	Serious[2]	RR 2.08 (0.71, 6.09)	Very low
Dopaminergic AE (Rasagiline)	1 (Olanow)	Serious[1]	N/A	Not serious	Serious[2]	IRR 0.72 (0.49, 1.07)	Low

[1]Selection of adverse events to report unclear; [2]Population not treatment-naïve; [3]Non-significant result

Levodopa versus dopamine agonists

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
UPDRS (ADL)	1 (Holloway)	Not serious	N/A	Serious[1]	Not serious[7]	MD -1.10 (-1.98, -0.22)	Moderate
UPDRS (motor) - short	1 (Rascol)	Not serious	N/A	Serious[1]	Serious[5]	MD -2.60 (-4.22, -0.98)	Low
UPDRS (motor)	2 (Holloway, Whone)	Not serious	Not serious	Serious[1]	Serious[5]	MD -4.69 (-6.29, -3.10)	Low
UPDRS (total)	1 (Holloway)	Not serious	N/A	Serious[1]	Serious[6]	MD -4.70 (-7.36, -2.04)	Low
Dyskinesia RR	1 (Whone)	No serious	N/A	Serious[1]	Not serious	RR 7.73 (2.39, 25.00)	Moderate
Any AE (Pramipexole)	1 (Holloway)	Serious[2]	N/A	Serious[1]	Not serious	IRR 0.55 (0.43, 0.70)	Low
Any AE (Ropinirole)	1 (Rascol)	Serious[2]	N/A	Serious[1]	Serious[3]	IRR 0.97 (0.84, 1.11)	Very low
SAE (Pramipexole)	1 (Holloway)	Serious[2]	N/A	Serious[1]	Serious[3]	IRR 0.40 (0.08, 2.08)	Very low

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
SAE (Ropinirole)	2 (Rascol, Whone)	Serious[2]	Not serious	Serious[1]	Serious[3]	RR 1.11 (0.69, 1.80)	Very low
AE discontinuation (Ropinirole)	2 (Rascol, Whone)	Serious[2]	Serious[4]	Serious[1]	Serious[3]	RR 0.73 (0.22, 2.39)	Very low

[1]Population not treatment-naïve; [2]Selection of adverse events to report unclear; [3]Non-significant result; [4]Considerable between study heterogeneity (i^2 >40%); [5]CI cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006); [6]CI cross MID of 7.3 points (Schrag et al., 2006); [7]CI do not cross MID of 3 points (Schrag et al., 2006)

Long-term follow-up (>2.5 years)

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
UPDRS (ADL)	2 (Holloway, Rascol)	Not serious	N/A	Serious[1]	Not serious[4]	MD -1.32 (-2.28, -0.35)	Moderate
UPDRS (motor)	2 (Holloway, Rascol)	Not serious	N/A	Serious[1]	Serious[2]	MD -4.39 (-6.55, -2.23)	Low
UPDRS (total)	2 (Holloway, Rascol)	Not serious	N/A	Serious[1]	Serious[3]	MD -5.20 (-8.90, -1.50)	Low
Dyskinesia	2 (Holloway, Rascol)	Not serious	N/A	Serious[1]	Not serious	RR 2.22 (1.74, 2.82)	Moderate

[1]Population not treatment-naïve; [2]CI cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006); [3]CI cross MID of 7.3 points (Schrag et al., 2006); [4]CI do not cross MID of 3 points (Schrag et al., 2006)

Levodopa versus monoamine oxidase inhibitors

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
UPDRS (ADL) - short	1 (Caraceni)	Serious[1]	N/A	Serious[2]	Not serious[3]	MD -1.10 (-1.62, -0.58)	Low

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
UPDRS (motor) - short	1 (Caraceni)	Serious[1]	N/A	Serious[2]	Not serious[4]	MD -1.00 (-2.07, 0.07)	Low

[1]Included studies at high risk of bias; [2]Population not treatment-naïve; [3]CI do not cross MID of 3 points (Schrag et al., 2006); [4]CI do not cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006)

Long-term follow-up (>2.5 years)

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Need for add-on therapy	1 (Caraceni)	Not serious	N/A	Serious[1]	Not serious	RR 0.20 (0.13, 0.31)	Moderate
Motor fluctuations	1 (Caraceni)	Not serious	N/A	Serious[1]	Not serious	RR 1.58 (1.05, 2.37)	Moderate
Dyskinesia	1 (Caraceni)	Not serious	N/A	Serious[1]	Serious[2]	RR 1.30 (0.87, 1.95)	Low

[1]Population not treatment-naïve; [2]Non-significant result

Dopamine agonists versus monoamine oxidase inhibitors

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Epworth sleep scale	1 (Viallet)	Not serious	N/A	Serious[2]	Not serious	MD 1.92 (0.73, 3.11)	Moderate
Any AE (Pramipexole-Rasagiline)	1 (Viallet)	Serious[1]	N/A	Serious[2]	Serious[3]	RR 1.13 (0.89, 1.43)	Very low
SAE (Pramipexole-Rasagiline)	1 (Viallet)	Serious[1]	N/A	Serious[2]	Serious[3]	RR 0.95 (0.06, 14.75)	Very low
AE discontinuation (Pramipexole-Rasagiline)	1 (Viallet)	Serious[1]	N/A	Serious[2]	Serious[3]	RR 2.83 (0.79, 10.06)	Very low

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
[1]Selection of adverse events to report unclear; [2]Population not treatment-naïve; [3]Non-significant result							

Levodopa versus dopamine agonists versus monoamine oxidase inhibitors

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Levodopa versus levodopa sparing (dopamine agonists and MAOBs)							
Mobility*	1 (PD MED)	Not serious	N/A	Not serious	Serious[1]	MD 1.8 [0.5, 3.0]	Moderate
Activities of daily living*	1 (PD MED)	Not serious	N/A	Not serious	Serious[1]	MD 1.9 [0.7, 3.0]	Moderate
Emotional wellbeing*	1 (PD MED)	Not serious	N/A	Not serious	Serious[2]	MD -0.2 [-1.1, 0.7]	Moderate
Stigma*	1 (PD MED)	Not serious	N/A	Not serious	Serious[1]	MD 1.3 [0.2, 2.3]	Moderate
Social support*	1 (PD MED)	Not serious	N/A	Not serious	Serious[2]	MD 0.1 [-0.6, 0.8]	Moderate
Cognition*	1 (PD MED)	Not serious	N/A	Not serious	Serious[2]	MD 1.0 [0.0, 2.0]	Moderate
Communication*	1 (PD MED)	Not serious	N/A	Not serious	Serious[2]	MD 0.9 [0.0, 1.8]	Moderate
Bodily discomfort*	1 (PD MED)	Not serious	N/A	Not serious	Serious[1]	MD 1.4 [0.3, 2.4]	Moderate
PDQ summary index	1 (PD MED)	Not serious	N/A	Not serious	Not serious	MD 1.0 [0.3, 1.7]	High
EQ-5D utility	1 (PD MED)	Not serious	N/A	Not serious	Not serious	MD 0.03 [0.01, 0.05]	High
Dyskinesia	1 (PD MED)	Not serious	N/A	Not serious	Not serious	HR 1.52 [1.16, 2.00]	High

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Discontinuation due to adverse events	1 (PD MED)	Not serious	N/A	Not serious	Not serious	RR 0.08 [0.04, 0.15]	High
Discontinuation due to lack of efficacy	1 (PD MED)	Not serious	N/A	Not serious	Not serious	RR 0.09 [0.04, 0.22]	High
Dopamine agonists versus MAOBs							
Mobility*	1 (PD MED)	Not serious	N/A	Not serious	Serious[2]	MD 1.4 [0.0, 2.9]	Moderate
Activities of daily living*	1 (PD MED)	Not serious	N/A	Not serious	Serious[2]	MD 0.3 [-1.1, 1.7]	Moderate
Emotional wellbeing*	1 (PD MED)	Not serious	N/A	Not serious	Serious[2]	MD 0.3 [-0.8, 1.4]	Moderate
Stigma*	1 (PD MED)	Not serious	N/A	Not serious	Serious[2]	MD 1.3 [0.0, 2.5]	Moderate
Social support*	1 (PD MED)	Not serious	N/A	Not serious	Serious[2]	MD 0.8 [-0.1, 1.7]	Moderate
Cognition*	1 (PD MED)	Not serious	N/A	Not serious	Serious[1]	MD 1.7 [0.5, 2.9]	Moderate
Communication*	1 (PD MED)	Not serious	N/A	Not serious	Serious[2]	MD 0.5 [-0.6, 1.5]	Moderate
Bodily discomfort*	1 (PD MED)	Not serious	N/A	Not serious	Serious[2]	MD 0.7 [-0.6, 2.0]	Moderate
PDQ summary index	1 (PD MED)	Not serious	N/A	Not serious	Serious[2]	MD 0.8 [0.0, 1.7]	Moderate
EQ-5D utility	1 (PD MED)	Not serious	N/A	Not serious	Serious[2]	MD 0.004 [-0.01, 0.02]	Moderate

*PDQ subscale [1]Significant result but mean difference below trials defined MID [2]Non-significant result

Network meta-analyses

UPDRS II (ADL): <6 months follow-up

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in UPDRS II (ADL) score					
10 Mally et al., 1995; Caraceni et al., 2001; Hauser et al., 2010; Jankovic et al., 2007; Mizuno et al., 2013; Hubble et al., 1995; Palhågen et al., 1998; Parkinson Study Group 1997; Parkinson Study Group 2002; Zhang et al., 2016	Not serious	Serious[1]	Serious[2]	Not serious	Low
[1]Considerable between study heterogeneity (i^2 >40%) [2]Population not treatment-naïve					

UPDRS II (ADL): 6 months to 2.5 years follow-up

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in UPDRS II (ADL) score					
6 Fahn et al., 1995; Schapira et al., 2013; Palhågen et al., 1998; Poewe et al., 2011; Pahwa et al., 2014; Parkinson Study Group 2002	Not serious	Serious[1]	Serious[2]	Not serious	Low
[1]Considerable between study heterogeneity (i^2 >40%) [2]Population not treatment-naïve					

UPDRS III (motor): <6 months follow-up

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in UPDRS III (motor) score					
10 Mally et al., 1995; Caraceni et al., 2001; Hauser et al., 2010; Jankovic et al., 2007;	Not serious	Serious[1]	Serious[2]	Not serious	Low

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Mizuno et al., 2013; Hubble et al., 1995; Palhågen et al., 1998; Parkinson Study Group 1997; Parkinson Study Group 2002; Rascol et al., 2000					
[1]Considerable between study heterogeneity (i^2 >40%) [2]Population not treatment-naïve					

UPDRS III (motor): 6 months to 2.5 years follow-up

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in UPDRS III (motor) score					
7 Fahn et al., 1995; Schapira et al., 2013; Palhågen et al., 1998; Poewe et al., 2011; Pahwa et al., 2014; Parkinson Study Group 2002; Whone et al., 2003	Not serious	Serious[1]	Serious[2]	Not serious	Low
[1]Considerable between study heterogeneity (i^2 >40%) [2]Population not treatment-naïve					

UPDRS total: <6 months follow-up

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in UPDRS total score					
5 Adler et al., 1997; Mally et al., 1995; Hubble et al., 1995; Palhågen et al., 1998; Parkinson Study Group 1997	Not serious	Not serious	Serious[1]	Not serious	Moderate
[2]Population not treatment-naïve					

UPDRS total: 6 months to 2.5 years follow-up

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in UPDRS total score					

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
5 Fahn et al., 1995; Schapira et al., 2013; Palhågen et al., 1998; Parkinson Study Group 2002; Olanow et al., 2009	Not serious	Not serious	Serious[2]	Not serious	Moderate
[2]Population not treatment-naïve					

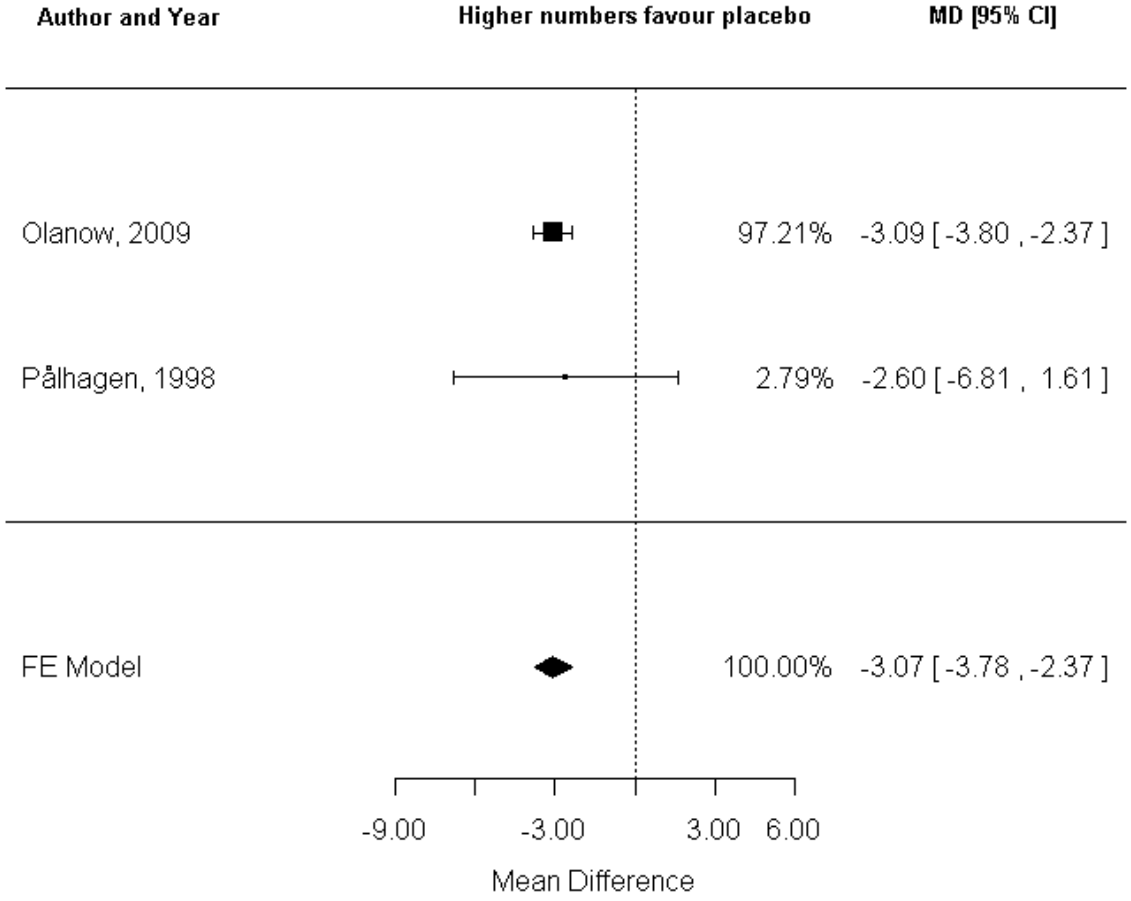
Epworth Sleep Scale

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in ESS score					
3 Hauser et al., 2010; Jankovic et al., 2007; Viallet et al., 2013	Not serious	Serious[1]	Serious[2]	Not serious	Low
[2]Population not treatment-naïve					

Meta-analyses

Treatment-naïve population

Direct meta-analysis - change in UPDRS (total) from baseline to 36 weeks/12 months (MAOBs vs placebo)



tau^2 (estimated amount of total heterogeneity): 0 (SE = 3.3554)

tau (square root of estimated tau^2 value): 0

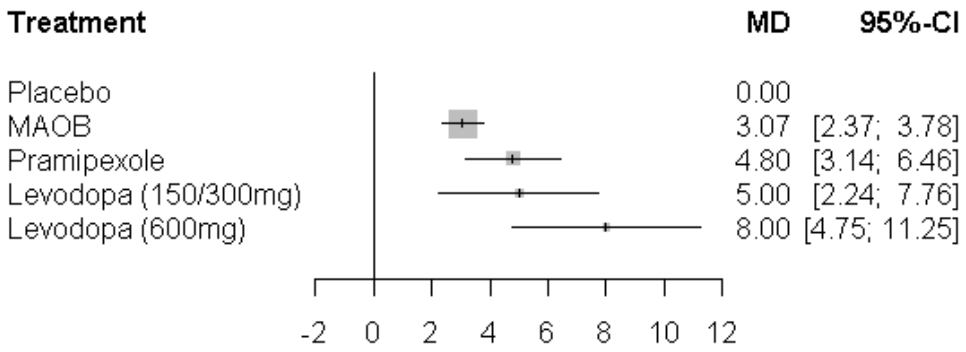
I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 1) = 0.0497, p-val = 0.8236

Network meta-analysis - UPDRS (total) – FE model



Differences between treatments – mean and 95% confidence interval

	Treatment A					
Treatment B		Placebo	MAOB	Pramipexole	Levodopa (150/300mg)	Levodopa (600mg)
	Placebo	N/A				
	MAOB	3.07 (2.37, 3.78)	N/A			
	Pramipexole	4.80 (3.14, 6.46)	1.73 (-0.08, 3.53)	N/A		
	Levodopa (150/300mg)	5.00 (2.25, 7.76)	1.93 (-0.92, 4.77)	0.20 (-3.02, 3.42)	N/A	
	Levodopa (600mg)	8.00 (4.75, 11.25)	4.93 (1.60, 8.26)	3.20 (-0.45, 6.85)	3.00 (0.49, 5.51)	N/A

Quantifying heterogeneity/inconsistency:

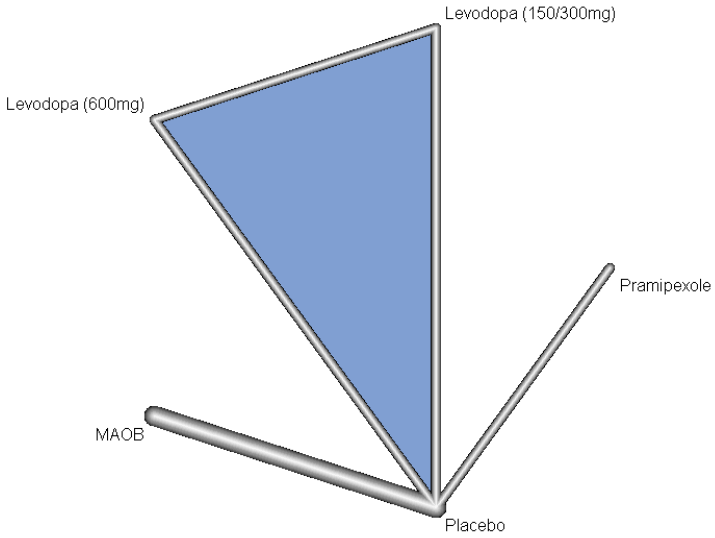
tau^2 < 0.0001; I^2 = 0%

Test of heterogeneity/inconsistency:

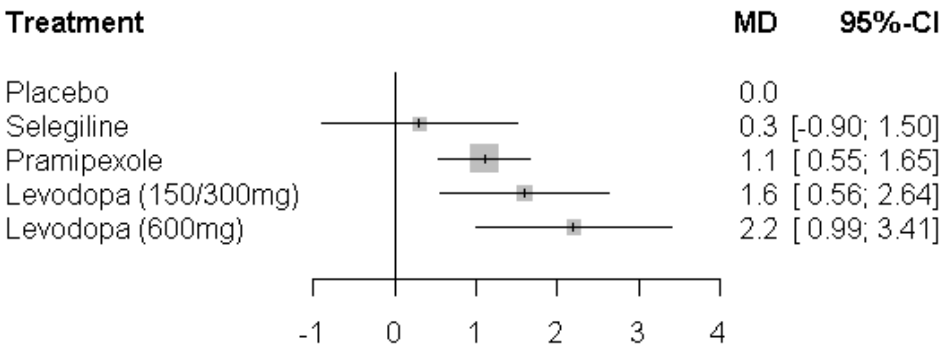
Q d.f. p.value

0.05 1 0.8236

Network graph:



Network meta-analysis - UPDRS 2 (ADL) – FE model



Differences between treatments – mean and 95% confidence interval

	Treatment A					
Treatment B		Placebo	Selegiline	Pramipexole	Levodopa (150/300mg)	Levodopa (600mg)
	Placebo	N/A				
	Selegiline	0.30 (-0.90, 1.50)	N/A			
	Pramipexole	1.10 (0.55, 1.65)	0.80 (-0.52, 2.12)	N/A		
	Levodopa (150/300mg)	1.60 (0.56, 2.64)	1.30 (-0.29, 2.89)	0.50 (-0.68, 1.68)	N/A	
	Levodopa (600mg)	2.20 (0.99, 3.41)	1.90 (0.20, 3.60)	1.10 (-0.23, 2.43)	0.60 (-0.29, 1.49)	N/A

Quantifying heterogeneity/inconsistency:

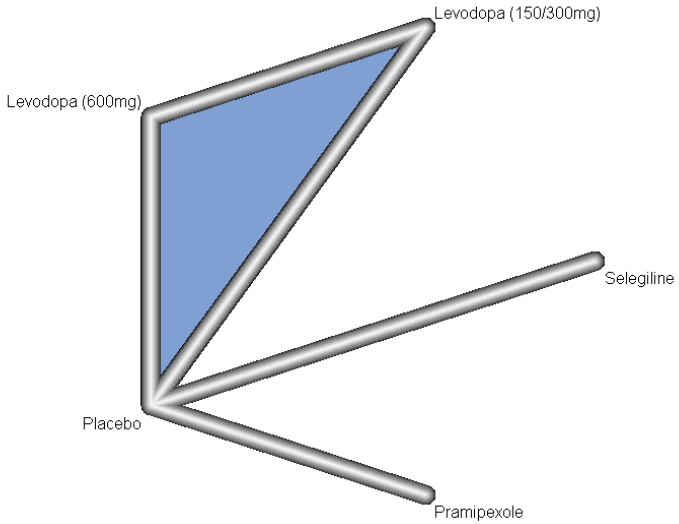
tau^2 < 0.0001; I^2 = 100%

Test of heterogeneity/inconsistency:

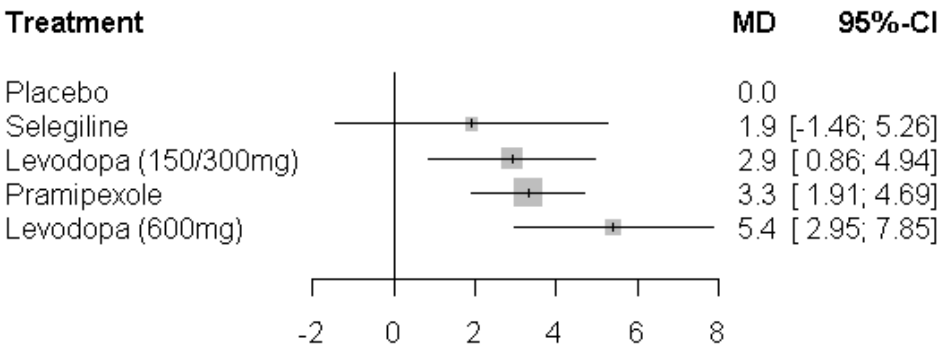
Q d.f. p.value

0 0 <0.0001

Network graph:



Network meta-analysis - UPDRS 3 (motor) – FE model



Differences between treatments – mean and 95% confidence interval

	Treatment A					
Treatment B		Placebo	Selegiline	Levodopa (150/300mg)	Pramipexole	Levodopa (600mg)
	Placebo	N/A				
	Selegiline	1.90 (-1.46, 5.26)	N/A			
	Levodopa (150/300mg)	2.90 (0.86, 4.94)	1.00 (-2.92, 4.93)	N/A		
	Pramipexole	3.30 (1.91, 4.69)	1.40 (-2.23, 5.03)	0.40 (-2.07, 2.86)	N/A	
	Levodopa (600mg)	5.40 (2.95, 7.85)	3.50 (-0.65, 7.65)	2.50 (0.55, 4.45)	2.10 (-0.71, 4.91)	N/A

Quantifying heterogeneity/inconsistency:

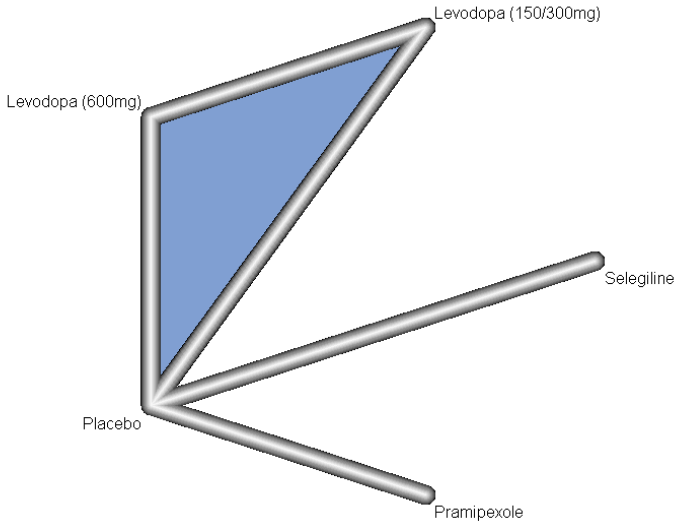
tau^2 < 0.0001; I^2 = 100%

Test of heterogeneity/inconsistency:

Q d.f. p.value

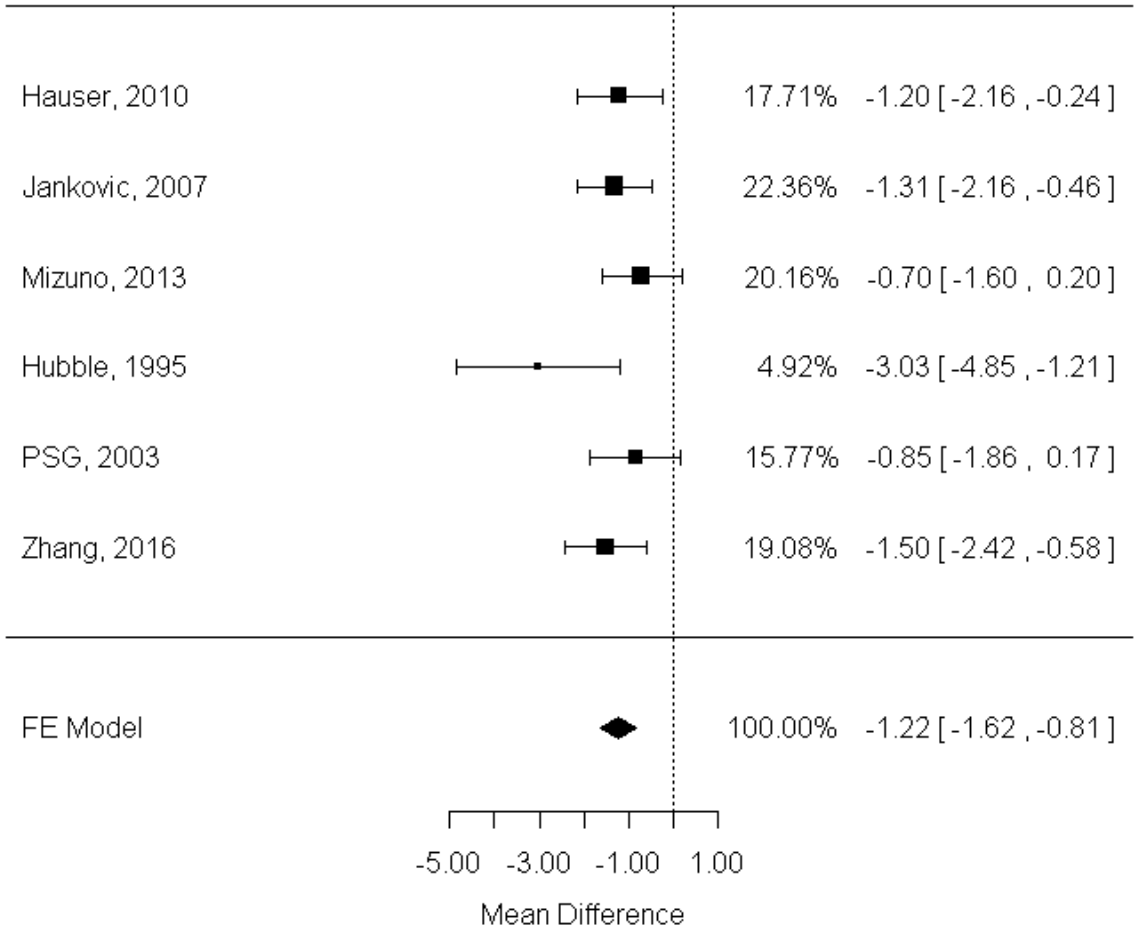
0 0 <0.0001

Network graph:



Full population

Direct meta-analysis – short-term (≤6 months) change in UPDRS (ADL) (dopamine agonists vs placebo)



Random-Effects Model (k = 5; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0000 (SE = 0.1561)

tau (square root of estimated tau^2 value): 0.0001

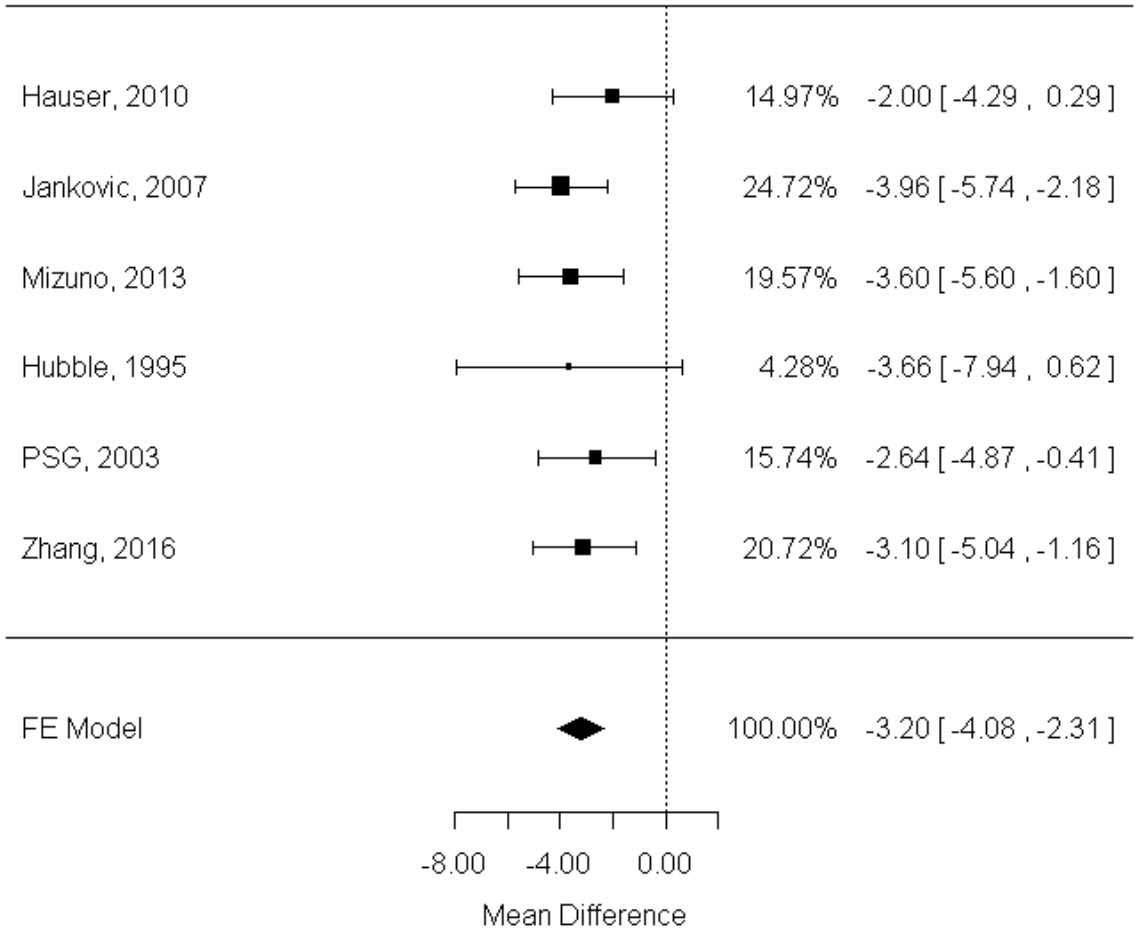
I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 4) = 5.9902, p-val = 0.3072

Direct meta-analysis – short-term (≤6 months) change in UPDRS (motor) (dopamine agonists vs placebo)



Random-Effects Model (k = 6; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.7433)

tau (square root of estimated tau^2 value): 0

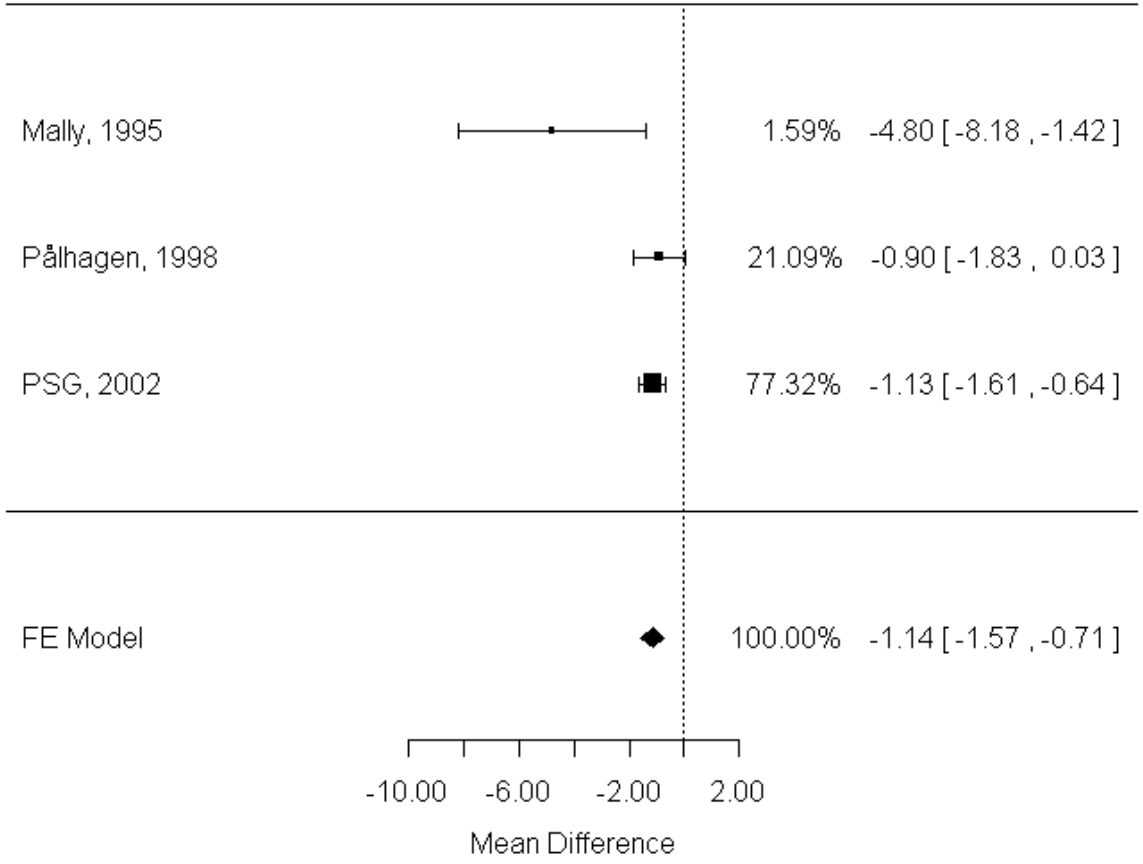
I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 5) = 2.2088, p-val = 0.8196

Direct meta-analysis – short-term (≤6 months) change in UPDRS (ADL) (MAOBs vs placebo)



Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0000 (SE = 0.2004)

tau (square root of estimated tau^2 value): 0.0012

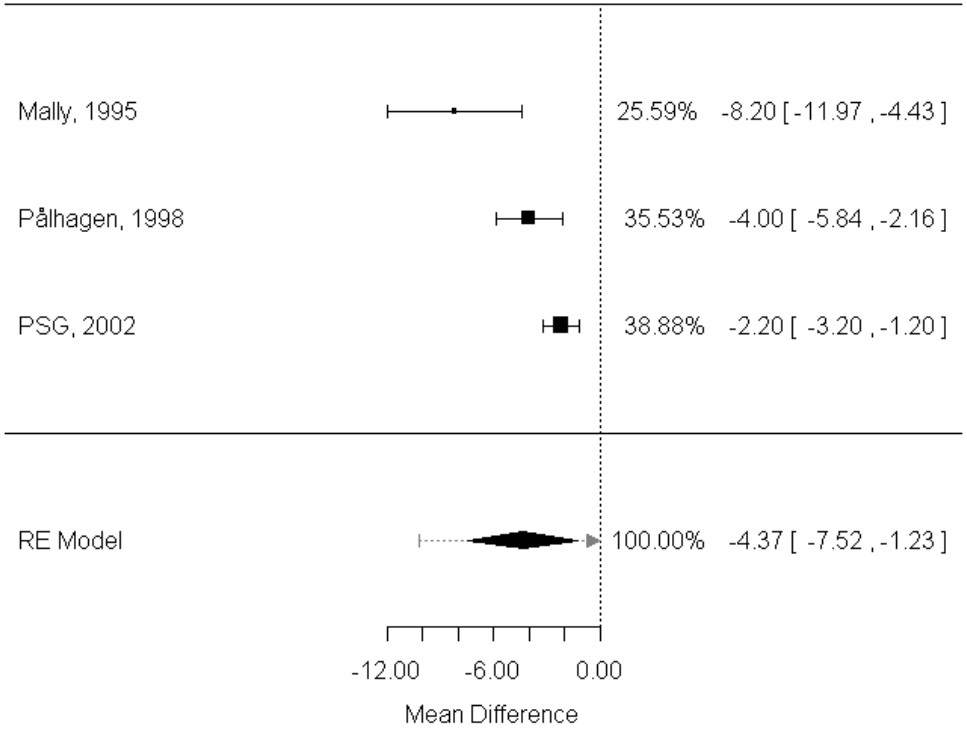
I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 2) = 4.7529, p-val = 0.0929

Direct meta-analysis – short-term (≤6 months) change in UPDRS (motor) (MAOBs vs placebo)



Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 6.3590 (SE = 7.7656)

tau (square root of estimated tau^2 value): 2.5217

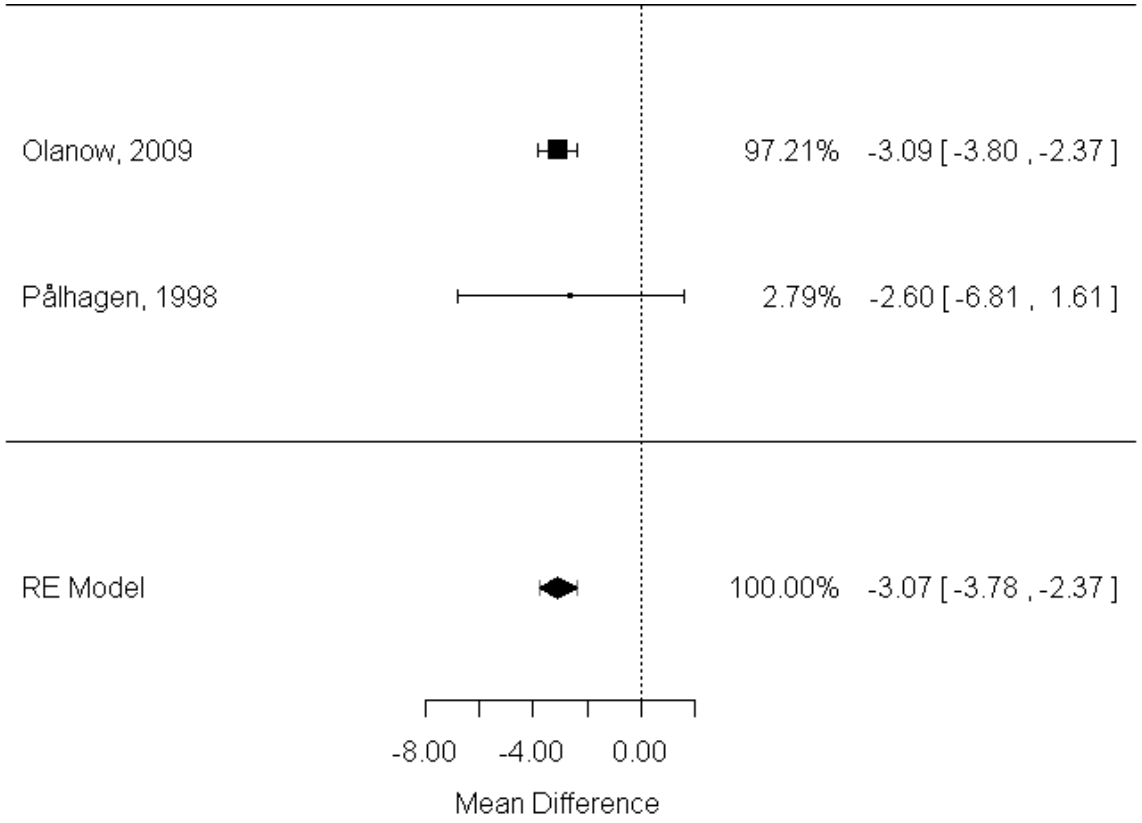
I^2 (total heterogeneity / total variability): 87.34%

H^2 (total variability / sampling variability): 7.90

Test for Heterogeneity:

Q(df = 2) = 10.8437, p-val = 0.0044

Direct meta-analysis – medium term (6 months – 2.5 years) change in UPDRS (total) (MAOBs vs placebo)



Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 3.3554)

tau (square root of estimated tau^2 value): 0

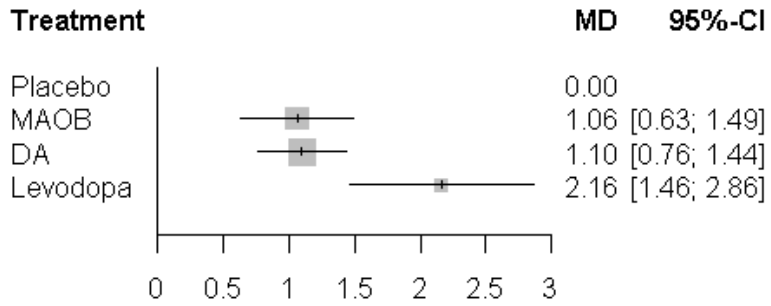
I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 1) = 0.0497, p-val = 0.8236

Network meta-analysis - UPDRS 2 (ADL) – short – RE model



Differences between treatments – mean and 95% confidence interval

	Treatment A				
Treatment B		Placebo	MAOB	Dopamine agonists	Levodopa
	Placebo	N/A			
	MAOB	1.06 (0.63, 1.49)	N/A		
	Dopamine agonists	1.10 (0.76, 1.44)	0.04 (-0.51, 0.58)	N/A	
	Levodopa	2.16 (1.46, 2.86)	1.10 (0.55, 1.65)	1.06 (0.29, 1.84)	N/A

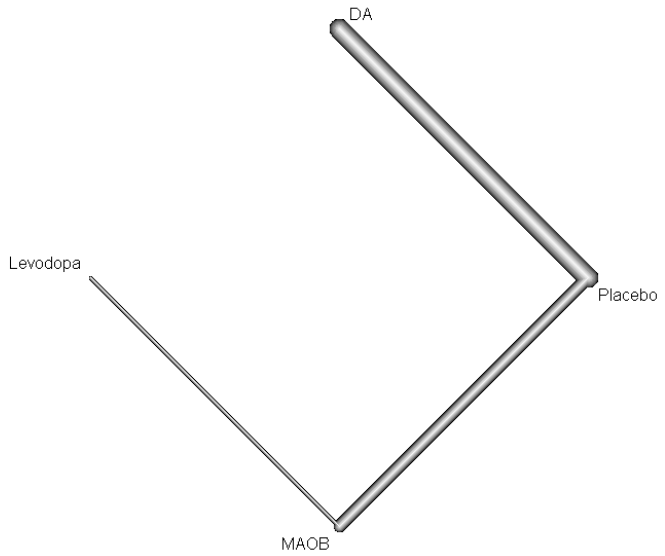
Quantifying heterogeneity/inconsistency:

tau^2 = 0.0743; I^2 = 54.9%

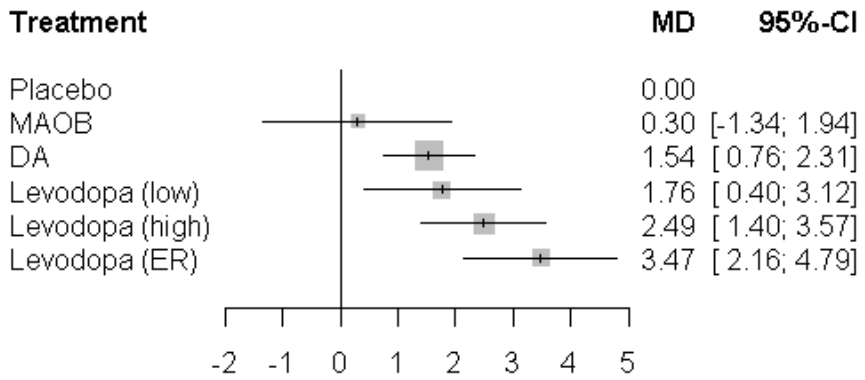
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
13.3	6	0.0385

Network graph:



Network meta-analysis - UPDRS 2 (ADL) – medium – RE model



Differences between treatments – mean and 95% confidence interval

	Placebo	MAOB	Dopamine agonists	Levodopa (low)	Levodopa (high)	Levodopa (ER)
Placebo	N/A					
MAOB	0.30 (-1.34, 1.94)	N/A				
Dopamine agonists	1.54 (0.76, 2.31)	1.54 (0.32, 2.77)	N/A			
Levodopa (low)	1.76 (0.40, 3.12)	1.70 (0.26, 3.14)	0.22 (-1.20, 1.64)	N/A		
Levodopa (high)	2.49 (1.40, 3.57)	2.57 (1.29, 3.85)	0.95 (-0.04, 1.94)	0.73 (-0.58, 2.04)	N/A	
Levodopa (ER)	3.47 (2.16, 4.79)	3.17 (1.78, 4.57)	1.94 (0.41, 3.47)	1.72 (-0.18, 3.61)	0.99 (-0.72, 2.69)	N/A

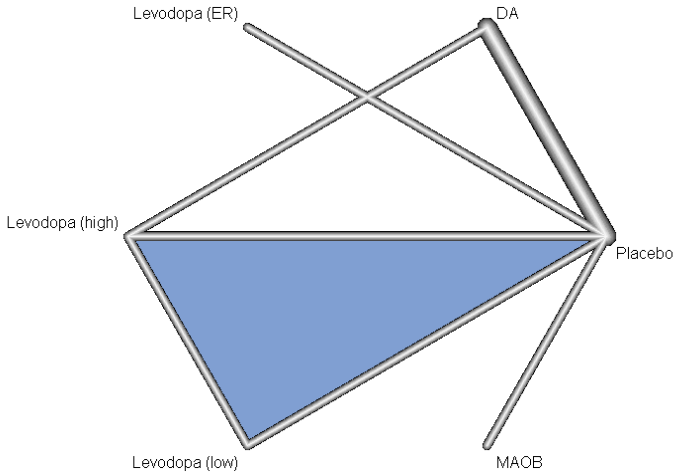
Quantifying heterogeneity/inconsistency:

tau^2 = 0.3201; I^2 = 80.9%

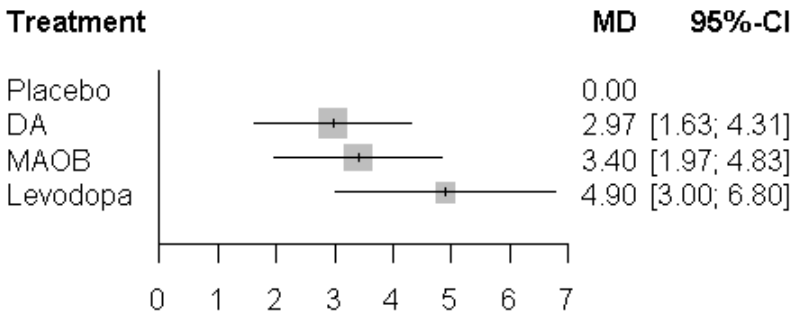
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
10.47	2	0.0053

Network graph:



Network meta-analysis - UPDRS 3 (motor) – short – RE model



Differences between treatments – mean and 95% confidence interval

	Treatment A				
Treatment B		Placebo	MAOB	Dopamine agonists	Levodopa
	Placebo	N/A			
	Dopamine agonists	2.97 (1.63, 4.31)	N/A		
	MAOB	3.40 (1.97, 4.83)	0.43 (-1.34, 2.20)	N/A	
	Levodopa	4.90 (3.00, 6.80)	1.93 (0.07, 3.79)	1.50 (-0.23, 3.23)	N/A

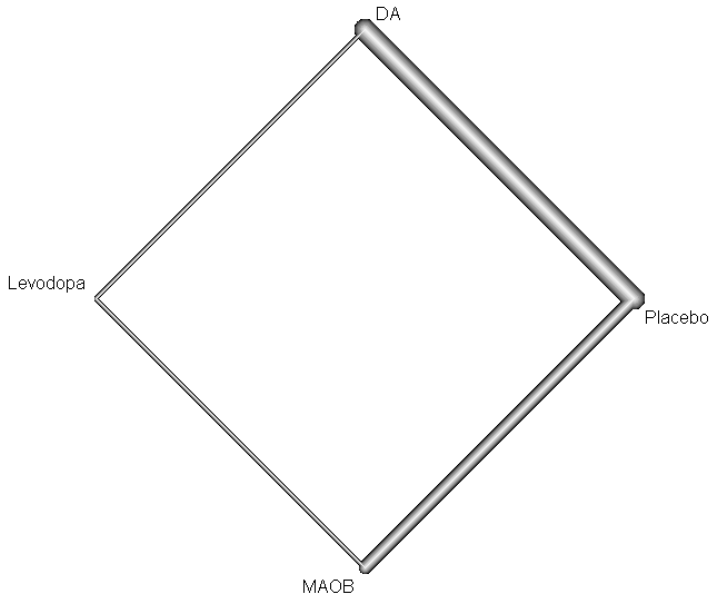
Quantifying heterogeneity/inconsistency:

tau^2 = 1.0095; I^2 = 55.2%

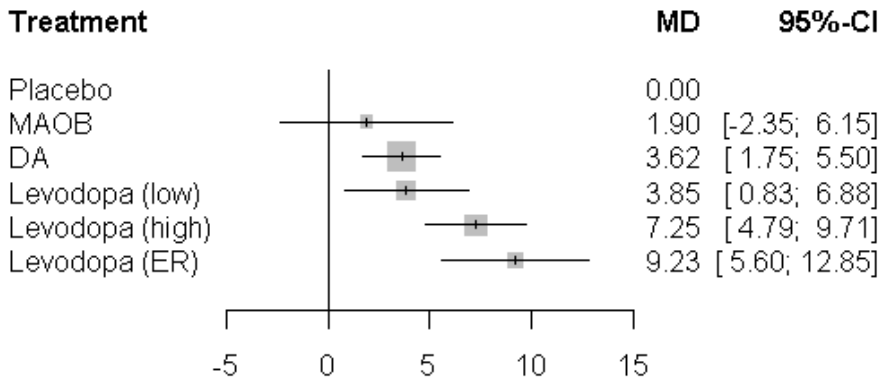
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
15.6	7	0.0289

Network graph:



Network meta-analysis - UPDRS 3 (motor) – medium – RE model



Differences between treatments – mean and 95% confidence interval

	Placebo	MAOB	Dopamine agonists	Levodopa (low)	Levodopa (high)	Levodopa (ER)
Placebo	N/A					
MAOB	1.90 (-2.35, 6.15)	N/A				
Dopamine agonists	3.62 (1.75, 5.50)	1.72 (-2.92, 6.37)	N/A			
Levodopa (low)	3.85 (0.83, 6.88)	1.95 (-3.26, 7.17)	0.23 (-2.99, 3.45)	N/A		
Levodopa (high)	7.25 (4.79, 9.71)	5.35 (0.44, 10.26)	3.63 (1.38, 5.88)	3.40 (0.40, 6.40)	N/A	
Levodopa (ER)	9.23 (5.60, 12.85)	7.33 (1.74, 12.91)	5.60 (1.52, 9.68)	5.37 (0.65, 10.10)	1.98 (-2.41, 6.36)	N/A

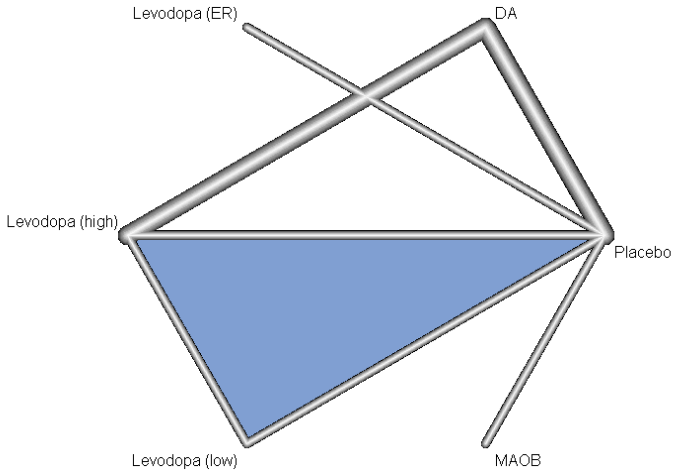
Quantifying heterogeneity/inconsistency:

tau^2 = 1.7971; I^2 = 67.0%

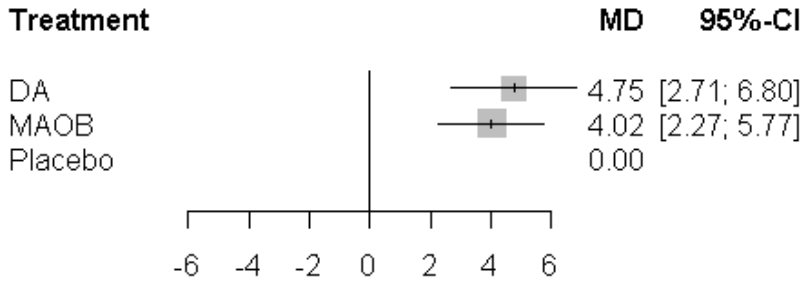
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
9.09	3	0.0282

Network graph:



Network meta-analysis - UPDRS (total) – short – FE model



Differences between treatments – mean and 95% confidence interval

	Treatment A			
Treatment B		Placebo	MAOB	Dopamine agonists
	Placebo	N/A		
	MAOB	4.02 (2.27, 5.77)	N/A	
	Dopamine agonists	4.75 (2.71, 6.80)	0.74 (-1.96, 3.43)	N/A

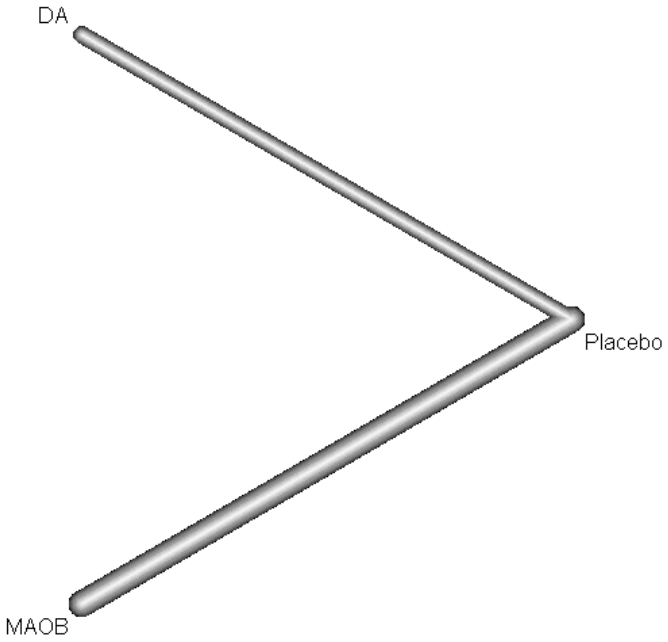
Quantifying heterogeneity/inconsistency:

tau^2 = 0.0732; I^2 = 1.8%

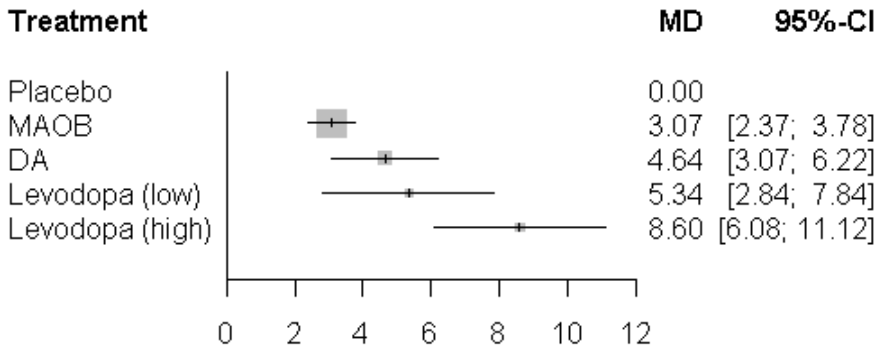
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
3.06	3	0.383

Network graph:



Network meta-analysis - UPDRS (total) – medium – FE model



Differences between treatments – mean and 95% confidence interval

	Placebo	MAOB	Dopamine agonists	Levodopa (low)	Levodopa (high)
Placebo	N/A				
MAOB	3.07 (2.37, 3.78)	N/A			
Dopamine agonists	4.64 (3.07, 6.22)	1.57 (-0.15, 3.29)	N/A		
Levodopa (low)	5.34 (2.84, 7.84)	2.26 (-0.33, 4.86)	0.69 (-2.04, 3.43)	N/A	
Levodopa (high)	8.60 (6.08, 11.12)	5.53 (2.91, 8.14)	3.96 (1.39, 6.53)	3.26 (0.92, 5.51)	N/A

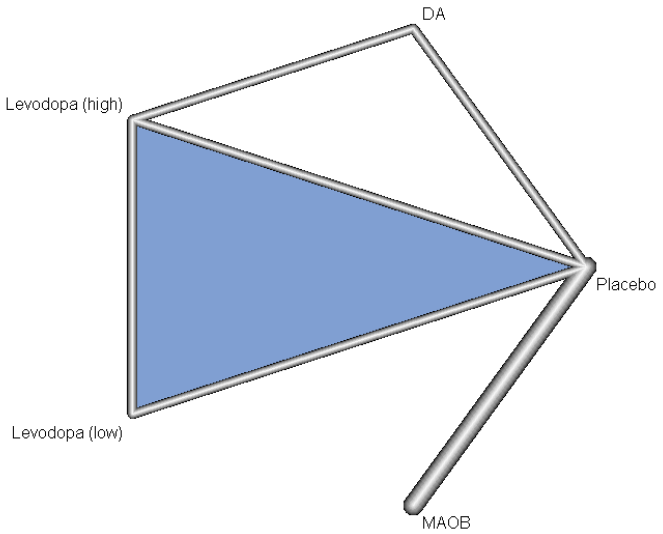
Quantifying heterogeneity/inconsistency:

tau^2 < 0.0001; I^2 = 0%

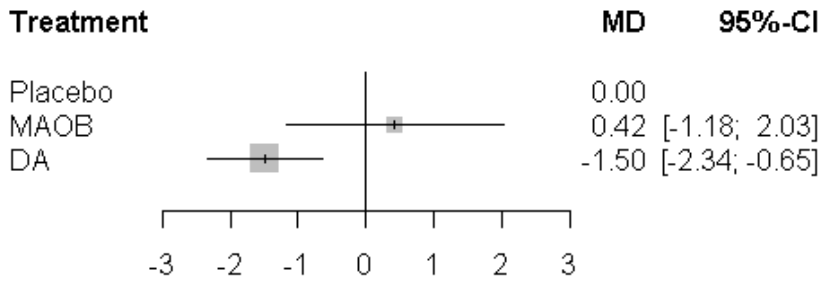
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
0.38	2	0.8283

Network graph:



Network meta-analysis - Epworth sleep scale – RE model



Differences between treatments – mean and 95% confidence interval

	Treatment A			
Treatment B		Placebo	MAOB	Dopamine agonists
	Placebo	N/A		
	MAOB	0.42 (-1.18, 2.03)	N/A	
	Dopamine agonists	-1.50 (-2.34, -0.65)	-1.92 (-2.64, -1.20)	N/A

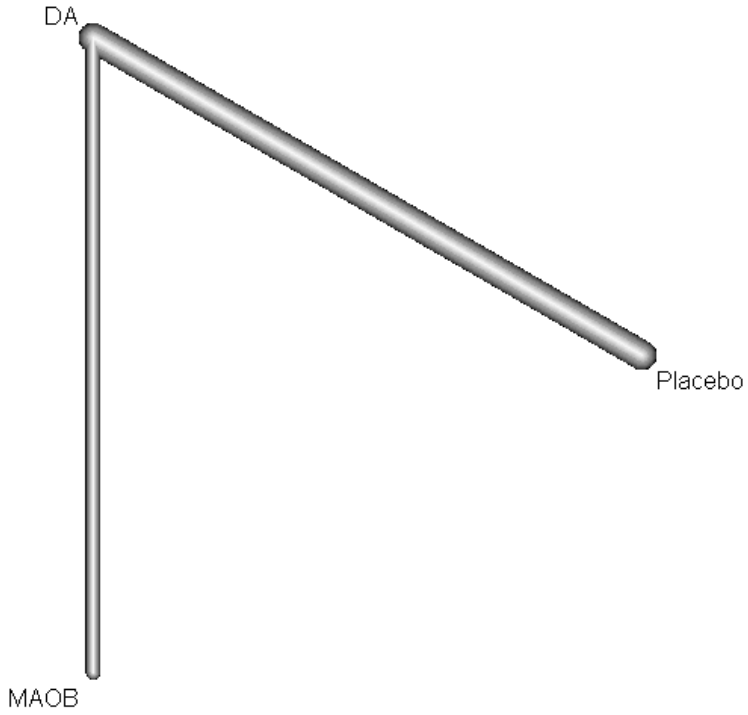
Quantifying heterogeneity/inconsistency:

tau^2 = 0.3508; I^2 = 94.4%

Test of heterogeneity/inconsistency:

Q	d.f.	p.value
17.81	1	<0.0001

Network graph:



E.2.2 Adjuvant treatment of motor symptoms

Efficacy outcomes by drug classes – Pairwise meta-analyses

Dopamine agonists vs. placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Off time	19[a]	Not serious	Serious[2]	Serious[3]	Not serious	MD -1.42 [-1.83, -1.01]	Low
UPDRS II (ADL)	14[b]	Not serious	Serious[2]	Serious[3]	Not serious[7]	MD -1.72 [-2.16, -1.27]	Low
UPDRS III (motor)	15[c]	Not serious	Serious[2]	Serious[3]	Serious[5]	MD -4.09 [-5.25, -2.92]	Very low
PDQ-39	2[d]	Serious[1]	Serious[2]	Serious[3]	Serious[6]	MD -1.88 [-5.40, 1.64]	Very Low
PDQUALIF	1[e]	Serious[1]	N/A	Serious[3]	Serious[4]	MD -3.22 [-6.86, 0.42]	Very Low

a Stowe Cochrane review 2010 (n=15: Interntl; Germany; Spain; UK; USA I; N America; Aust/Germ; CLEOPATRA; Denmark; Europe; US/Canada; EASE-PD; France/Eng; UK/Israel; USA) Nicholas 2014; Nomoto 2014; Pahwa 2007; Poewe 2007

b Stowe Cochrane review 2010 (n=6: Spain; USA I; Aust/Germ; CLEOPATRA; Europe; H Kong/Taiw); Mizuno 2003; Mizuno 2007; Nicholas 2014; Nomoto 2014; Pahwa 2007; Poewe 2007; PSG 2007; Watts 2010

c Stowe Cochrane review 2010 (n=7: Spain; USA I; Aust/Germ; CLEOPATRA; Europe; H Kong/Taiw; EASE-PD); Mizuno 2003; Mizuno 2007; Nicholas 2014; Nomoto 2014; Pahwa 2007; Poewe 2007; PSG 2007; Watts 2010

d Poewe 2007; Watts 2010

e PSG 2007

[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result; [5]CI cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006); [6]CI cross MID of 1.6 points (Peto et al., 2001); [7]CI do not cross MID of 3 points (Schrag et al., 2006)

COMTIs versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Off time	13[a]	Not serious	Not serious	Serious[3]	Not serious	MD -0.81 [-1.01, -0.60]	Moderate
UPDRS II (ADL)	12[b]	Not serious	Not serious	Serious[3]	Not serious[5]	MD -0.99 [-1.35, -0.63]	Moderate
UPDRS III (motor)	13[c]	Not serious	Not serious	Serious[3]	Not serious[6]	MD -2.11 [-2.74, -1.47]	Moderate
PDQ-39	1[d]	Not serious	N/A	Serious[3]	Serious[4]	MD 6.90 [-4.05, 17.85]	Low

a Stowe Cochrane review 2010 (n=12: Celomen; ComQol; INT-01; LARGO; Nomecomt; Sth Korea; UK/Irish; China; Europe; TFSG I; TFSG 3; TIPS I); Fenelon 2003

b Stowe Cochrane review 2010 (n=10: Celomen; ComQol; INT-02; Nomecomt; Sth Korea; UK/Irish; TFSG 3; TIPS I; TIPS II; US/Canada); Fenelon 2003; Tolosa 2014

c Stowe Cochrane review 2010 (n=12: Celomen; ComQol; Interntl; LARGO; Nomecomt; Sth Korea; UK/Irish; Europe; TFSG 3; TIPS I; TIPS II; US/Canada); Tolosa 2014

d Tolosa 2014

[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]CI cross MID of 1.6 points (Peto et al., 2001); [5]CI do not cross MID of 3 points (Schrag et al., 2006); [6]CI do not cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006)

MAOBIs versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Off time	4[a]	Not serious	Not serious	Serious[1]	Not serious	MD -0.98 [-1.22, -0.74]	Moderate
UPDRS II (ADL)	1[b]	Not serious	Not serious	Serious[1]	Not serious[2]	MD -1.85 [-2.62, -1.08]	Moderate
UPDRS III (motor)	2[c]	Not serious	N/A	Serious[1]	Not serious[3]	MD -2.29 [-3.05, -1.54]	Moderate

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
a Stowe Cochrane review 2010 (n=3: LARGO; PRESTO; USA); Zhang 2013 b Zhang 2013 c Stowe 2010 (n=1: LARGO); Zhang 2013							
[1]Population not as defined in protocol; [2]CI do not cross MID of 3 points (Schrag et al., 2006); [3]CI do not cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006)							

Dopamine agonists versus COMTIs

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
UPDRS II (ADL)	2[a]	Serious[1]	Not serious	Serious[2]	Not serious[4]	MD 0.40 [-0.48, 1.27]	Low
UPDRS III (motor)	2[a]	Serious[1]	Not serious	Serious[2]	Not serious[5]	MD -0.10 [-2.06, 1.86]	Low
Off time	2[a]	Serious[1]	Not serious	Serious[2]	Serious[3]	MD -0.11 [-0.83, 0.60]	Very Low
PDQ-39	1[b]	Serious[1]	N/A	Serious[2]	Serious[6]	MD -2.90 [-6.38, 0.58]	Very low
a Deane 2004 (n=1); Deuschl 2007 b Deuschl 2007							
[1]Individual study(ies) at risk of bias; [2]Population not as defined in protocol; [3]Non-significant result; [4]CI do not cross MID of 3 points (Schrag et al., 2006); [5]CI do not cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006); [6]CI cross MID of 1.6 points (Peto et al., 2001)							

Amantadine versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Hyperkinesia (CDRS)	1[a]	Not serious	N/A	Serious[3]	Serious[4]	MD -6.20 [-14.37, 1.97]	Low
Dystonia (CDRS)	1[a]	Not serious	N/A	Serious[3]	Serious[4]	MD -0.40 [-4.06, 3.26]	Low
UPDRS II	1[a]	Not serious	N/A	Serious[3]	Serious[5]	MD -1.70	Low

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
						[-9.05, 5.65]	
UPDRS III	1[a]	Not serious	N/A	Serious[3]	Serious[6]	MD -2.40 [-9.39, 4.59]	Low
a da Silvia-Junior 2005							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; 4Non-significant result; 5CI cross MID of 3 points (Schrag et al., 2006); 6CI cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006)							

Safety outcomes by individual drugs – Pairwise meta-analyses

Ropinirole versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	7[a]	Serious[1]	Serious[2]	Serious[3]	Serious[4]	RR 2.36 [0.77, 7.22]	Very Low
Hallucinations	3[b]	Not serious	Not serious	Serious[3]	Not serious	RR 5.97 [2.23, 16.02]	Moderate
Mortality	3[c]	Not serious	Not serious	Serious[3]	Serious[4]	RR 0.29 [0.03, 2.77]	Low
Any AEs	7[d]	Serious[1]	Not serious	Serious[3]	Not serious	RR 1.15 [1.08, 1.23]	Low
SAEs	3[e]	Serious[1]	Not serious	Serious[3]	Serious[4]	RR 0.94 [0.56, 1.57]	Very Low
AE discontinuation	7[f]	Not serious	Not serious	Serious[3]	Serious[4]	RR 1.11 [0.80, 1.53]	Low
Psychosis (Parkinson's Psychosis Rating Scale)	1[g]	Not serious	N/A	Serious[3]	Serious[4]	MD 0.30 [-0.20, 0.80]	Low
a Stowe Cochrane review 2010 (n=3: EASE-PD; France/Eng; USA); Lieberman 1997; Mizuno 2010; Mizuno 2014; Watts 2010 b Stowe Cochrane review 2010 (n=1: EASE-PD); Mizuno 2010; Mizuno 2014 c Stowe Cochrane review 2010 (n=3: EASE-PD; France/Eng; UK/Israel)							

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
d Stowe Cochrane review 2010 (n=3: EASE-PD; France/Eng; UK/Israel); Mizuno 2010; Mizuno 2014; Pahwa 2007; Watts 2010 e Mizuno 2010; Mizuno 2014; Watts 2010 f Stowe Cochrane review 2010 (n=4: EASE-PD; France/Eng; UK/Israel; USA)); Mizuno 2010; Mizuno 2014; Watts 2010 g Watts 2010							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity; [3]Population not as defined in protocol; [4]Non-significant result							

Rotigotine versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	5[a]	Not serious	Not serious	Serious[3]	Not serious	RR 3.06 [1.95, 4.81]	Moderate
Hallucinations	5[a]	Not serious	Not serious	Serious[3]	Not serious	RR 3.89 [1.82, 8.30]	Moderate
Any AEs	4[b]	Not serious	Serious[2]	Serious[3]	Serious[4]	RR 1.09 [0.99, 1.20]	Low
SAEs	3[c]	Not serious	Not serious	Serious[3]	Serious[4]	RR 0.61 [0.31, 1.19]	Low
AE discontinuation	5[a]	Serious[1]	Not serious	Serious[3]	Serious[4]	RR 0.87 [0.63, 1.21]	Very Low
Mortality	1[d]	Not serious	N/A	Serious[3]	Serious[4]	RR 1.34 [0.06, 27.69]	Low
Impulse Control Disorder	1[d]	Not serious	N/A	Serious[3]	Serious[4]	RR 2.93 [0.16, 52.61]	Low
a Lewitt 2007; Mizuno 2014; Nicholas 2014; Nomoto 2014; Poewe 2007 b Mizuno 2014; Nicholas 2014; Nomoto 2014; Poewe 2007 c Mizuno 2014; Nicholas 2014; Nomoto 2014 d Nicholas 2014							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result							

Pramipexole versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	10[a]	Not serious	Not serious	Serious[3]	Not serious	RR 1.92 [1.61, 2.29]	Moderate
Hallucinations	9[b]	Not serious	Not serious	Serious[3]	Not serious	RR 2.86 [1.99, 4.09]	Moderate
Any AEs	8[c]	Not serious	Not serious	Serious[3]	Not serious	RR 1.08 [1.01, 1.14]	Moderate
SAEs	3[d]	Serious[1]	Not serious	Serious[3]	Serious[4]	1.49 [0.64, 3.44]	Very Low
AE discontinuation	8[c]	Not serious	Not serious	Serious[3]	Serious[4]	RR 0.86 [0.66, 1.12]	Low

a Stowe Cochrane review 2010 (n=6: Aust/Germ; CLEOPATRA; Denmark; Europe; Interntl; US/Canada); Mizuno 2003; Poewe 2007; PSG 2007; Schapira 2011
b Stowe Cochrane review 2010 (n=5: Aust/Germ; CLEOPATRA; Europe; Interntl; US/Canada); Mizuno 2003; Poewe 2007; PSG 2007; Schapira 2011
c Stowe Cochrane review 2010 (n=5: Aust/Germ; CLEOPATRA; Denmark; Interntl; US/Canada); Mizuno 2003; Poewe 2007; Schapira 2011
d Mizuno 2003; PSG 2007; Schapira 2011

[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result

Cabergoline versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	3[a]	Not serious	Not serious	Serious[3]	Not serious	RR 1.29 [1.01, 1.64]	Moderate
Hallucinations	3[a]	Not serious	Not serious	Serious[3]	Serious[4]	RR 2.18 [0.74, 6.46]	Low
Mortality	1[b]	Not serious	N/A	Serious[3]	Serious[4]	RR 0.33 [0.01, 7.72]	Low
Any AEs	3[a]	Not serious	Not serious	Serious[3]	Not serious	RR 1.17 [1.03, 1.34]	Moderate
AE discontinuation	3[a]	Not serious	Serious[2]	Serious[3]	Serious[4]	RR 1.25	Very Low

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
						[0.48, 3.22]	

a Stowe Cochrane review 2010 (n=3: Spain; USA I; USA 2)
b Stowe Cochrane review 2010 (n=1: Spain)

[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result

Bromocriptine versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	3[a]	Not serious	Not serious	Serious[3]	Not serious	RR 1.82 [1.20, 2.76]	Moderate
Hallucinations	3[a]	Not serious	Serious[2]	Serious[3]	Serious[4]	RR 1.93 [0.49, 7.56]	Low
Any AEs	3[a]	Not serious	Not serious	Serious[3]	Not serious	RR 1.17 [1.03, 1.34]	Moderate
AE discontinuation	5[b]	Not serious	Not serious	Serious[3]	Serious[4]	RR 1.02 [0.71, 1.47]	Low

a Stowe Cochrane review 2010 (n=2: Interntl; Japan); Mizuno 2003
b Stowe Cochrane review 2010 (n=4: Interntl; Japan; Rotterdam; South Africa); Mizuno 2003

[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result

Pergolide versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	1[a]	Not serious	N/A	Serious[3]	Not serious	RR 2.54 [1.93, 3.34]	Moderate
Hallucinations	1[a]	Not serious	N/A	Serious[3]	Not serious	RR 4.29 [1.81, 10.18]	Moderate
Mortality	1[a]	Not serious	N/A	Serious[3]	Serious[4]	RR 0.49 [0.05, 5.41]	Low
AE discontinuation	1[a]	Not serious	N/A	Serious[3]	Serious[4]	RR 2.23	Low

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
						0.99, 4.99]	
a Stowe Cochrane review (n=1: N America)							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result							

Entacapone versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	11[a]	Not serious	Not serious	Serious[3]	Not serious	RR 2.01 [1.67, 2.42]	Moderate
Hallucinations	8[b]	Not serious	Serious[2]	Serious[3]	Serious[4]	RR 0.43 [0.03, 6.84]	Very Low
Mortality	1[c]	Not serious	N/A	Serious[3]	Serious[4]	RR 0.40 [0.09, 1.79]	Low
Any AEs	10[d]	Serious1	Serious[2]	Serious[3]	Not serious	RR 1.39 [1.07, 1.81]	Very Low
SAEs	3[e]	Not serious	Not serious	Serious[3]	Serious[4]	RR 0.91 [0.39, 2.12]	Low
AE discontinuation	12[f]	Not serious	Not serious	Serious[3]	Not serious	RR 1.51 [1.17, 1.95]	Moderate
a Stowe Cochrane review 2010 (n=10: Celomen; ComQol; Filomen; INT-02; Japan; LARGO; Nomecomt; Seesaw; Sth Korea; UK/Irish); Fenelon 2003							
b Stowe Cochrane review 2010 (n=7: Celomen; INT-02; LARGO; Nomecomt; Seesaw; Sth Korea; UK/Irish); Fenelon 2003							
c Stowe Cochrane review 2010 (n=1: Filomen)							
d Stowe Cochrane review 2010 (n=7: Celomen; ComQol; INT-02; Japan; LARGO; Seesaw; UK/Irish;); Fenelon 2003; Destee 2009; Tolosa 2014							
e Fenelon 2003; Destee 2009; Tolosa 2014							
f Stowe Cochrane review 2010 (n=9: Celomen; ComQol; Filomen; INT-02; Interntl; Japan; LARGO; Nomecomt; Seesaw); Fenelon 2003; Destee 2009; Tolosa 2014							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result							

Tolcapone versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	6[a]	Not serious	Not serious	Serious[3]	Not serious	RR 2.58 [1.93, 3.44]	Moderate
Hallucinations	4[b]	Not serious	Serious[2]	Serious[3]	Not serious	RR 2.50 [1.23, 5.06]	Low
Any AEs	4[b]	Not serious	Not serious	Serious[3]	Not serious	RR 1.22 [1.10, 1.34]	Moderate
AE discontinuation	5[c]	Not serious	Serious[2]	Serious[3]	Serious[4]	RR 1.47 [0.88, 2.46]	Very Low

a Stowe Cochrane review 2010 (n=6: China; Europe; TFSG 3; TIPS I; TIPS II; US/Canada)
b Stowe Cochrane review 2010 (n=4: TFSG 3; TIPS I; TIPS II; US/Canada)
c Stowe Cochrane review 2010 (n=5: Europe; TFSG 3; TIPS I; TIPS II; US/Canada)

[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result

Rasagiline versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	2[a]	Not serious	Not serious	Serious[3]	Serious[4]	RR 1.19 [0.53, 2.65]	Low
Hallucinations	1[b]	Not serious	N/A	Serious[3]	Serious[4]	RR 1.65 [0.40, 6.83]	Low
Any AEs	3[c]	Not serious	Not serious	Serious[3]	Serious[4]	RR 1.06 [0.93, 1.22]	Low
SAEs	1[d]	Not serious	N/A	Serious[3]	Serious[4]	RR 1.05 [0.07, 16.60]	Low
AE discontinuation	2[a]	Not serious	Not serious	Serious[3]	Serious[4]	RR 0.59 [0.28, 1.28]	Low

a Stowe Cochrane review 2010 (n=1: LARGO); Zhang 2013
b Stowe Cochrane review 2010 (n=1: LARGO)
c Stowe Cochrane review 2010 (n=2: LARGO; PRESTO); Zhang 2013

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
d Zhang 2013							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity; [3]Population not as defined in protocol; [4]Non-significant result							

Selegiline versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	3[a]	Serious[1]	Serious[2]	Serious[3]	Serious[4]	RR 0.86 [0.44, 1.69]	Very Low
Hallucinations	1[b]	Not serious	N/A	Serious[3]	Serious[4]	RR 2.76 [0.30, 25.60]	Low
Any AEs	3[a]	Serious[1]	Not serious	Serious[3]	Serious[4]	RR 1.08 [0.88, 1.33]	Very Low
SAEs	1[c]	Serious[1]	N/A	Serious[3]	Serious[4]	RR 4.00 [0.51, 31.10]	Very Low
AE discontinuation	3[a]	Serious[1]	Serious[2]	Serious[3]	Serious[4]	RR 1.72 [0.14, 20.91]	Very Low
a Stowe Cochrane review 2010 (n=2: Norw/Fin; USA); Ondo 2007 b Stowe Cochrane review 2010 (n=1: USA) c Ondo 2007							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result							

Amantadine versus placebo

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Hyperkinesia (CDRS)	1[a]	Not serious	N/A	Serious[3]	Serious[4]	MD -6.20 [-14.37, 1.97]	Low
Dystonia (CDRS)	1[a]	Not serious	N/A	Serious[3]	Serious[4]	MD -0.40 -4.06, 3.26]	Low

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
UPDRS II	1[a]	Not serious	N/A	Serious[3]	Serious[5]	MD -1.70 -9.05, 5.65]	Low
UPDRS III	1[a]	Not serious	N/A	Serious[3]	Serious[6]	MD -2.40 [-9.39, 4.59]	Low
a da Silvia-Junior 2005							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result; [5]CI cross MID of 3 points (Schrag et al., 2006); [6]CI cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006)							

Ropinirole versus Rotigotine

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	1[a]	Not serious	N/A	Serious[3]	Serious[4]	RR 0.86 [0.51, 1.43]	Low
Hallucinations	1[a]	Not serious	N/A	Serious[3]	Serious[4]	RR 2.01 [0.51, 7.91]	Low
Any AEs	1[a]	Not serious	N/A	Serious[3]	Not serious	RR 0.88 [0.80, 0.97]	Moderate
SAEs	1[a]	Not serious	N/A	Serious[3]	Serious[4]	RR 0.72 [0.23, 2.22]	Low
AE discontinuation	1[a]	Not serious	N/A	Serious[3]	Serious[4]	RR 1.01 [0.48, 2.10]	Low
a Mizuno 2014							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result							

Ropinirole versus Bromocriptine

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	2[a]	Serious[1]	Not serious	Serious[3]	Serious[4]	RR 1.44 [0.66, 3.16]V	Very Low
Hallucinations	2[a]	Serious[1]	Not serious	Serious[3]	Serious[4]	RR 0.76	Very Low

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
						[0.27, 2.15]	
a Clarke Cochrane review 2001b (n=2)							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result							

Pramipexole versus Bromocriptine

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	2[a]	Not serious	Not serious	Serious[3]	Not serious	RR 2.33 [1.14, 4.74]	Moderate
Hallucinations	1[b]	Not serious	N/A	Serious[3]	Serious[4]	RR 0.90 [0.46, 1.75]	Low
Any AEs	1[b]	Not serious	N/A	Serious[3]	Serious[4]	RR 0.94 [0.85, 1.04]	Low
SAEs	1[b]	Not serious	N/A	Serious[3]	Serious[4]	RR 7.14 [0.37, 136.43]	Low
AE discontinuation	1[b]	Not serious	N/A	Serious[3]	Serious[4]	RR 0.69 [0.29, 1.61]	Low
a Stowe Cochrane review 2010 (n=1: Interntl); Mizuno 2003 b Mizuno 2003							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result							

Rotigotine versus Pramipexole

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	1[a]	Not serious	N/A	Serious[3]	Serious[4]	RR 0.76 [0.46, 1.25]	Low
Hallucinations	1[a]	Not serious	N/A	Serious[3]	Serious[4]	RR 0.70 [0.32, 1.55]	Low
Any AEs	1[a]	Not serious	N/A	Serious[3]	Serious[4]	RR 1.00 [0.88, 1.14]	Low

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
AE discontinuation	1[a]	Not serious	N/A	Serious[3]	Serious[4]	RR 0.77 [0.36, 1.66]	Low
a Poewe 2007							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result							

Pramipexole versus Pergolide

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Any AEs	1[a]	Serious[1]	N/A	Serious[3]	Serious[4]	RR 0.80 [0.52, 1.24]	Very Low
AE discontinuation	1[a]	Serious[1]	N/A	Serious[3]	Serious[4]	RR 1.30 [0.24, 6.96]	Very Low
a Rektorova 2003							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result							

Cabergoline versus Bromocriptine

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	5[a]	Not serious	Not serious	Serious[3]	Not serious	RR 1.49 [1.04, 2.13]	Moderate
Hallucinations	5[a]	Not serious	Not serious	Serious[3]	Serious[4]	RR 1.31 [0.89, 1.94]	Low
a Clarke Cochrane review 2001a (n=5)							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result							

Cabergoline versus Entacapone

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Hallucinations	1[a]	Serious[1]	N/A	Serious[3]	Serious[4]	RR 1.04 [0.22, 4.99]	Very Low

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Any AEs	1[a]	Serious[1]	N/A	Serious[3]	Serious[4]	RR 0.99 [0.74, 1.32]	Very Low
SAEs	1[a]	Serious[1]	N/A	Serious[3]	Serious[4]	RR 0.52 [0.13, 2.00]	Very Low
AE discontinuation	1[a]	Serious[1]	N/A	Serious[3]	Serious[4]	RR 1.63 [0.67, 4.00]	Very Low
a Deuschl 2007							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result							

Bromocriptine versus Tolcapone

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	1[a]	Serious[1]	N/A	Serious[3]	Serious[4]	RR 0.74 [0.51, 1.06]	Very Low
Hallucinations	1[a]	Serious[1]	N/A	Serious[3]	Serious[4]	RR 6.81 [0.86, 53.98]	Very Low
a Dean Cochrane review 2004 (n=1)							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result							

Pergolide versus Tolcapone

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	1[a]	Serious[1]	N/A	Serious[3]	Not serious	RR 0.51 [0.34, 0.78]	Low
AE discontinuation	1[a]	Serious[1]	N/A	Serious[3]	Not serious	RR 2.97 [1.12, 7.87]	Low
a Dean Cochrane review 2004 (n=1)							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result							

Entacapone versus Tolcapone

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Dyskinesia	1[a]	Not serious	N/A	Serious[3]	Serious[4]	RR 0.96 [0.59, 1.56]	Low
Hallucinations	1[a]	Not serious	N/A	Serious[3]	Serious[4]	RR 7.00 [0.37, 133.22]	Low
Any AEs	1[a]	Not serious	N/A	Serious[3]	Serious[4]	RR 0.93 [0.70, 1.24]	Low
SAEs	1[a]	Not serious	N/A	Serious[3]	Serious[4]	RR 0.17 [0.02, 1.35]	Low
AE discontinuation	1[a]	Not serious	N/A	Serious[3]	Serious[4]	RR 3.00 [0.12, 72.49]	Low
a ESS 2007							
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2 >40%); [3]Population not as defined in protocol; [4]Non-significant result							

Network meta-analyses

OFF time (hours)

Quality assessment					
Number of RCTs	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Change in OFF time					
35 DAs vs. placebo n=19 COMTIs vs. placebo n=13 MAOBIs vs. placebo n=3	Not serious[1]	Not serious	Serious[2]	Not serious	Moderate
[1]Individual studies at risk of bias, but overall risk of bias rated low due to consistency of effect between studies at high and low risk of bias [2]Considered serious as population is not as defined in protocol					

UPDRS II (ADL)

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in UPDRS II score					
30 DA vs. placebo n=14 COMTIs vs. placebo n=12 Amantadine vs. placebo n=3 DA vs. COMTIs n=3	Not serious[1]	Serious[2]	Serious[3]	Not serious	Low

[1]Individual studies at risk of bias, but overall risk of bias rated low due to consistency of effect between studies at high and low risk of bias
[2]Considerable between study heterogeneity (I^2>40%)
[3]Considered serious as population is not as defined in protocol

UPDRS III (motor)

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in UPDRS III score					
34 DAs vs. placebo n=15 COMTIs vs. placebo n=13 MAOBIs vs. placebo n=2 Amantadine vs. placebo n=1 DAs vs. COMTIs n=3	Not serious[1]	Serious[2]	Serious[3]	Not serious	Low

[1]Individual studies at risk of bias, but overall risk of bias rated low due to consistency of effect between studies at high and low risk of bias
[2]Considerable between study heterogeneity (I^2>40%)
[3]Considered serious as population is not as defined in protocol

PDQ-39

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in PDQ-39 score					
4	Serious[1]	Serious[2]	Serious[3]	Not serious	Very Low

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
DA vs. placebo n=2 COMTIs vs. placebo n=1 DAs vs. COMTIs n=1					
[1]Individual studies at risk of bias [2]Considerable between study heterogeneity (I^2>40%) [3]Considered serious as population is not as defined in protocol					

Dyskinesia

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Dyskinesia					
65 DAs vs. placebo=29 COMTIs vs. placebo n=17 MAOBIs vs. placebo n=5 DAs vs. DAs n=11 DAs vs. COMTIs n=2 COMTI vs. COMTI n=1	Not serious[1]	Not serious	Serious[2]	Not serious	Moderate
[1]Individual studies at risk of bias, but overall risk of bias rated low due to consistency of effect between studies at high and low risk of bias [2]Considered serious as population is not as defined in protocol					

Hallucinations

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Hallucinations					
51 DA vs. placebo n=24 COMTIs vs. placebo n=12 MAOBIs vs. placebo =2 DA vs. DA n=10	Not serious[1]	Not serious	Serious[2]	Not serious	Moderate

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
DA vs. COMT n=2 COMT vs. COMT n=1					
[1]Individual studies at risk of bias, but overall risk of bias rated low due to consistency of effect between studies at high and low risk of bias [2]Considered serious as population is not as defined in protocol					

Mortality

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Mortality					
8 DAs vs. placebo n=6 COMTIs vs. placebo n=2	Not serious	Not serious	Serious[1]	Not serious	Moderate
[1]Considered serious as population is not as defined in protocol					

Serious adverse events (SAEs)

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
SAEs					
18 DAs vs. placebo n=9 COMTIs vs. placebo n=3 MAOBIs vs. placebo n=2 DAs vs. DAs n=2 COMTIs vs. COMTIs n=1 DA vs. COMTI n=1	Not serious[1]	Not serious	Serious[2]	Not serious	Moderate
[1]Individual studies at risk of bias, but overall risk of bias rated low due to consistency of effect between studies at high and low risk of bias [2]Considered serious as population is not as defined in protocol					

Any adverse events

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Any AEs					
51 DAs vs. placebo n=25 COMTIs vs. placebo n=14 MAOBIs vs. placebo n=6 DAs vs. DAs n=4 DA vs. COMTI n=1 COMTI vs. COMTI n=1	Not serious[1]	Not serious	Serious[2]	Not serious	Moderate
[1]Individual studies at risk of bias, but overall risk of bias rated low due to consistency of effect between studies at high and low risk of bias [2]Considered serious as population is not as defined in protocol					

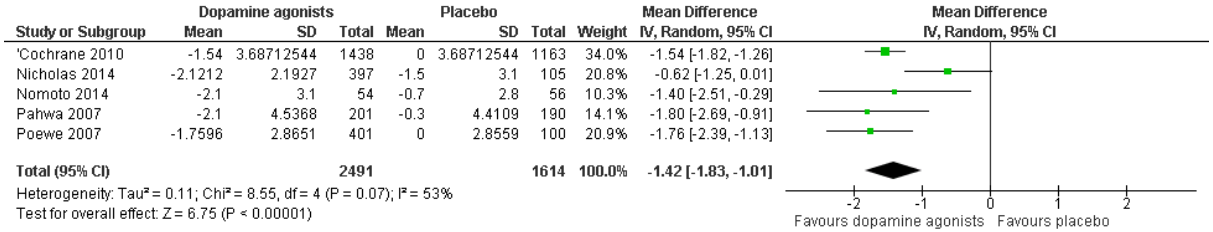
Adverse event discontinuations

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
AE discontinuations					
58 DAs vs. placebo n=29 COMTIs vs. placebo n=17 MAOBIs vs. placebo n=5 DAs vs. DAs n=4 DAs vs. COMTIs n=2 COMTI vs. COMTI n=1	Not serious[1]	Not serious	Serious[2]	Not serious	Moderate
[1]Individual studies at risk of bias, but overall risk of bias rated low due to consistency of effect between studies at high and low risk of bias [2]Considered serious as population is not as defined in protocol					

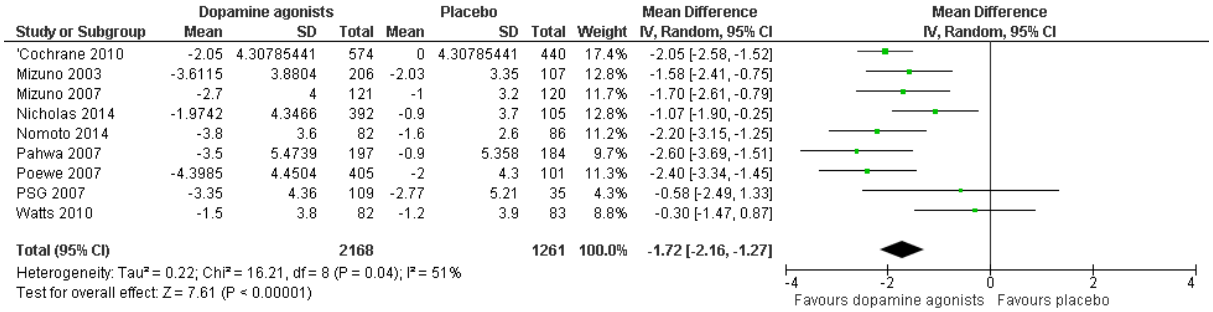
Pairwise meta-analyses

Dopamine agonists vs. Placebo

Off time



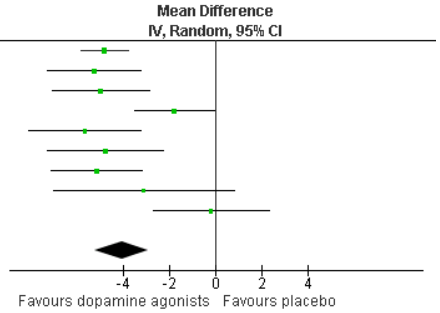
UPDRS II



UPDRS III

Study or Subgroup	Dopamine agonists Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
'Cochrane 2010	-4.86	9.8842617	774	0	9.8842617	629	16.0%	-4.86 [-5.90, -3.82]
Mizuno 2003	-10.8564	10.1089	206	-5.55	8.08	107	11.7%	-5.31 [-7.37, -3.24]
Mizuno 2007	-9.5	8.7	121	-4.5	7.9	120	11.5%	-5.00 [-7.10, -2.90]
Nicholas 2014	-4.3087	7.8591	393	-2.5	8.2	105	13.0%	-1.81 [-3.56, -0.06]
Nomoto 2014	-10.1	9	86	-4.4	7.4	86	10.1%	-5.70 [-8.16, -3.24]
Pahwa 2007	-6.5	12.6052	194	-1.7	12.3779	183	9.9%	-4.80 [-7.32, -2.28]
Poewe 2007	-9.4941	8.3474	405	-4.3	9.3	101	12.0%	-5.19 [-7.18, -3.21]
PSG 2007	-6.92	9.3	109	-3.77	10.66	35	5.9%	-3.15 [-7.09, 0.79]
Watts 2010	-3.7	9.3	83	-3.5	7	81	9.9%	-0.20 [-2.72, 2.32]
Total (95% CI)			**2371**			**1447**	**100.0%**	**-4.09 [-5.25, -2.92]**

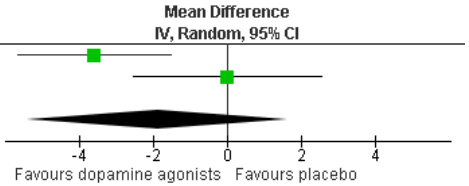
Heterogeneity: Tau² = 1.92; Chi² = 22.93, df = 8 (P = 0.003); I² = 65%
Test for overall effect: Z = 6.88 (P < 0.00001)



PDQ-39

Study or Subgroup	Dopamine agonists Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
Poewe 2007	-4.8985	9.6467	405	-1.3	9.4	101	52.3%	-3.60 [-5.66, -1.54]
Watts 2010	-2.4	7.2	66	-2.4	7.3	59	47.7%	0.00 [-2.55, 2.55]
Total (95% CI)			**471**			**160**	**100.0%**	**-1.88 [-5.40, 1.64]**

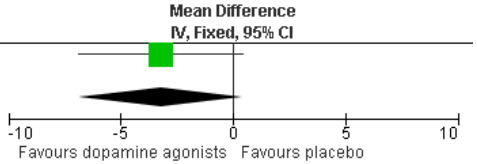
Heterogeneity: Tau² = 5.08; Chi² = 4.64, df = 1 (P = 0.03); I² = 78%
Test for overall effect: Z = 1.05 (P = 0.30)



PDQUALIF

Study or Subgroup	Dopamine agonists Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
PSG 2007	-5.41	11.75	109	-2.19	8.74	35	100.0%	-3.22 [-6.86, 0.42]
Total (95% CI)			**109**			**35**	**100.0%**	**-3.22 [-6.86, 0.42]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.73 (P = 0.08)

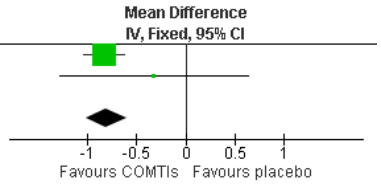


COMTIs vs. Placebo

Off time

Study or Subgroup	COMTIs Mean	COMTIs SD	COMTIs Total	Placebo Mean	Placebo SD	Placebo Total	Weight	Mean Difference IV, Fixed, 95% CI
'Cochrane 2010	-0.83	2.41805517	1138	0	2.41805517	922	95.4%	-0.83 [-1.04, -0.62]
Fenelon 2003	-0.8633	2.903	99	-0.53666	3.13167	63	4.6%	-0.33 [-1.29, 0.64]
Total (95% CI)			**1237**			**985**	**100.0%**	**-0.81 [-1.01, -0.60]**

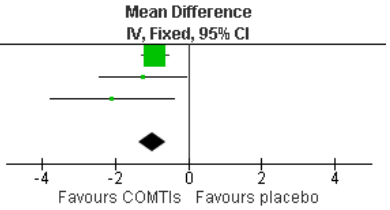
Heterogeneity: Chi² = 1.00, df = 1 (P = 0.32); I² = 0%
Test for overall effect: Z = 7.71 (P < 0.00001)



UPDRS II

Study or Subgroup	COMTIs Mean	COMTIs SD	COMTIs Total	Placebo Mean	Placebo SD	Placebo Total	Weight	Mean Difference IV, Fixed, 95% CI
'Cochrane 2010	-0.91	3.95430496	944	0	3.95430496	697	86.5%	-0.91 [-1.30, -0.52]
Fenelon 2003	-1.6	6.5133	198	-0.35	4.4573	126	9.1%	-1.25 [-2.45, -0.05]
Tolosa 2014	9.1	3.6	39	11.2	4.2	42	4.5%	-2.10 [-3.80, -0.40]
Total (95% CI)			**1181**			**865**	**100.0%**	**-0.99 [-1.35, -0.63]**

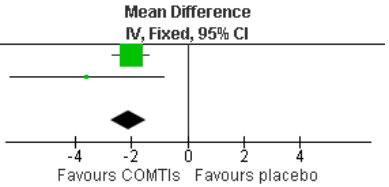
Heterogeneity: Chi² = 1.98, df = 2 (P = 0.37); I² = 0%
Test for overall effect: Z = 5.41 (P < 0.00001)



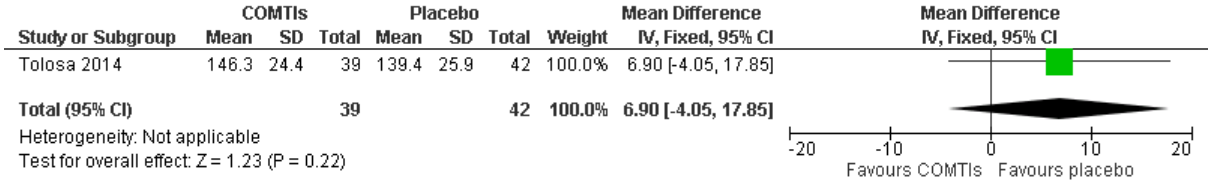
UPDRS III

Study or Subgroup	COMTIs Mean	COMTIs SD	COMTIs Total	Placebo Mean	Placebo SD	Placebo Total	Weight	Mean Difference IV, Fixed, 95% CI
'Cochrane 2010	-2.02	7.53757201	1135	0	7.53757201	922	94.5%	-2.02 [-2.67, -1.37]
Tolosa 2014	13.9	6.4	39	17.5	6.1	42	5.5%	-3.60 [-6.33, -0.87]
Total (95% CI)			**1174**			**964**	**100.0%**	**-2.11 [-2.74, -1.47]**

Heterogeneity: Chi² = 1.22, df = 1 (P = 0.27); I² = 18%
Test for overall effect: Z = 6.48 (P < 0.00001)

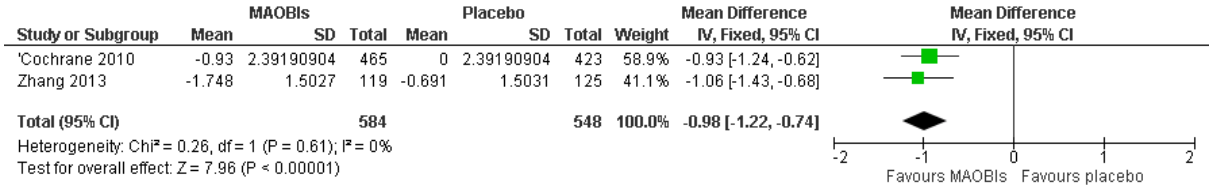


PDQ-39

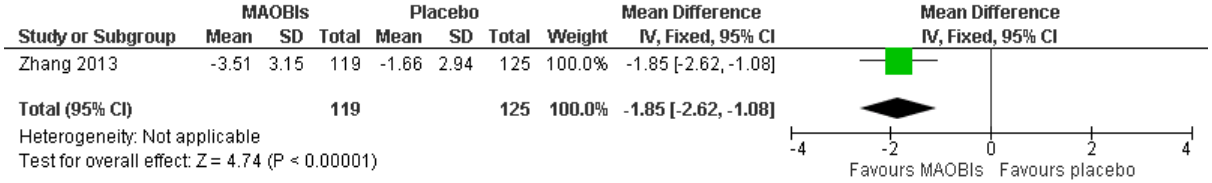


MAOBIs vs. Placebo

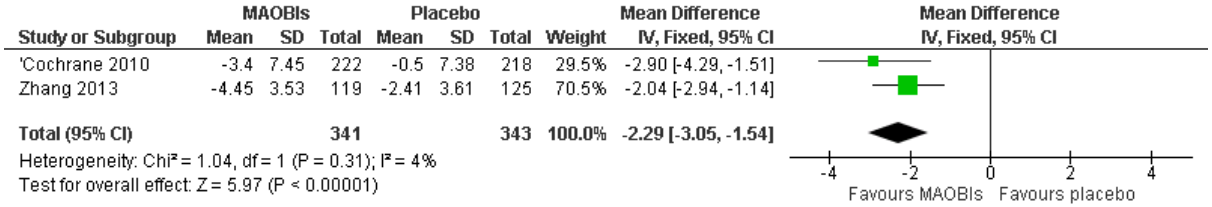
Off time



UPDRS II

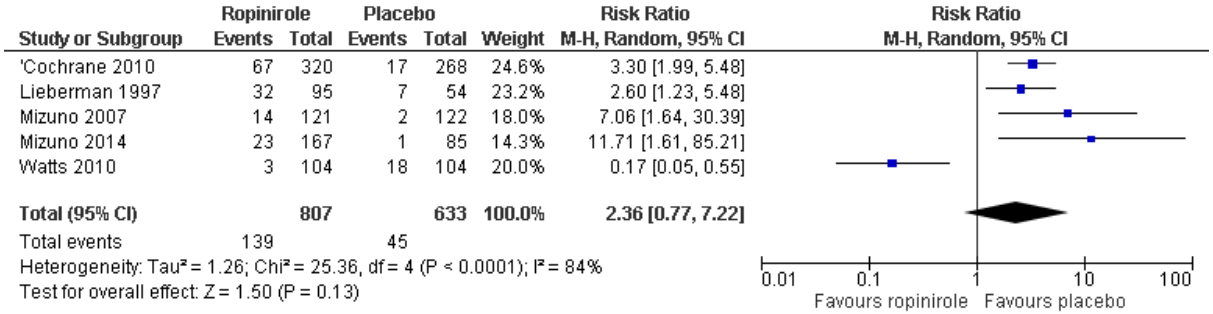


UPDRS III

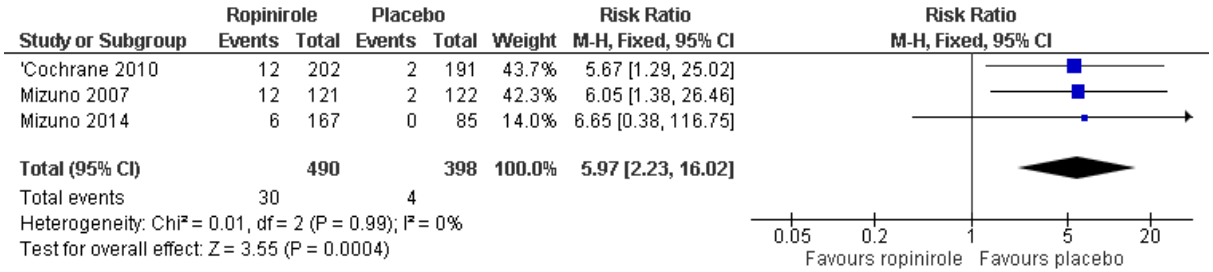


Ropinirole – Placebo

Dyskinesia



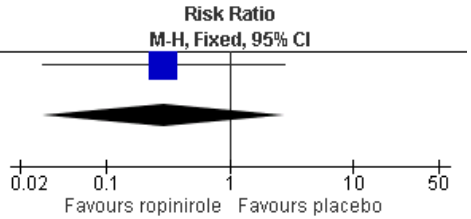
Hallucinations



Mortality

Study or Subgroup	Ropinirole Events	Ropinirole Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	1	271	3	236	100.0%	0.29 [0.03, 2.77]
Total (95% CI)		**271**		**236**	**100.0%**	**0.29 [0.03, 2.77]**
Total events	1		3			

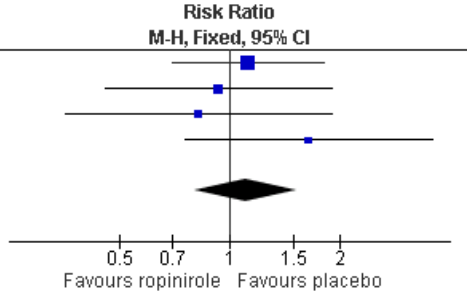
Heterogeneity: Not applicable
Test for overall effect: Z = 1.07 (P = 0.28)



AE discontinuation

Study or Subgroup	Ropinirole Events	Ropinirole Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	37	366	26	290	46.3%	1.13 [0.70, 1.82]
Mizuno 2007	13	121	14	122	22.3%	0.94 [0.46, 1.91]
Mizuno 2014	13	167	8	85	16.9%	0.83 [0.36, 1.92]
Watts 2010	15	105	9	104	14.4%	1.65 [0.76, 3.60]
Total (95% CI)		**759**		**601**	**100.0%**	**1.11 [0.80, 1.53]**
Total events	78		57			

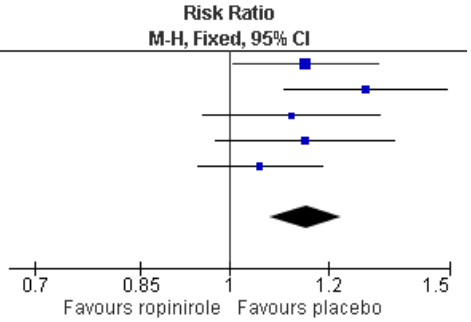
Heterogeneity: Chi² = 1.69, df = 3 (P = 0.64); I² = 0%
Test for overall effect: Z = 0.63 (P = 0.53)



Any AEs

Study or Subgroup	Ropinirole Events	Ropinirole Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	185	271	140	236	29.8%	1.15 [1.01, 1.31]
Mizuno 2007	102	121	80	122	15.9%	1.29 [1.11, 1.49]
Mizuno 2014	130	167	59	85	15.6%	1.12 [0.95, 1.32]
Pahwa 2007	129	202	106	191	21.7%	1.15 [0.98, 1.36]
Watts 2010	91	104	86	104	17.1%	1.06 [0.94, 1.19]
Total (95% CI)		**865**		**738**	**100.0%**	**1.15 [1.08, 1.23]**
Total events	637		471			

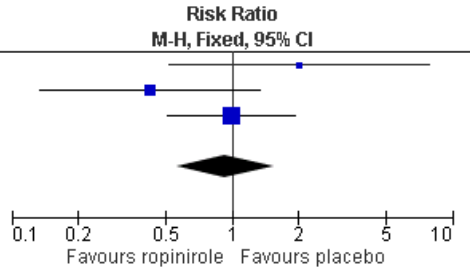
Heterogeneity: Chi² = 4.29, df = 4 (P = 0.37); I² = 7%
Test for overall effect: Z = 4.15 (P < 0.0001)



SAEs

Study or Subgroup	Ropinirole Events	Ropinirole Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Mizuno 2007	6	121	3	122	11.5%	2.02 [0.52, 7.88]
Mizuno 2014	5	167	6	85	30.6%	0.42 [0.13, 1.35]
Watts 2010	15	105	15	104	57.9%	0.99 [0.51, 1.92]
Total (95% CI)		**393**		**311**	**100.0%**	**0.94 [0.56, 1.57]**
Total events	26		24			

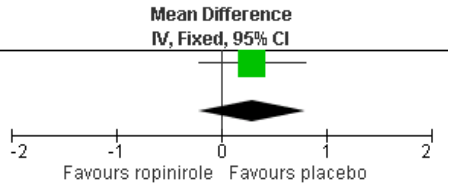
Heterogeneity: Chi² = 3.04, df = 2 (P = 0.22); I² = 34%
Test for overall effect: Z = 0.25 (P = 0.80)



Psychosis (PPRS)

Study or Subgroup	Ropinirole Mean	Ropinirole SD	Ropinirole Total	Placebo Mean	Placebo SD	Placebo Total	Weight	Mean Difference IV, Fixed, 95% CI
Watts 2010	0.3	1.7	83	0	1.6	82	100.0%	0.30 [-0.20, 0.80]
Total (95% CI)			**83**			**82**	**100.0%**	**0.30 [-0.20, 0.80]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.17 (P = 0.24)

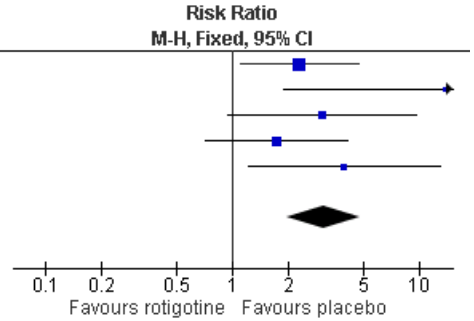


Rotigotine - Placebo

Dyskinesia

Study or Subgroup	Rotigotine Events	Rotigotine Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
LeWitt 2007	35	229	8	120	38.1%	2.29 [1.10, 4.78]
Mizuno 2014	27	168	1	85	4.8%	13.66 [1.89, 98.82]
Nicholas 2014	34	406	3	108	17.2%	3.01 [0.94, 9.63]
Nomoto 2014	12	87	7	87	25.4%	1.71 [0.71, 4.15]
Poewe 2007	24	204	3	101	14.6%	3.96 [1.22, 12.84]
Total (95% CI)		**1094**		**501**	**100.0%**	**3.06 [1.95, 4.81]**
Total events	132		22			

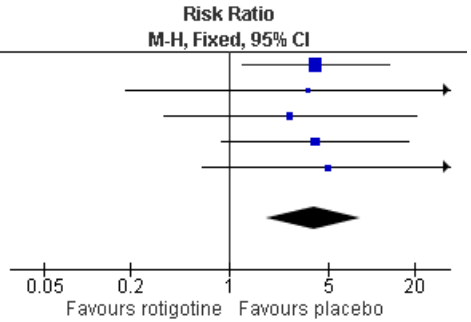
Heterogeneity: Chi² = 4.63, df = 4 (P = 0.33); I² = 14%
Test for overall effect: Z = 4.85 (P < 0.00001)



Hallucinations

Study or Subgroup	Rotigotine Events	Rotigotine Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
LeWitt 2007	23	229	3	120	41.4%	4.02 [1.23, 13.11]
Mizuno 2014	3	168	0	85	7.0%	3.56 [0.19, 68.18]
Nicholas 2014	10	406	1	108	16.6%	2.66 [0.34, 20.55]
Nomoto 2014	8	87	2	87	21.0%	4.00 [0.87, 18.30]
Poewe 2007	10	204	1	101	14.1%	4.95 [0.64, 38.14]
Total (95% CI)		**1094**		**501**	**100.0%**	**3.89 [1.82, 8.30]**
Total events	54		7			

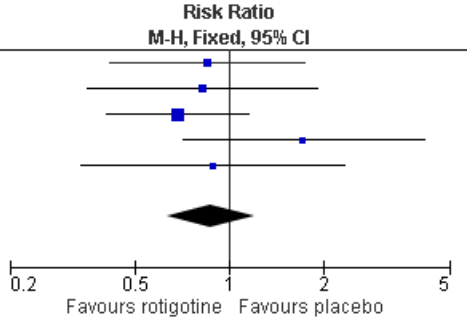
Heterogeneity: Chi² = 0.19, df = 4 (P = 1.00); I² = 0%
Test for overall effect: Z = 3.51 (P = 0.0004)



AE discontinuation

Study or Subgroup	Rotigotine Events	Rotigotine Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
LeWitt 2007	18	231	11	120	21.6%	0.85 [0.42, 1.74]
Mizuno 2014	13	168	8	85	15.8%	0.82 [0.35, 1.91]
Nicholas 2014	44	408	17	108	40.1%	0.69 [0.41, 1.15]
Nomoto 2014	12	87	7	87	10.4%	1.71 [0.71, 4.15]
Poewe 2007	11	205	6	99	12.1%	0.89 [0.34, 2.32]
Total (95% CI)		**1099**		**499**	**100.0%**	**0.87 [0.63, 1.21]**
Total events	98		49			

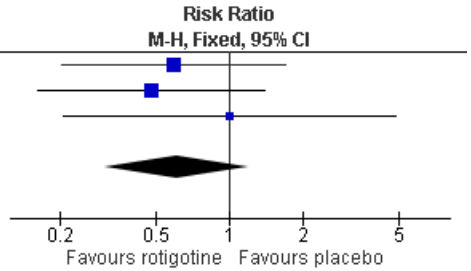
Heterogeneity: Chi² = 3.11, df = 4 (P = 0.54); I² = 0%
Test for overall effect: Z = 0.82 (P = 0.41)



SAEs

Study or Subgroup	Rotigotine Events	Rotigotine Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Mizuno 2014	7	168	6	85	42.2%	0.59 [0.20, 1.70]
Nicholas 2014	9	408	5	108	41.9%	0.48 [0.16, 1.39]
Nomoto 2014	3	87	3	87	15.9%	1.00 [0.21, 4.82]
Total (95% CI)		**663**		**280**	**100.0%**	**0.61 [0.31, 1.19]**
Total events	19		14			

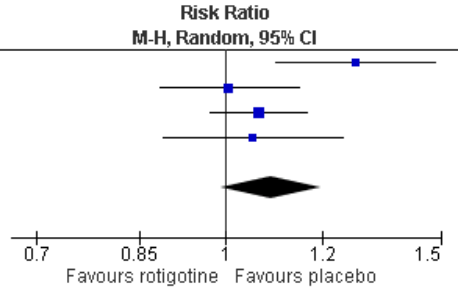
Heterogeneity: Chi² = 0.59, df = 2 (P = 0.75); I² = 0%
Test for overall effect: Z = 1.45 (P = 0.15)



Any AEs

Study or Subgroup	Rotigotine Events	Rotigotine Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Random, 95% CI
Mizuno 2014	149	168	59	85	21.8%	1.28 [1.10, 1.49]
Nicholas 2014	295	406	78	108	25.3%	1.01 [0.88, 1.15]
Nomoto 2014	82	87	77	87	33.9%	1.06 [0.97, 1.17]
Poewe 2007	141	204	65	99	19.0%	1.05 [0.89, 1.25]
Total (95% CI)		**865**		**379**	**100.0%**	**1.09 [0.99, 1.20]**
Total events	667		279			

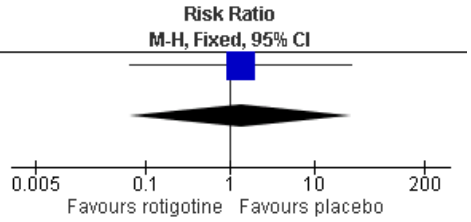
Heterogeneity: Tau² = 0.00; Chi² = 6.07, df = 3 (P = 0.11); I² = 51%
Test for overall effect: Z = 1.80 (P = 0.07)



Mortality

Study or Subgroup	Rotigotine Events	Rotigotine Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Nicholas 2014	2	406	0	108	100.0%	1.34 [0.06, 27.69]
Total (95% CI)		**406**		**108**	**100.0%**	**1.34 [0.06, 27.69]**
Total events	2		0			

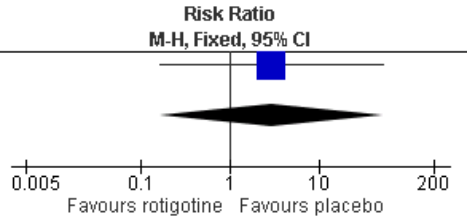
Heterogeneity: Not applicable
Test for overall effect: Z = 0.19 (P = 0.85)



ICD

Study or Subgroup	Rotigotine Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Nicholas 2014	5	408	0	108	100.0%	2.93 [0.16, 52.61]
Total (95% CI)		**408**		**108**	**100.0%**	**2.93 [0.16, 52.61]**
Total events	5		0			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.73 (P = 0.47)

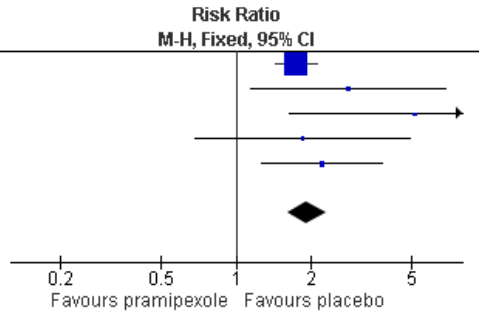


Pramipexole vs. Placebo

Dyskinesia

Study or Subgroup	Pramipexole Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	233	706	118	620	78.6%	1.73 [1.43, 2.10]
Mizuno 2003	16	102	6	108	3.6%	2.82 [1.15, 6.93]
Poewe 2007	31	201	3	101	2.5%	5.19 [1.63, 16.58]
PSG 2007	23	109	4	35	3.8%	1.85 [0.69, 4.97]
Schapira 2011	59	339	14	178	11.5%	2.21 [1.27, 3.85]
Total (95% CI)		**1457**		**1042**	**100.0%**	**1.92 [1.61, 2.29]**
Total events	362		145			

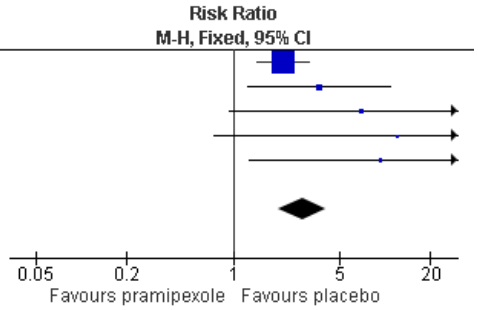
Heterogeneity: $Chi^2 = 4.85$, df = 4 (P = 0.30); $I^2 = 18\%$
Test for overall effect: Z = 7.29 (P < 0.00001)



Hallucinations

Study or Subgroup	Pramipexole Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	84	658	30	502	82.4%	2.14 [1.43, 3.19]
Mizuno 2003	14	102	4	108	9.4%	3.71 [1.26, 10.89]
Poewe 2007	14	201	1	101	3.2%	7.03 [0.94, 52.75]
PSG 2007	18	109	0	35	1.8%	12.11 [0.75, 195.91]
Schapira 2011	18	339	1	178	3.2%	9.45 [1.27, 70.22]
Total (95% CI)		**1409**		**924**	**100.0%**	**2.86 [1.99, 4.09]**
Total events	148		36			

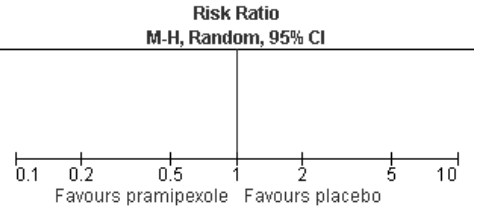
Heterogeneity: Chi² = 5.42, df = 4 (P = 0.25); I² = 26%
Test for overall effect: Z = 5.72 (P < 0.00001)



Mortality

Study or Subgroup	Pramipexole Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
'Cochrane 2010	0	120	0	131		Not estimable
Total (95% CI)		**120**		**131**		**Not estimable**
Total events	0		0			

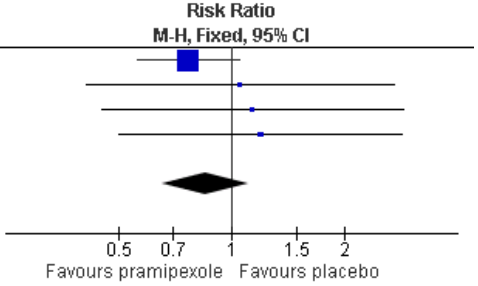
Heterogeneity: Not applicable
Test for overall effect: Not applicable



AE discontinuations

Study or Subgroup	Pramipexole Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	67	675	76	590	76.4%	0.77 [0.57, 1.05]
Mizuno 2003	8	102	8	108	7.3%	1.06 [0.41, 2.72]
Poewe 2007	14	202	6	99	7.6%	1.14 [0.45, 2.89]
Schapira 2011	16	339	7	178	8.7%	1.20 [0.50, 2.86]
Total (95% CI)		**1318**		**975**	**100.0%**	**0.86 [0.66, 1.12]**
Total events	105		97			

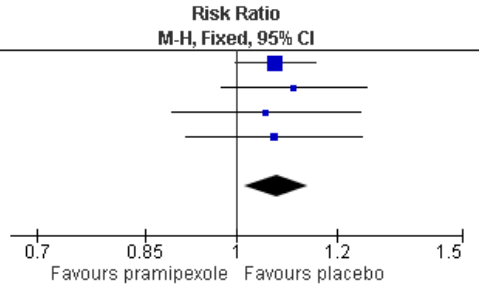
Heterogeneity: Chi² = 1.60, df = 3 (P = 0.66); I² = 0%
Test for overall effect: Z = 1.14 (P = 0.26)



Any AEs

Study or Subgroup	Pramipexole Events	Pramipexole Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	370	452	280	367	50.9%	1.07 [1.00, 1.15]
Mizuno 2003	87	102	83	108	13.3%	1.11 [0.97, 1.27]
Poewe 2007	140	202	65	99	14.4%	1.06 [0.89, 1.25]
Schapira 2011	202	339	99	178	21.4%	1.07 [0.91, 1.25]
Total (95% CI)		**1095**		**752**	**100.0%**	**1.08 [1.01, 1.14]**
Total events	799		527			

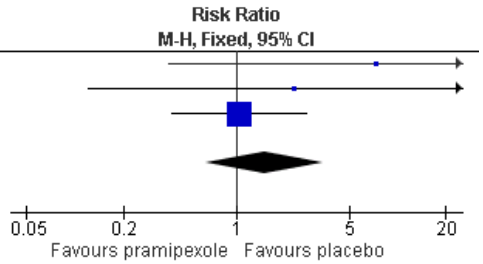
Heterogeneity: Chi² = 0.28, df = 3 (P = 0.96); I² = 0%
Test for overall effect: Z = 2.44 (P = 0.01)



SAEs

Study or Subgroup	Pramipexole Events	Pramipexole Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Mizuno 2003	3	102	0	107	5.4%	7.34 [0.38, 140.36]
PSG 2007	3	109	0	35	8.3%	2.29 [0.12, 43.30]
Schapira 2011	12	339	6	178	86.4%	1.05 [0.40, 2.75]
Total (95% CI)		**550**		**320**	**100.0%**	**1.49 [0.64, 3.44]**
Total events	18		6			

Heterogeneity: Chi² = 1.71, df = 2 (P = 0.43); I² = 0%
Test for overall effect: Z = 0.93 (P = 0.35)

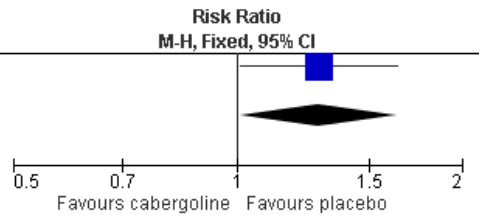


Cabergoline vs. Placebo

Dyskinesia

Study or Subgroup	Cabergoline Events	Cabergoline Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	135	291	57	158	100.0%	1.29 [1.01, 1.64]
Total (95% CI)		**291**		**158**	**100.0%**	**1.29 [1.01, 1.64]**
Total events	135		57			

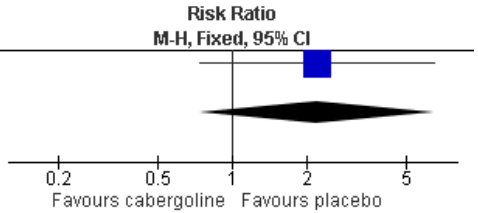
Heterogeneity: Not applicable
Test for overall effect: Z = 2.04 (P = 0.04)



Hallucinations

Study or Subgroup	Cabergoline Events	Cabergoline Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	14	165	4	103	100.0%	2.18 [0.74, 6.46]
Total (95% CI)		**165**		**103**	**100.0%**	**2.18 [0.74, 6.46]**
Total events	14		4			

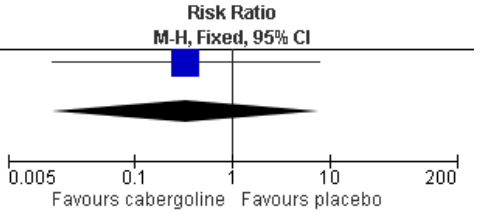
Heterogeneity: Not applicable
Test for overall effect: Z = 1.41 (P = 0.16)



Mortality

Study or Subgroup	Cabergoline Events	Cabergoline Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	0	20	1	20	100.0%	0.33 [0.01, 7.72]
Total (95% CI)		**20**		**20**	**100.0%**	**0.33 [0.01, 7.72]**
Total events	0		1			

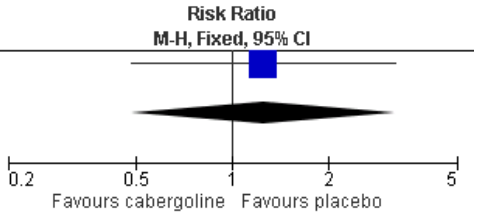
Heterogeneity: Not applicable
Test for overall effect: Z = 0.69 (P = 0.49)



AE discontinuation

Study or Subgroup	Cabergoline Events	Cabergoline Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	12	165	6	103	100.0%	1.25 [0.48, 3.22]
Total (95% CI)		**165**		**103**	**100.0%**	**1.25 [0.48, 3.22]**
Total events	12		6			

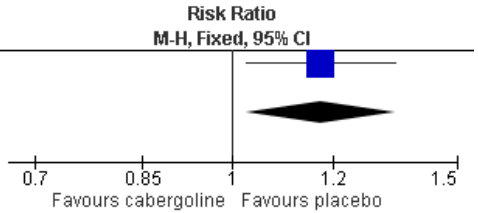
Heterogeneity: Not applicable
Test for overall effect: Z = 0.46 (P = 0.65)



Any AEs

Study or Subgroup	Cabergoline Events	Cabergoline Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	141	165	75	103	100.0%	1.17 [1.03, 1.34]
Total (95% CI)		**165**		**103**	**100.0%**	**1.17 [1.03, 1.34]**
Total events	141		75			

Heterogeneity: Not applicable
Test for overall effect: Z = 2.35 (P = 0.02)

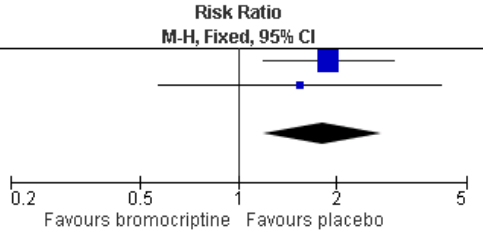


Bromocriptine vs. Placebo

Dyskinesia

Study or Subgroup	Bromocriptine Events	Bromocriptine Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	45	198	23	191	79.8%	1.89 [1.19, 2.99]
Mizuno 2003	9	105	6	108	20.2%	1.54 [0.57, 4.18]
Total (95% CI)		**303**		**299**	**100.0%**	**1.82 [1.20, 2.76]**
Total events	54		29			

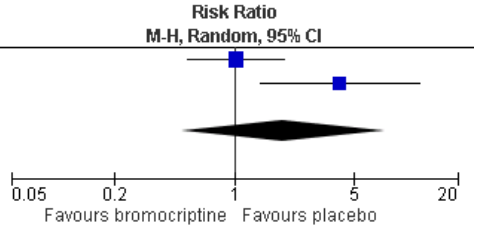
Heterogeneity: Chi² = 0.13, df = 1 (P = 0.72); I² = 0%
Test for overall effect: Z = 2.80 (P = 0.005)



Hallucinations

Study or Subgroup	Bromocriptine Events	Bromocriptine Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Random, 95% CI
'Cochrane 2010	17	198	16	191	54.6%	1.02 [0.53, 1.97]
Mizuno 2003	16	105	4	108	45.4%	4.11 [1.42, 11.90]
Total (95% CI)		**303**		**299**	**100.0%**	**1.93 [0.49, 7.56]**
Total events	33		20			

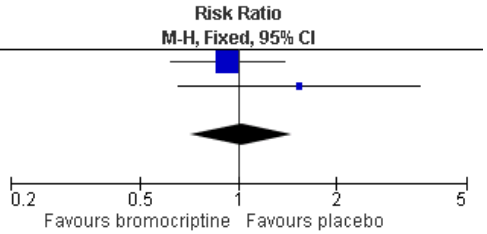
Heterogeneity: Tau² = 0.78; Chi² = 4.85, df = 1 (P = 0.03); I² = 79%
Test for overall effect: Z = 0.94 (P = 0.35)



AE discontinuation

Study or Subgroup	Bromocriptine Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	39	232	40	220	83.9%	0.92 [0.62, 1.38]
Mizuno 2003	12	105	8	108	16.1%	1.54 [0.66, 3.62]
Total (95% CI)		**337**		**328**	**100.0%**	**1.02 [0.71, 1.47]**
Total events	51		48			

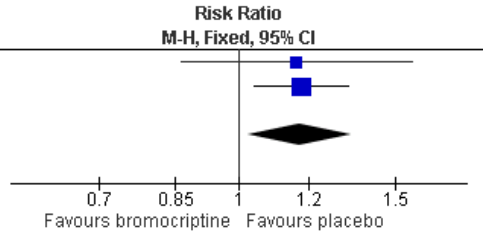
Heterogeneity: Chi² = 1.14, df = 1 (P = 0.29); I² = 12%
Test for overall effect: Z = 0.13 (P = 0.90)



Any AEs

Study or Subgroup	Bromocriptine Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	54	114	44	108	35.6%	1.16 [0.86, 1.57]
Mizuno 2003	95	105	83	108	64.4%	1.18 [1.04, 1.33]
Total (95% CI)		**219**		**216**	**100.0%**	**1.17 [1.03, 1.34]**
Total events	149		127			

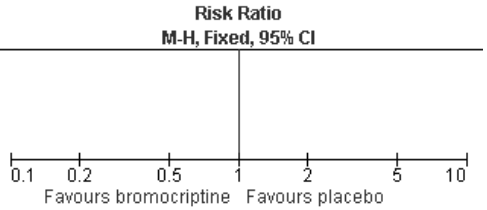
Heterogeneity: Chi² = 0.01, df = 1 (P = 0.93); I² = 0%
Test for overall effect: Z = 2.37 (P = 0.02)



SAEs

Study or Subgroup	Bromocriptine Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Mizuno 2003	0	104	0	107		Not estimable
Total (95% CI)		**104**		**107**		**Not estimable**
Total events	0		0			

Heterogeneity: Not applicable
Test for overall effect: Not applicable

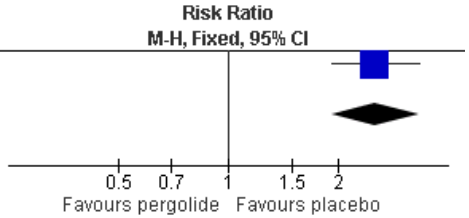


Pergolide vs. Placebo

Dyskinesia

Study or Subgroup	Pergolide Events	Pergolide Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	118	189	46	187	100.0%	2.54 [1.93, 3.34]
Total (95% CI)		**189**		**187**	**100.0%**	**2.54 [1.93, 3.34]**
Total events	118		46			

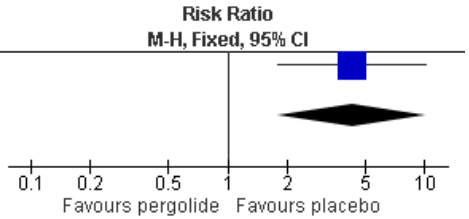
Heterogeneity: Not applicable
Test for overall effect: Z = 6.66 (P < 0.00001)



Hallucinations

Study or Subgroup	Pergolide Events	Pergolide Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	26	189	6	187	100.0%	4.29 [1.81, 10.18]
Total (95% CI)		**189**		**187**	**100.0%**	**4.29 [1.81, 10.18]**
Total events	26		6			

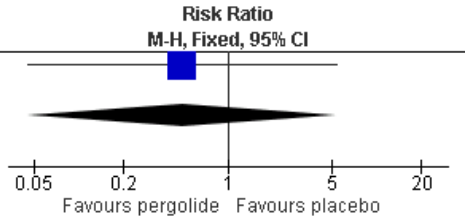
Heterogeneity: Not applicable
Test for overall effect: Z = 3.30 (P = 0.0010)



Mortality

Study or Subgroup	Pergolide Events	Pergolide Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	1	189	2	187	100.0%	0.49 [0.05, 5.41]
Total (95% CI)		**189**		**187**	**100.0%**	**0.49 [0.05, 5.41]**
Total events	1		2			

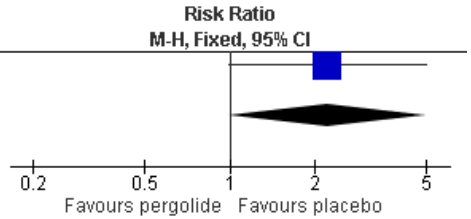
Heterogeneity: Not applicable
Test for overall effect: Z = 0.58 (P = 0.56)



AE discontinuation

Study or Subgroup	Pergolide Events	Pergolide Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	18	189	8	187	100.0%	2.23 [0.99, 4.99]
Total (95% CI)		**189**		**187**	**100.0%**	**2.23 [0.99, 4.99]**
Total events	18		8			

Heterogeneity: Not applicable
Test for overall effect: Z = 1.94 (P = 0.05)

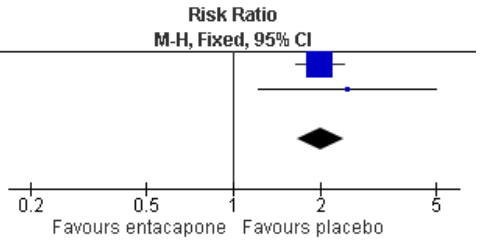


Entacapone vs. Placebo

Dyskinesia

Study or Subgroup	Entacapone Events	Entacapone Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	327	1430	122	1057	93.5%	1.98 [1.63, 2.40]
Fenelon 2003	31	99	8	63	6.5%	2.47 [1.21, 5.02]
Total (95% CI)		**1529**		**1120**	**100.0%**	**2.01 [1.67, 2.42]**
Total events	358		130			

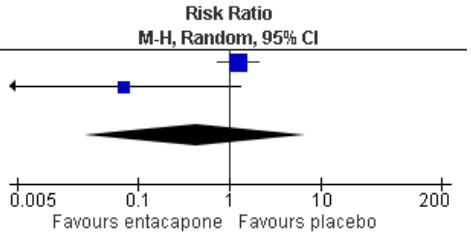
Heterogeneity: Chi² = 0.34, df = 1 (P = 0.56); I² = 0%
Test for overall effect: Z = 7.39 (P < 0.00001)



Hallucinations

Study or Subgroup	Entacapone Events	Entacapone Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Random, 95% CI
'Cochrane 2010	36	924	23	740	62.5%	1.25 [0.75, 2.10]
Fenelon 2003	0	99	4	63	37.5%	0.07 [0.00, 1.30]
Total (95% CI)		**1023**		**803**	**100.0%**	**0.43 [0.03, 6.84]**
Total events	36		27			

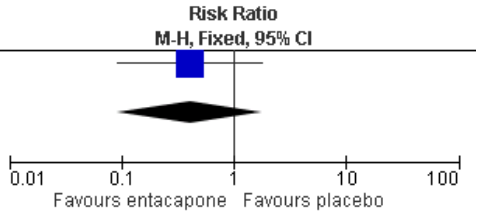
Heterogeneity: Tau² = 3.14; Chi² = 3.77, df = 1 (P = 0.05); I² = 73%
Test for overall effect: Z = 0.60 (P = 0.55)



Mortality

Study or Subgroup	Entacapone Events	Entacapone Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	3	317	4	171	100.0%	0.40 [0.09, 1.79]
Total (95% CI)		**317**		**171**	**100.0%**	**0.40 [0.09, 1.79]**
Total events	3		4			

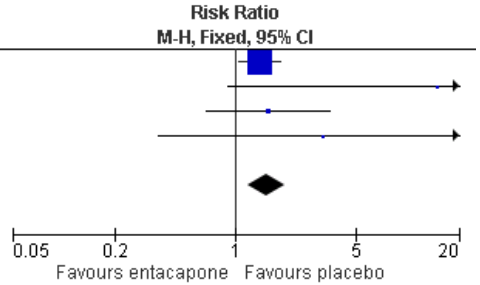
Heterogeneity: Not applicable
Test for overall effect: Z = 1.19 (P = 0.23)



AE discontinuation

Study or Subgroup	Entacapone Events	Entacapone Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	132	1232	71	917	88.9%	1.38 [1.05, 1.82]
Destee 2009	12	110	0	66	0.7%	15.09 [0.91, 250.74]
Fenelon 2003	17	99	7	63	9.3%	1.55 [0.68, 3.51]
Tolosa 2014	3	45	1	49	1.0%	3.27 [0.35, 30.28]
Total (95% CI)		**1486**		**1095**	**100.0%**	**1.51 [1.17, 1.95]**
Total events	164		79			

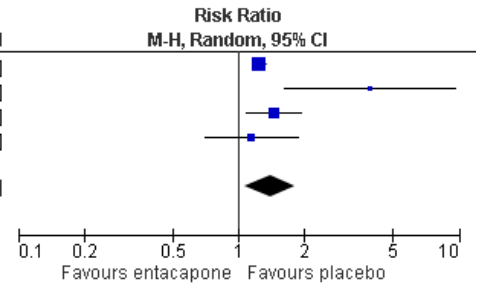
Heterogeneity: Chi² = 3.43, df = 3 (P = 0.33); I² = 13%
Test for overall effect: Z = 3.16 (P = 0.002)



Any AEs

Study or Subgroup	Entacapone Events	Entacapone Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Random, 95% CI
'Cochrane 2010	716	1029	431	764	44.9%	1.23 [1.15, 1.33]
Destee 2009	32	110	5	68	7.3%	3.96 [1.62, 9.66]
Fenelon 2003	68	99	30	63	29.6%	1.44 [1.08, 1.93]
Tolosa 2014	20	45	19	49	18.2%	1.15 [0.71, 1.85]
Total (95% CI)		**1283**		**944**	**100.0%**	**1.39 [1.07, 1.81]**
Total events	836		485			

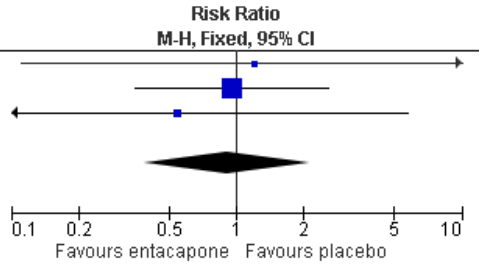
Heterogeneity: Tau² = 0.04; Chi² = 7.82, df = 3 (P = 0.05); I² = 62%
Test for overall effect: Z = 2.45 (P = 0.01)



SAEs

Study or Subgroup	Entacapone Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Destee 2009	2	110	1	66	11.9%	1.20 [0.11, 12.98]
Fenelon 2003	9	99	6	63	69.9%	0.95 [0.36, 2.55]
Tolosa 2014	1	45	2	49	18.2%	0.54 [0.05, 5.80]
Total (95% CI)		**254**		**178**	**100.0%**	**0.91 [0.39, 2.12]**
Total events	12		9			

Heterogeneity: Chi² = 0.24, df = 2 (P = 0.89); I² = 0%
Test for overall effect: Z = 0.22 (P = 0.82)

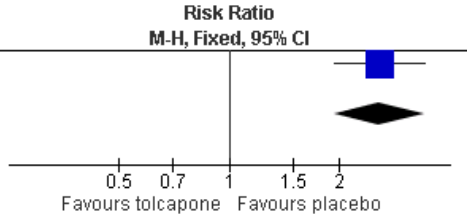


Tolcapone vs. Placebo

Dyskinesia

Study or Subgroup	Tolcapone Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	125	288	49	291	100.0%	2.58 [1.93, 3.44]
Total (95% CI)		**288**		**291**	**100.0%**	**2.58 [1.93, 3.44]**
Total events	125		49			

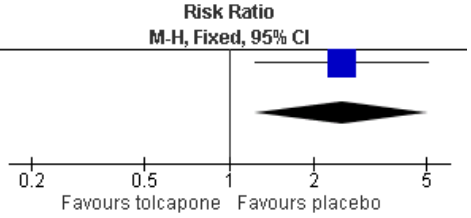
Heterogeneity: Not applicable
Test for overall effect: Z = 6.46 (P < 0.00001)



Hallucinations

Study or Subgroup	Tolcapone Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	20	155	11	213	100.0%	2.50 [1.23, 5.06]
Total (95% CI)		**155**		**213**	**100.0%**	**2.50 [1.23, 5.06]**
Total events	20		11			

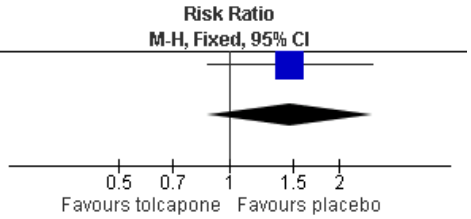
Heterogeneity: Not applicable
Test for overall effect: Z = 2.54 (P = 0.01)



AE discontinuation

Study or Subgroup	Tolcapone Events	Tolcapone Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	32	268	22	271	100.0%	1.47 [0.88, 2.46]
Total (95% CI)		**268**		**271**	**100.0%**	**1.47 [0.88, 2.46]**
Total events	32		22			

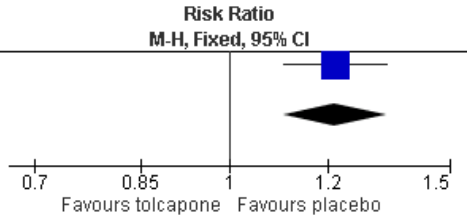
Heterogeneity: Not applicable
Test for overall effect: Z = 1.47 (P = 0.14)



Any AEs

Study or Subgroup	Tolcapone Events	Tolcapone Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	184	208	155	213	100.0%	1.22 [1.10, 1.34]
Total (95% CI)		**208**		**213**	**100.0%**	**1.22 [1.10, 1.34]**
Total events	184		155			

Heterogeneity: Not applicable
Test for overall effect: Z = 4.00 (P < 0.0001)

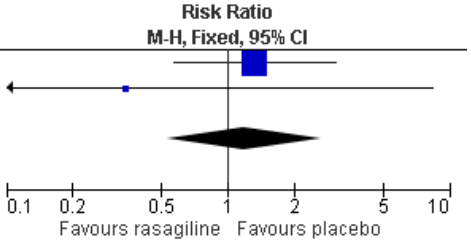


Rasagiline vs. Placebo

Dyskinesia

Study or Subgroup	Rasagiline Events	Rasagiline Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	12	231	9	229	86.0%	1.32 [0.57, 3.08]
Zhang 2013	0	119	1	124	14.0%	0.35 [0.01, 8.44]
Total (95% CI)		**350**		**353**	**100.0%**	**1.19 [0.53, 2.65]**
Total events	12		10			

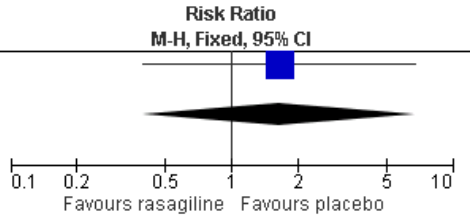
Heterogeneity: $Chi^2 = 0.63$, df = 1 (P = 0.43); $I^2 = 0\%$
Test for overall effect: Z = 0.41 (P = 0.68)



Hallucinations

Study or Subgroup	Rasagiline Events	Rasagiline Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	5	231	3	229	100.0%	1.65 [0.40, 6.83]
Total (95% CI)		**231**		**229**	**100.0%**	**1.65 [0.40, 6.83]**
Total events	5		3			

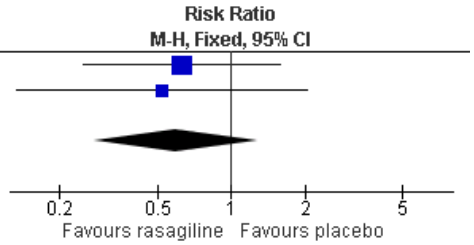
Heterogeneity: Not applicable
Test for overall effect: Z = 0.69 (P = 0.49)



AE discontinuation

Study or Subgroup	Rasagiline Events	Rasagiline Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	7	231	11	229	65.4%	0.63 [0.25, 1.60]
Zhang 2013	3	119	6	125	34.6%	0.53 [0.13, 2.05]
Total (95% CI)		**350**		**354**	**100.0%**	**0.59 [0.28, 1.28]**
Total events	10		17			

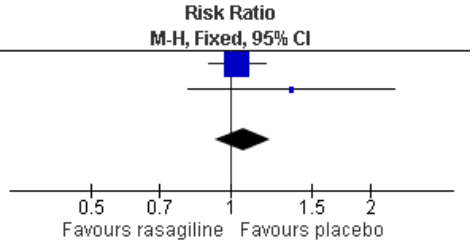
Heterogeneity: $Chi^2 = 0.05$, df = 1 (P = 0.83); $I^2 = 0\%$
Test for overall effect: Z = 1.33 (P = 0.18)



Any AEs

Study or Subgroup	Rasagiline Events	Rasagiline Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	193	380	191	388	90.2%	1.03 [0.90, 1.19]
Zhang 2013	27	119	21	125	9.8%	1.35 [0.81, 2.25]
Total (95% CI)		**499**		**513**	**100.0%**	**1.06 [0.93, 1.22]**
Total events	220		212			

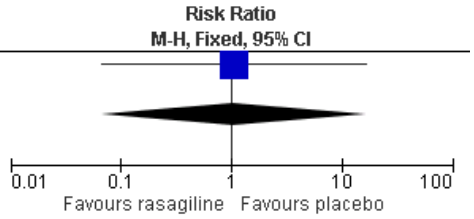
Heterogeneity: $Chi^2 = 1.01$, df = 1 (P = 0.31); $I^2 = 1\%$
Test for overall effect: Z = 0.87 (P = 0.39)



SAEs

Study or Subgroup	Rasagiline Events	Rasagiline Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Zhang 2013	1	119	1	125	100.0%	1.05 [0.07, 16.60]
Total (95% CI)		**119**		**125**	**100.0%**	**1.05 [0.07, 16.60]**
Total events	1		1			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.03 (P = 0.97)

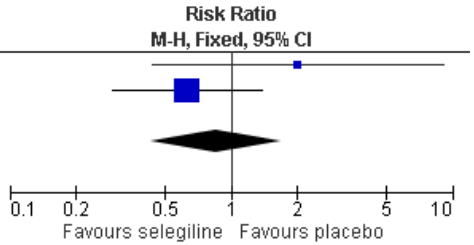


Selegiline vs. Placebo

Dyskinesia

Study or Subgroup	Selegiline Events	Selegiline Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Ondo 2007	8	100	2	50	16.4%	2.00 [0.44, 9.07]
'Cochrane 2010	9	70	13	64	83.6%	0.63 [0.29, 1.38]
Total (95% CI)		**170**		**114**	**100.0%**	**0.86 [0.44, 1.69]**
Total events	17		15			

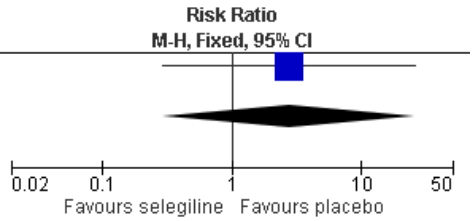
Heterogeneity: Chi² = 1.79, df = 1 (P = 0.18); I² = 44%
Test for overall effect: Z = 0.45 (P = 0.66)



Hallucinations

Study or Subgroup	Selegiline Events	Selegiline Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	3	50	1	46	100.0%	2.76 [0.30, 25.60]
Total (95% CI)		**50**		**46**	**100.0%**	**2.76 [0.30, 25.60]**
Total events	3		1			

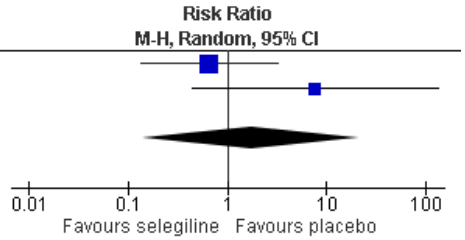
Heterogeneity: Not applicable
Test for overall effect: Z = 0.89 (P = 0.37)



AE discontinuation

Study or Subgroup	Selegiline Events	Selegiline Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Random, 95% CI
'Cochrane 2010	3	144	3	95	60.7%	0.66 [0.14, 3.20]
Ondo 2007	7	100	0	50	39.3%	7.57 [0.44, 130.01]
Total (95% CI)		**244**		**145**	**100.0%**	**1.72 [0.14, 20.91]**
Total events	10		3			

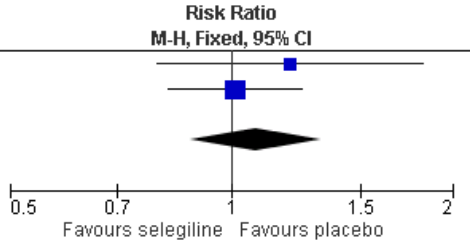
Heterogeneity: Tau² = 2.02; Chi² = 2.47, df = 1 (P = 0.12); I² = 60%
Test for overall effect: Z = 0.43 (P = 0.67)



Any AEs

Study or Subgroup	Selegiline Events	Selegiline Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	45	114	21	64	35.9%	1.20 [0.79, 1.83]
Ondo 2007	73	100	36	50	64.1%	1.01 [0.82, 1.25]
Total (95% CI)		**214**		**114**	**100.0%**	**1.08 [0.88, 1.33]**
Total events	118		57			

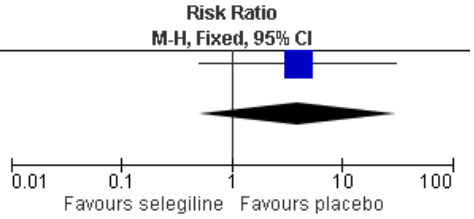
Heterogeneity: Chi² = 0.61, df = 1 (P = 0.43); I² = 0%
Test for overall effect: Z = 0.75 (P = 0.45)



SAEs

Study or Subgroup	Selegiline Events	Selegiline Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Ondo 2007	8	100	1	50	100.0%	4.00 [0.51, 31.10]
Total (95% CI)		**100**		**50**	**100.0%**	**4.00 [0.51, 31.10]**
Total events	8		1			

Heterogeneity: Not applicable
Test for overall effect: Z = 1.32 (P = 0.19)

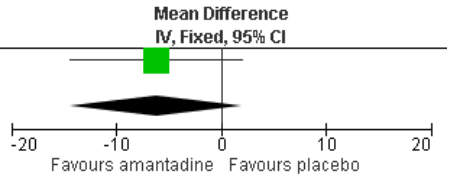


Amantadine vs. Placebo

Hyperkinesia (CDRS)

Study or Subgroup	Amantadine Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
da Silva-Junior 2005	6.8	4.9	9	13	11.5	9	100.0%	-6.20 [-14.37, 1.97]
Total (95% CI)			**9**			**9**	**100.0%**	**-6.20 [-14.37, 1.97]**

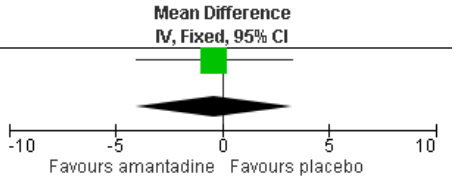
Heterogeneity: Not applicable
Test for overall effect: Z = 1.49 (P = 0.14)



Dystonia (CDRS)

Study or Subgroup	Amantadine Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
da Silva-Junior 2005	4.1	3.6	9	4.5	4.3	9	100.0%	-0.40 [-4.06, 3.26]
Total (95% CI)			**9**			**9**	**100.0%**	**-0.40 [-4.06, 3.26]**

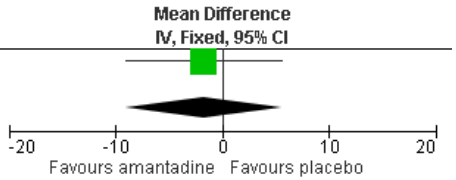
Heterogeneity: Not applicable
Test for overall effect: Z = 0.21 (P = 0.83)



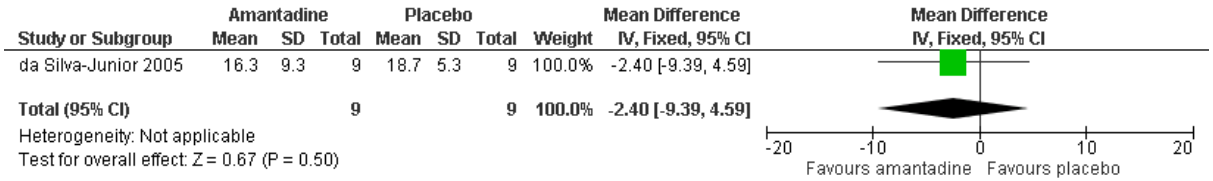
UPDRS II

Study or Subgroup	Amantadine Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
da Silva-Junior 2005	12.1	8	9	13.8	7.9	9	100.0%	-1.70 [-9.05, 5.65]
Total (95% CI)			**9**			**9**	**100.0%**	**-1.70 [-9.05, 5.65]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.45 (P = 0.65)

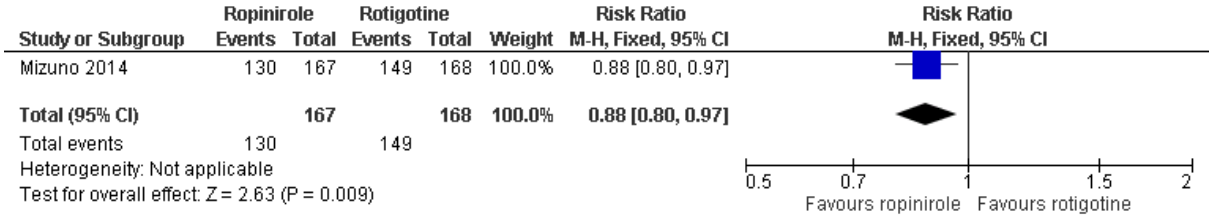


UPDRS III

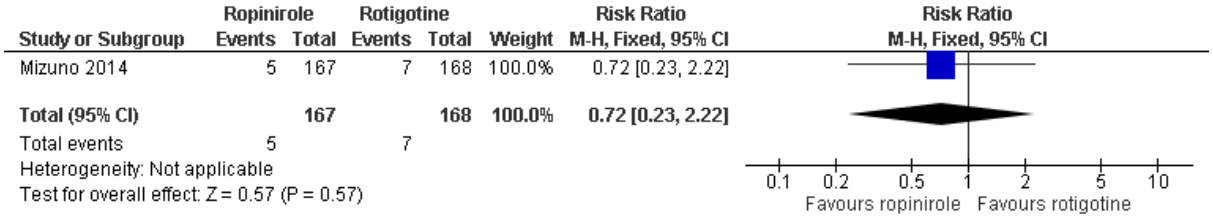


Ropinirole vs. Rotigotine

Any AEs



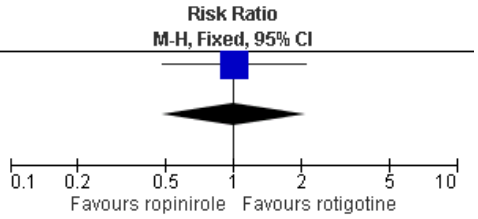
SAEs



AE discontinuation

Study or Subgroup	Ropinirole Events	Ropinirole Total	Rotigotine Events	Rotigotine Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Mizuno 2014	13	167	13	168	100.0%	1.01 [0.48, 2.10]
Total (95% CI)		**167**		**168**	**100.0%**	**1.01 [0.48, 2.10]**
Total events	13		13			

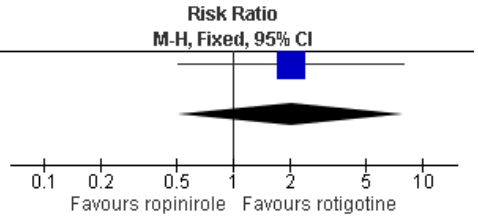
Heterogeneity: Not applicable
Test for overall effect: Z = 0.02 (P = 0.99)



Hallucinations

Study or Subgroup	Ropinirole Events	Ropinirole Total	Rotigotine Events	Rotigotine Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Mizuno 2014	6	167	3	168	100.0%	2.01 [0.51, 7.91]
Total (95% CI)		**167**		**168**	**100.0%**	**2.01 [0.51, 7.91]**
Total events	6		3			

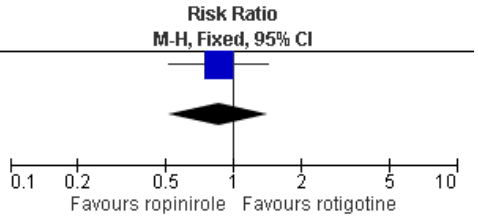
Heterogeneity: Not applicable
Test for overall effect: Z = 1.00 (P = 0.32)



Dyskinesia

Study or Subgroup	Ropinirole Events	Ropinirole Total	Rotigotine Events	Rotigotine Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Mizuno 2014	23	167	27	168	100.0%	0.86 [0.51, 1.43]
Total (95% CI)		**167**		**168**	**100.0%**	**0.86 [0.51, 1.43]**
Total events	23		27			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.59 (P = 0.56)

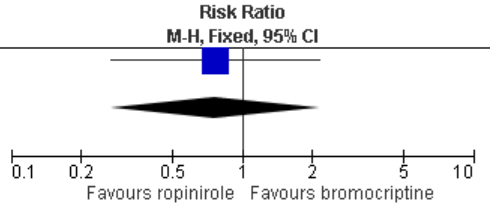


Ropinirole vs. Bromocriptine

Hallucinations

Study or Subgroup	Ropinirole Events	Ropinirole Total	Bromocriptine Events	Bromocriptine Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Clarke 2001 (B)	6	163	8	166	100.0%	0.76 [0.27, 2.15]
Total (95% CI)		**163**		**166**	**100.0%**	**0.76 [0.27, 2.15]**
Total events	6		8			

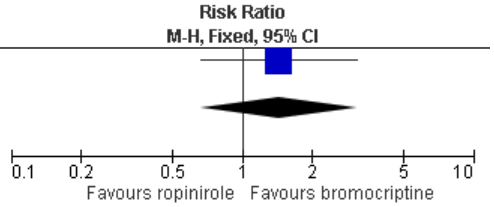
Heterogeneity: Not applicable
Test for overall effect: Z = 0.51 (P = 0.61)



Dyskinesia

Study or Subgroup	Ropinirole Events	Ropinirole Total	Bromocriptine Events	Bromocriptine Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Clarke 2001 (B)	14	169	10	174	100.0%	1.44 [0.66, 3.16]
Total (95% CI)		**169**		**174**	**100.0%**	**1.44 [0.66, 3.16]**
Total events	14		10			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.91 (P = 0.36)

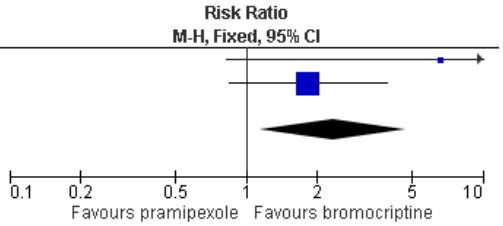


Pramipexole vs. Bromocriptine

Dyskinesia

Study or Subgroup	Pramipexole Events	Pramipexole Total	Bromocriptine Events	Bromocriptine Total	Weight	Risk Ratio M-H, Fixed, 95% CI
'Cochrane 2010	7	114	1	108	10.4%	6.63 [0.83, 53.01]
Mizuno 2003	16	102	9	105	89.6%	1.83 [0.85, 3.95]
Total (95% CI)		**216**		**213**	**100.0%**	**2.33 [1.14, 4.74]**
Total events	23		10			

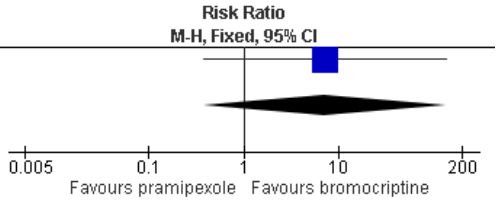
Heterogeneity: Chi² = 1.35, df = 1 (P = 0.25); I² = 26%
Test for overall effect: Z = 2.33 (P = 0.02)



SAEs

Study or Subgroup	Pramipexole Events	Pramipexole Total	Bromocriptine Events	Bromocriptine Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Mizuno 2003	3	102	0	104	100.0%	7.14 [0.37, 136.43]
Total (95% CI)		**102**		**104**	**100.0%**	**7.14 [0.37, 136.43]**
Total events	3		0			

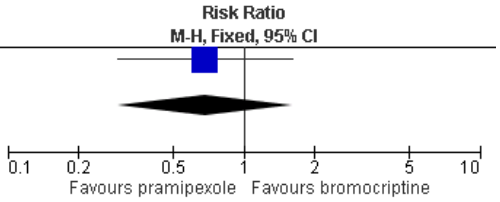
Heterogeneity: Not applicable
Test for overall effect: Z = 1.31 (P = 0.19)



AE discontinuation

Study or Subgroup	Pramipexole Events	Pramipexole Total	Bromocriptine Events	Bromocriptine Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Mizuno 2003	8	102	12	105	100.0%	0.69 [0.29, 1.61]
Total (95% CI)		**102**		**105**	**100.0%**	**0.69 [0.29, 1.61]**
Total events	8		12			

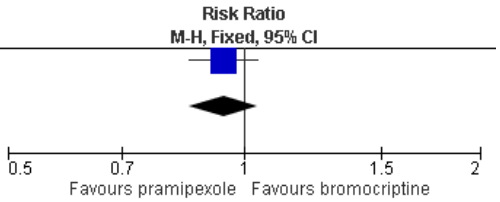
Heterogeneity: Not applicable
Test for overall effect: Z = 0.87 (P = 0.39)



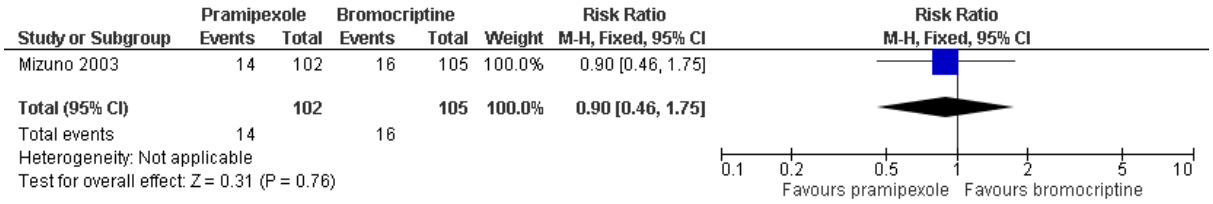
Any AEs

Study or Subgroup	Pramipexole Events	Pramipexole Total	Bromocriptine Events	Bromocriptine Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Mizuno 2003	87	102	95	105	100.0%	0.94 [0.85, 1.04]
Total (95% CI)		**102**		**105**	**100.0%**	**0.94 [0.85, 1.04]**
Total events	87		95			

Heterogeneity: Not applicable
Test for overall effect: Z = 1.14 (P = 0.26)

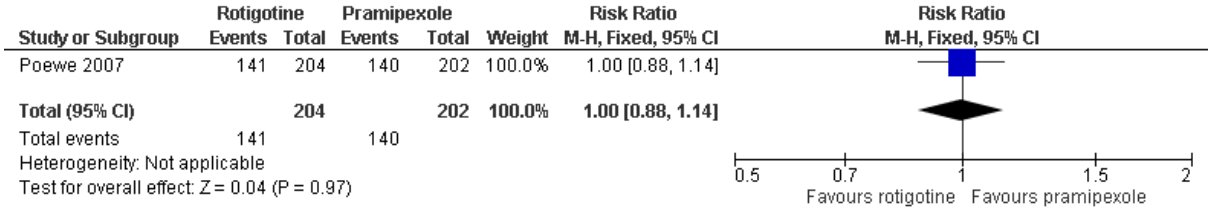


Hallucinations

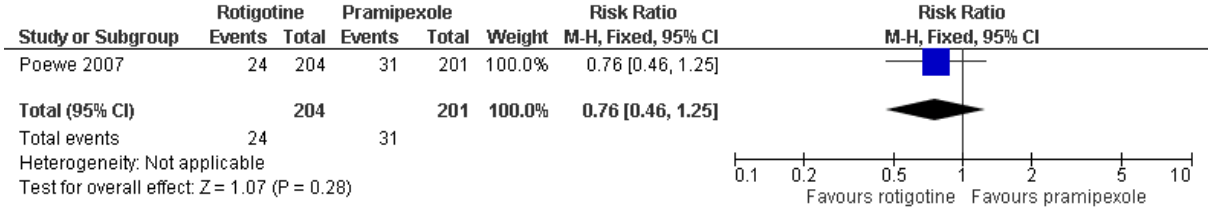


Rotigotine vs. Pramipexole

Any AEs



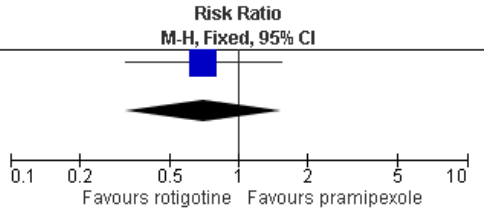
Dyskinesia



Hallucinations

Study or Subgroup	Rotigotine Events	Rotigotine Total	Pramipexole Events	Pramipexole Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Poewe 2007	10	204	14	201	100.0%	0.70 [0.32, 1.55]
Total (95% CI)		**204**		**201**	**100.0%**	**0.70 [0.32, 1.55]**
Total events	10		14			

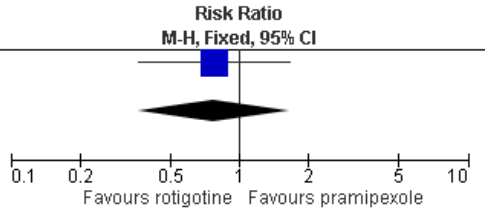
Heterogeneity: Not applicable
Test for overall effect: Z = 0.87 (P = 0.38)



AE discontinuation

Study or Subgroup	Rotigotine Events	Rotigotine Total	Pramipexole Events	Pramipexole Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Poewe 2007	11	205	14	202	100.0%	0.77 [0.36, 1.66]
Total (95% CI)		**205**		**202**	**100.0%**	**0.77 [0.36, 1.66]**
Total events	11		14			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.66 (P = 0.51)

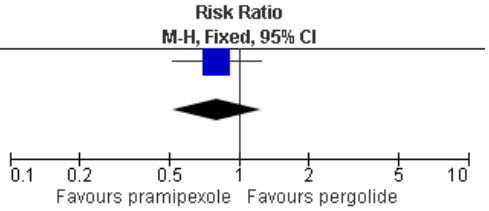


Pramipexole vs. Pergolide

Any AEs

Study or Subgroup	Pramipexole Events	Pramipexole Total	Pergolide Events	Pergolide Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Rektorova 2003	13	22	14	19	100.0%	0.80 [0.52, 1.24]
Total (95% CI)		**22**		**19**	**100.0%**	**0.80 [0.52, 1.24]**
Total events	13		14			

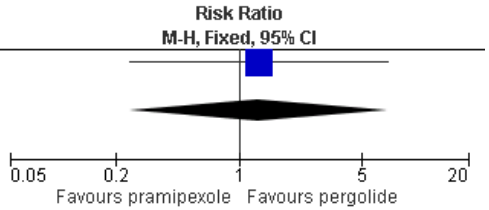
Heterogeneity: Not applicable
Test for overall effect: Z = 0.98 (P = 0.32)



AE discontinuation

Study or Subgroup	Pramipexole Events	Total	Pergolide Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Rektorova 2003	3	22	2	19	100.0%	1.30 [0.24, 6.96]
Total (95% CI)		**22**		**19**	**100.0%**	**1.30 [0.24, 6.96]**
Total events	3		2			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.30 (P = 0.76)

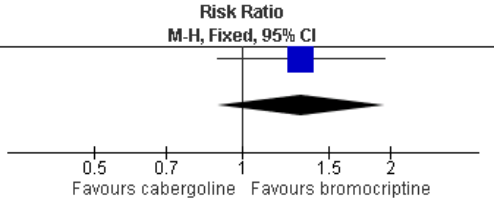


Cabergoline vs. Bromocriptine

Hallucinations

Study or Subgroup	Cabergoline Events	Total	Bromocriptine Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Clarke 2001 (A)	53	520	41	528	100.0%	1.31 [0.89, 1.94]
Total (95% CI)		**520**		**528**	**100.0%**	**1.31 [0.89, 1.94]**
Total events	53		41			

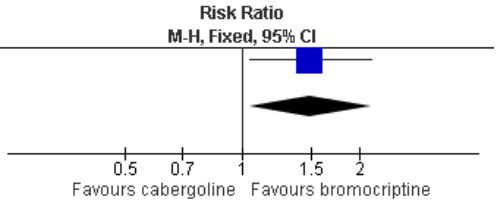
Heterogeneity: Not applicable
Test for overall effect: Z = 1.37 (P = 0.17)



Dyskinesia

Study or Subgroup	Cabergoline Events	Total	Bromocriptine Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Clarke 2001 (A)	66	520	45	528	100.0%	1.49 [1.04, 2.13]
Total (95% CI)		**520**		**528**	**100.0%**	**1.49 [1.04, 2.13]**
Total events	66		45			

Heterogeneity: Not applicable
Test for overall effect: Z = 2.17 (P = 0.03)

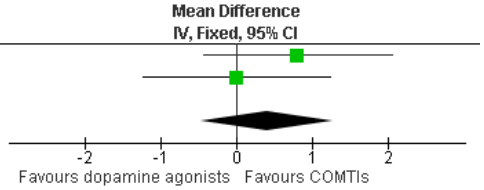


Dopamine Agonists vs. COMTIs

UPDRS II

	Dopamine agonists			COMTIs				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
Deane 2004	-0.1	3.4	74	-0.9	4.2	72	49.8%	0.80 [-0.44, 2.04]
Deuschl 2007	-2.5	3.9	69	-2.5	3.5	69	50.2%	0.00 [-1.24, 1.24]
Total (95% CI)			**143**			**141**	**100.0%**	**0.40 [-0.48, 1.27]**

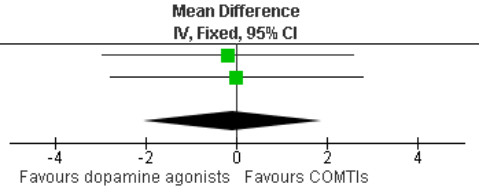
Heterogeneity: Chi² = 0.80, df = 1 (P = 0.37); I² = 0%
Test for overall effect: Z = 0.89 (P = 0.37)



UPDRS III

	Dopamine agonists			COMTIs				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
Deane 2004	-3.3	8.6	74	-3.1	8.5	72	50.0%	-0.20 [-2.97, 2.57]
Deuschl 2007	-6.3	7.9	69	-6.3	8.7	69	50.0%	0.00 [-2.77, 2.77]
Total (95% CI)			**143**			**141**	**100.0%**	**-0.10 [-2.06, 1.86]**

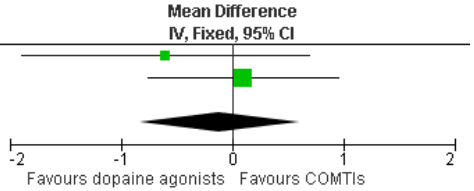
Heterogeneity: Chi² = 0.01, df = 1 (P = 0.92); I² = 0%
Test for overall effect: Z = 0.10 (P = 0.92)



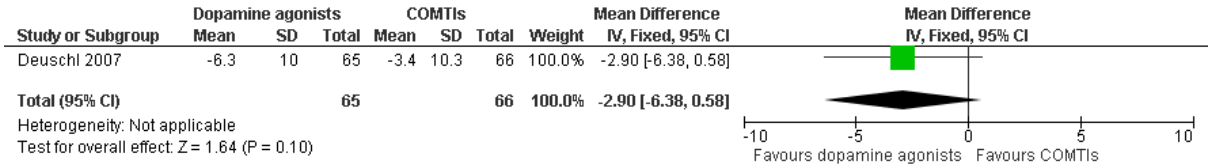
Off time

	Dopamine agonists			COMTIs				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
Deane 2004	2.4	3.9	74	3	4.1	72	30.5%	-0.60 [-1.90, 0.70]
Deuschl 2007	-1.7	2.4	64	-1.8	2.7	71	69.5%	0.10 [-0.76, 0.96]
Total (95% CI)			**138**			**143**	**100.0%**	**-0.11 [-0.83, 0.60]**

Heterogeneity: Chi² = 0.78, df = 1 (P = 0.38); I² = 0%
Test for overall effect: Z = 0.31 (P = 0.76)

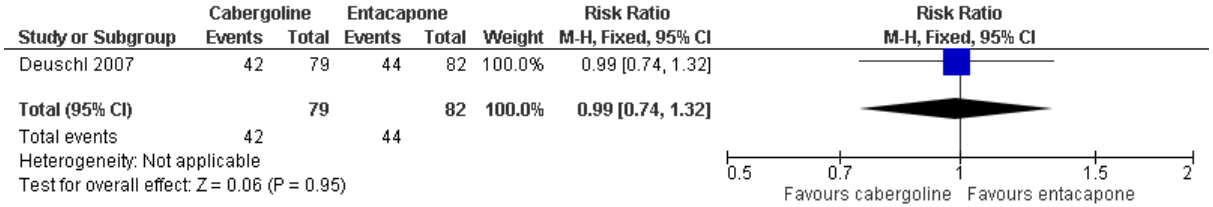


PDQ-39

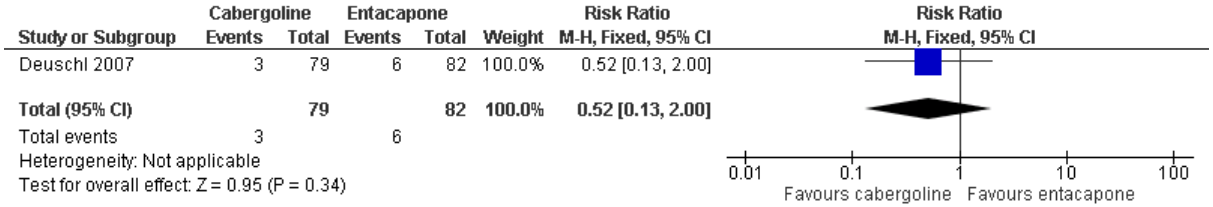


Cabergoline vs. Entacapone

Any AEs



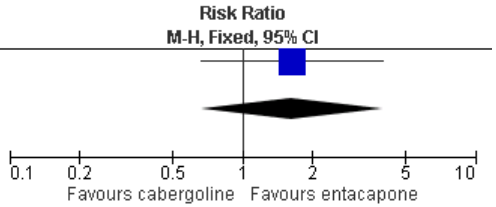
SAEs



AE discontinuation

Study or Subgroup	Cabergoline Events	Cabergoline Total	Entacapone Events	Entacapone Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Deuschl 2007	11	79	7	82	100.0%	1.63 [0.67, 4.00]
Total (95% CI)		**79**		**82**	**100.0%**	**1.63 [0.67, 4.00]**
Total events	11		7			

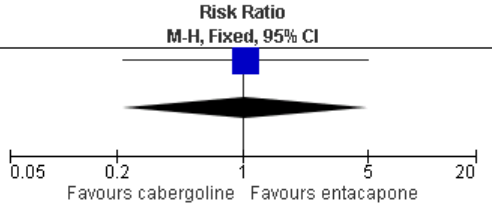
Heterogeneity: Not applicable
Test for overall effect: Z = 1.07 (P = 0.28)



Hallucinations

Study or Subgroup	Cabergoline Events	Cabergoline Total	Entacapone Events	Entacapone Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Deuschl 2007	3	79	3	82	100.0%	1.04 [0.22, 4.99]
Total (95% CI)		**79**		**82**	**100.0%**	**1.04 [0.22, 4.99]**
Total events	3		3			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.05 (P = 0.96)

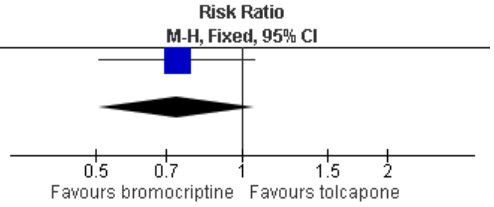


Bromocriptine vs. Tolcapone

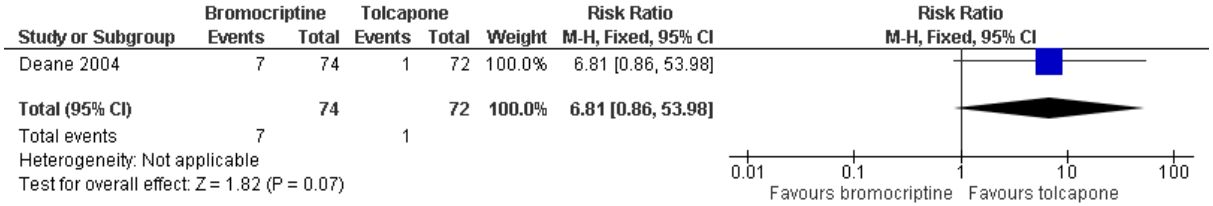
Dyskinesia

Study or Subgroup	Bromocriptine Events	Bromocriptine Total	Tolcapone Events	Tolcapone Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Deane 2004	28	74	37	72	100.0%	0.74 [0.51, 1.06]
Total (95% CI)		**74**		**72**	**100.0%**	**0.74 [0.51, 1.06]**
Total events	28		37			

Heterogeneity: Not applicable
Test for overall effect: Z = 1.63 (P = 0.10)

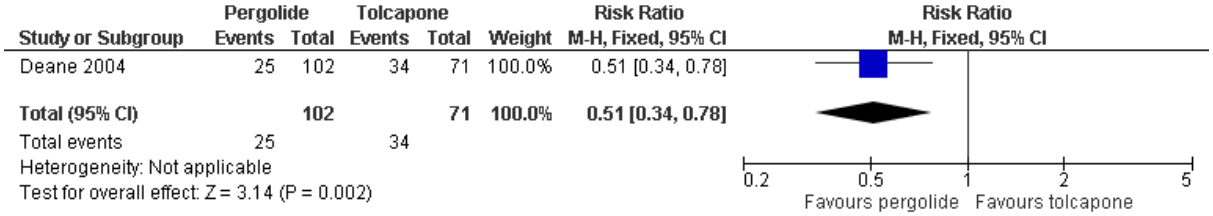


Hallucinations

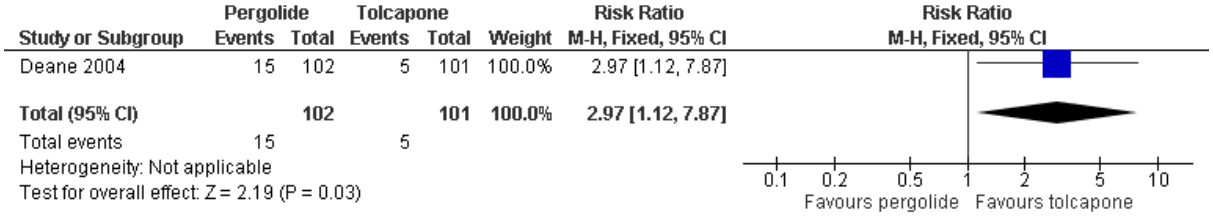


Pergolide vs. Tolcapone

Dyskinesia



AE discontinuation

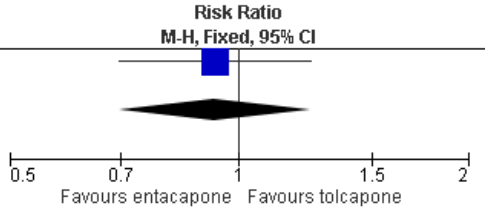


Entacapone vs. Tolcapone

Any AEs

Study or Subgroup	Entacapone Events	Entacapone Total	Tolcapone Events	Tolcapone Total	Weight	Risk Ratio M-H, Fixed, 95% CI
ESS 2007	40	75	43	75	100.0%	0.93 [0.70, 1.24]
Total (95% CI)		**75**		**75**	**100.0%**	**0.93 [0.70, 1.24]**
Total events	40		43			

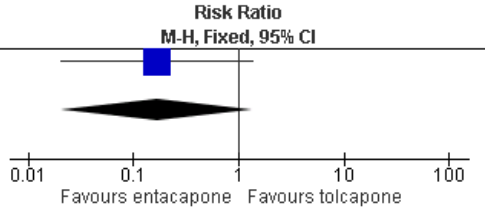
Heterogeneity: Not applicable
Test for overall effect: Z = 0.49 (P = 0.62)



SAEs

Study or Subgroup	Entacapone Events	Entacapone Total	Tolcapone Events	Tolcapone Total	Weight	Risk Ratio M-H, Fixed, 95% CI
ESS 2007	1	75	6	75	100.0%	0.17 [0.02, 1.35]
Total (95% CI)		**75**		**75**	**100.0%**	**0.17 [0.02, 1.35]**
Total events	1		6			

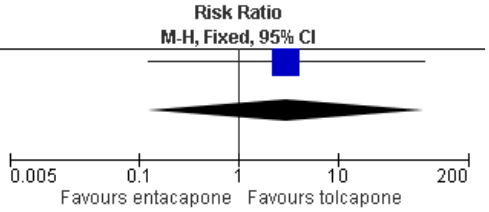
Heterogeneity: Not applicable
Test for overall effect: Z = 1.68 (P = 0.09)



AE discontinuation

Study or Subgroup	Entacapone Events	Entacapone Total	Tolcapone Events	Tolcapone Total	Weight	Risk Ratio M-H, Fixed, 95% CI
ESS 2007	1	75	0	75	100.0%	3.00 [0.12, 72.49]
Total (95% CI)		**75**		**75**	**100.0%**	**3.00 [0.12, 72.49]**
Total events	1		0			

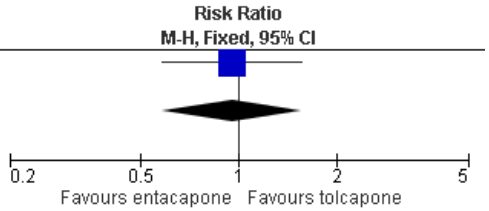
Heterogeneity: Not applicable
Test for overall effect: Z = 0.68 (P = 0.50)



Dyskinesia

Study or Subgroup	Entacapone Events	Entacapone Total	Tolcapone Events	Tolcapone Total	Weight	Risk Ratio M-H, Fixed, 95% CI
ESS 2007	22	75	23	75	100.0%	0.96 [0.59, 1.56]
Total (95% CI)		**75**		**75**	**100.0%**	**0.96 [0.59, 1.56]**
Total events	22		23			

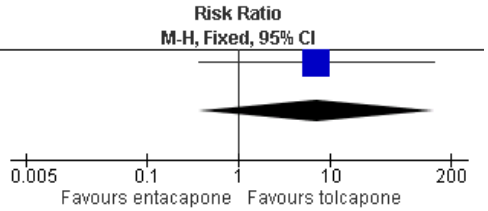
Heterogeneity: Not applicable
Test for overall effect: Z = 0.18 (P = 0.86)



Hallucinations

Study or Subgroup	Entacapone Events	Entacapone Total	Tolcapone Events	Tolcapone Total	Weight	Risk Ratio M-H, Fixed, 95% CI
ESS 2007	3	75	0	75	100.0%	7.00 [0.37, 133.22]
Total (95% CI)		**75**		**75**	**100.0%**	**7.00 [0.37, 133.22]**
Total events	3		0			

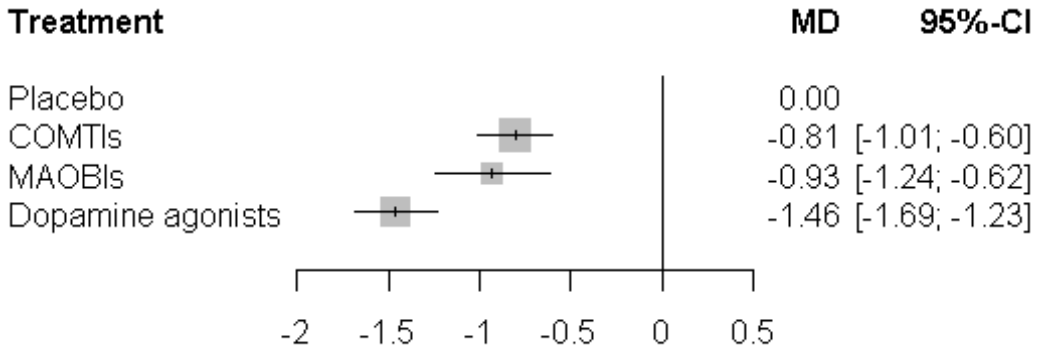
Heterogeneity: Not applicable
Test for overall effect: Z = 1.29 (P = 0.20)



Network meta-analyses

Efficacy outcomes by drug classes

Off time (hours) – FE model



Quantifying heterogeneity/inconsistency:

tau^2 = 0.0914; I^2 = 47.7%

Test of heterogeneity/inconsistency:

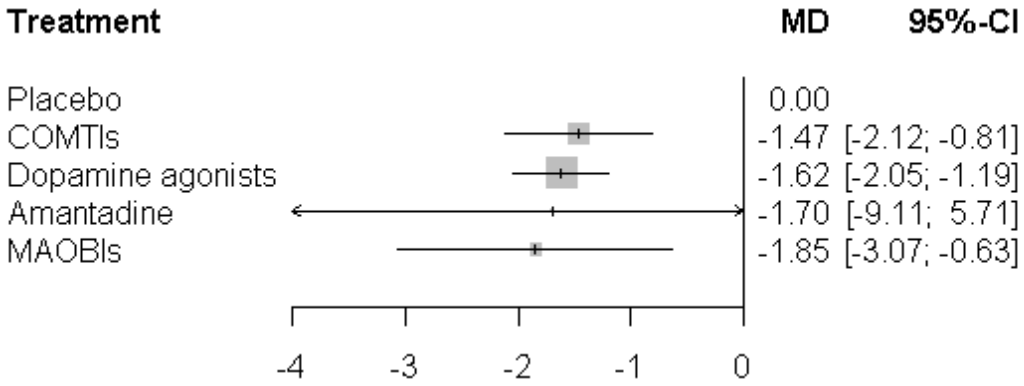
Q	d.f.	p.value
9.55	5	0.089

Differences between treatments – mean and 95% confidence interval

	Treatment A				
Treatment B		Placebo	COMTIs	MAOBIs	Dopamine agonists

	Treatment A				
	Placebo	N/A			
	COMTIs	-0.81 (-1.01, -0.60)	N/A		
	MAOBIs	-0.93 (-1.25, -0.62)	-0.12 (-0.50, 0.25)	N/A	
	Dopamine agonists	-1.46 (-1.69, -1.23)	-0.65 (-0.96, -0.35)	-0.53 (-0.92, -0.14)	N/A

UPDRS II (ADL) – RE model



Quantifying heterogeneity/inconsistency:

tau^2 = 0.2352; I^2 = 50.9%

Test of heterogeneity/inconsistency:

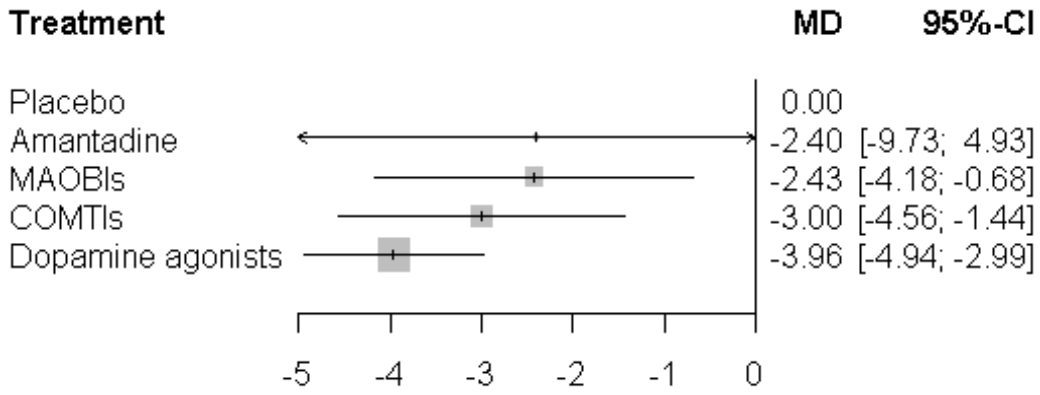
Q	d.f.	p.value
24.45	12	0.0176

Differences between treatments – mean and 95% confidence interval

	Treatment A					
Treatment B		Placebo	COMTIs	Dopamine agonists	Amantadine	MAOBIs
	Placebo	N/A				
	COMTIs	-1.47 (-2.12, -0.81)	N/A			
	Dopamine agonists	-1.62 (-2.05, -1.19)	-0.15 (-0.85, 0.54)	N/A		
	Amantadine	-1.70 (-9.11, 5.71)	-0.23 (-7.67, 7.20)	-0.08 (-7.50, 7.34)	N/A	

	Treatment A					
	MAOBIs	-1.85 (-3.07, -0.63)	-0.38 (-1.77, 1.00)	-0.23 (-1.52, 1.06)	-0.15 (-7.66, 7.36)	N/A

UPDRS III (motor) – RE model



Quantifying heterogeneity/inconsistency:

tau^2 = 1.2468; I^2 = 58.2%

Test of heterogeneity/inconsistency:

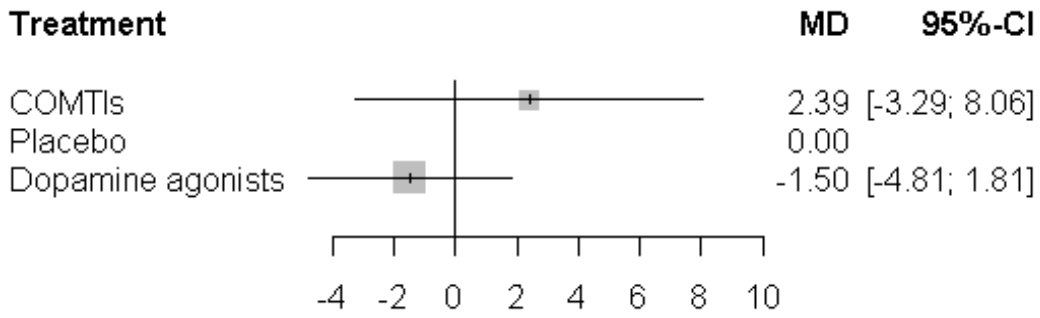
Q	d.f.	p.value
28.71	12	0.0044

Differences between treatments – mean and 95% confidence interval

	Treatment A					
Treatment B		Placebo	Amantadine	MAOBIs	COMTIs	Dopamine agonists
	Placebo	N/A				
	Amantadine	-2.40 (-9.73, 4.93)	N/A			
	MAOBIs	-2.43 (-4.18, -0.68)	-0.03 (-7.56, 7.50)	N/A		
	COMTIs	-3.00	-0.60	-0.57	N/A	

	Treatment A					
		(4.56, -1.44)	(-8.09, 6.89)	(-2.91, 1.77)		
	Dopamine agonists	-3.96 (-4.94, -2.99)	-1.56 (-8.95, 5.83)	-1.53 (-3.53, 0.47)	-0.96 (-2.60, 0.67)	N/A

PDQ-39 – RE model



Quantifying heterogeneity/inconsistency:

tau^2 = 4.7260; I^2 = 65.1%

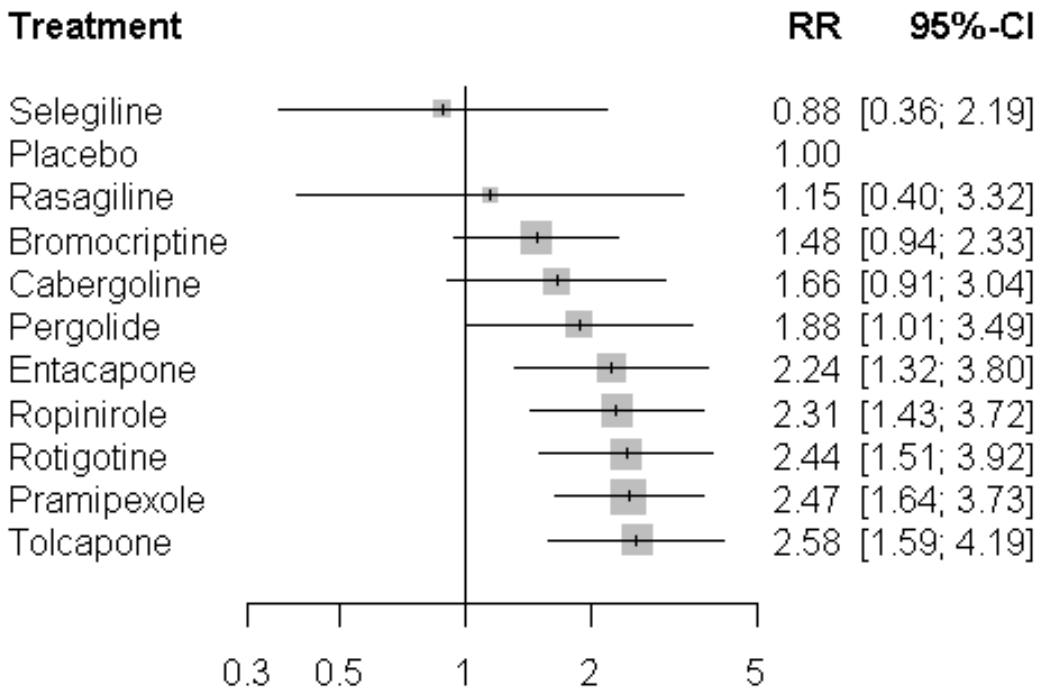
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
5.72	2	0.0572

Differences between treatments – mean and 95% confidence interval

	Treatment A			
Treatment B		COMTIs	Placebo	Dopamine agonists
	COMTIs	N/A		
	Placebo	-2.39 (-8.06, 3.29)	N/A	
	Dopamine agonists	-3.89 (-8.90, 1.13)	-1.50 (-4.81, 1.81)	N/A

Dyskinesia – RE model

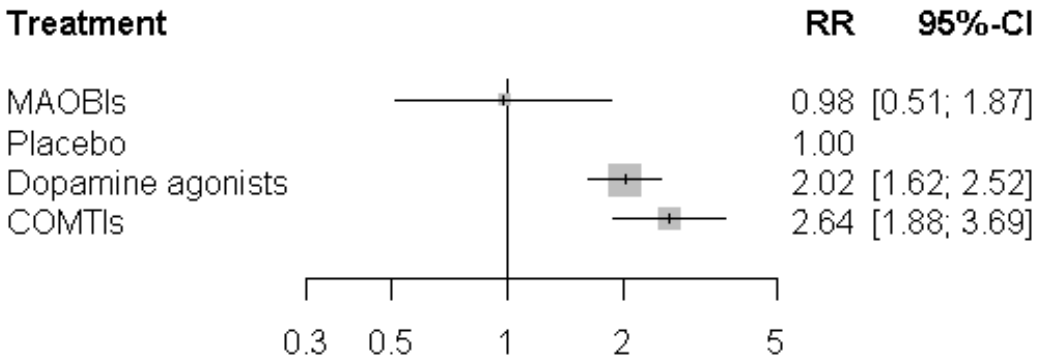


Quantifying heterogeneity/inconsistency:

tau^2 = 0.1426; I^2 = 62.1%

Test of heterogeneity/inconsistency:

Q	d.f.	p.value
58	22	<0.0001



Quantifying heterogeneity/inconsistency:

tau^2 = 0.0992; I^2 = 63.7%

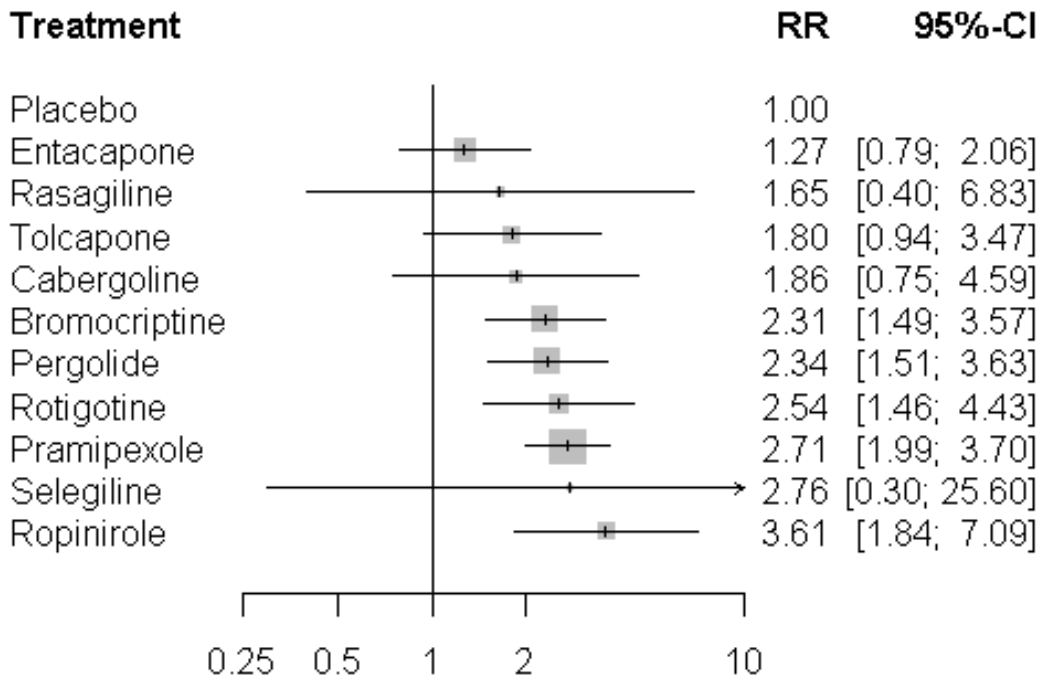
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
60.58	22	<0.0001

Differences between treatments – relative risk and 95% confidence interval

	Treatment A				
Treatment B		MAOBIs	Placebo	Dopamine agonists	COMTIs
	MAOBIs	N/A			
	Placebo	1.02 (0.53, 1.95)	N/A		
	Dopamine agonists	2.06 (1.04, 4.08)	2.02 (1.62, 2.52)	N/A	
	COMTIs	2.69 (1.30, 5.57)	2.64 (1.88, 3.69)	1.30 (0.92, 1.85)	N/A

Hallucinations – FE model

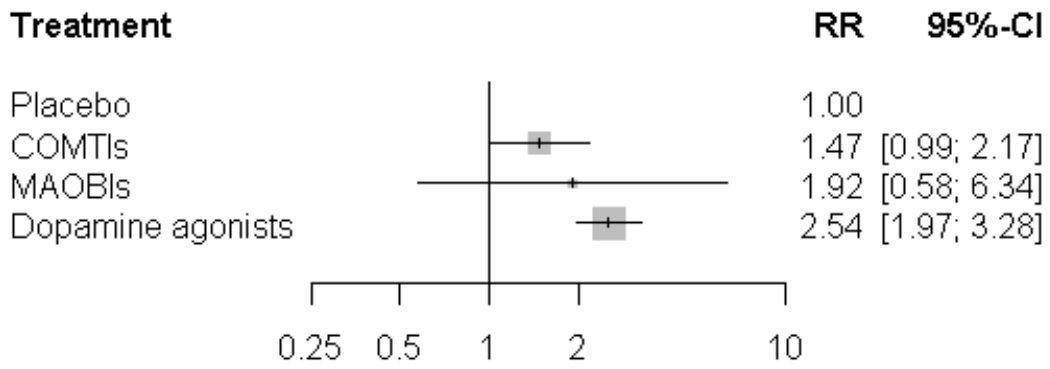


Quantifying heterogeneity/inconsistency:

tau^2 = 0.2206; I^2 = 40.2%

Test of heterogeneity/inconsistency:

Q	d.f.	p.value
28.42	17	0.0403



Quantifying heterogeneity/inconsistency:

tau^2 = 0.1407; I^2 = 31.9%

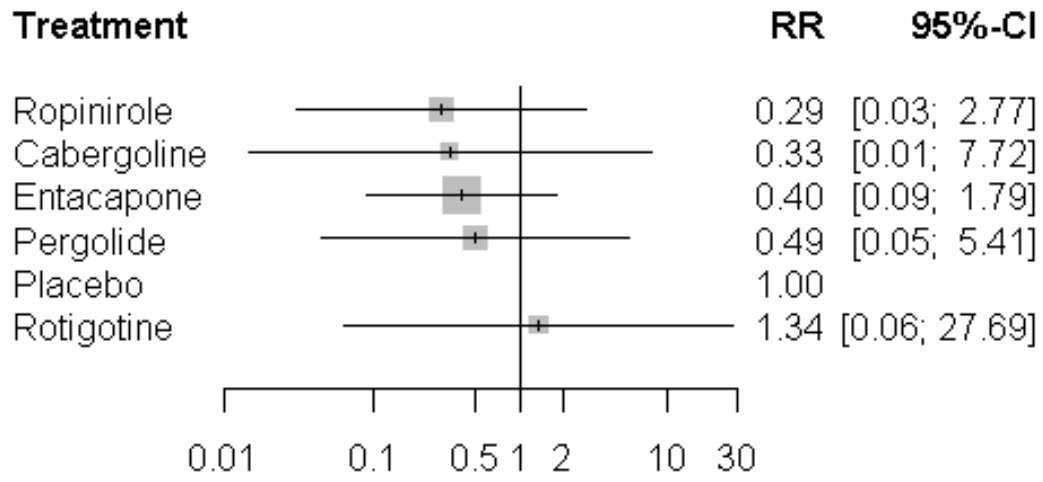
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
26.41	18	0.0907

Differences between treatments – relative risk and 95% confidence interval

	Treatment A				
Treatment B		Placebo	COMTIs	MAOBIs	Dopamine agonists
	Placebo	N/A			
	COMTIs	1.47 (0.99, 2.17)	N/A		
	MAOBIs	1.92 (0.58, 6.34)	1.31 (0.37, 4.60)	N/A	
	Dopamine agonists	2.54 (1.97, 3.28)	1.73 (1.10, 2.73)	1.33 (0.39, 4.51)	N/A

Mortality – FE model



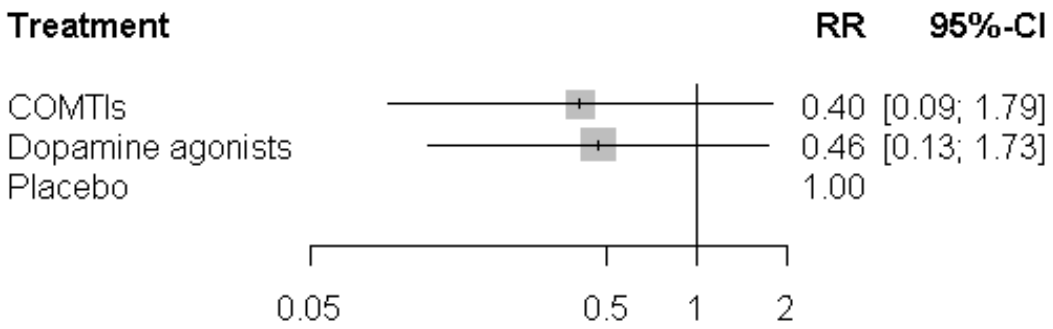
Quantifying heterogeneity/inconsistency:

tau^2 < 0.0001; I^2 = 100%

Test of heterogeneity/inconsistency:

Q d.f. p.value

0 0 <0.0001



Quantifying heterogeneity/inconsistency:

tau^2 < 0.0001; I^2 = 100%

Test of heterogeneity/inconsistency:

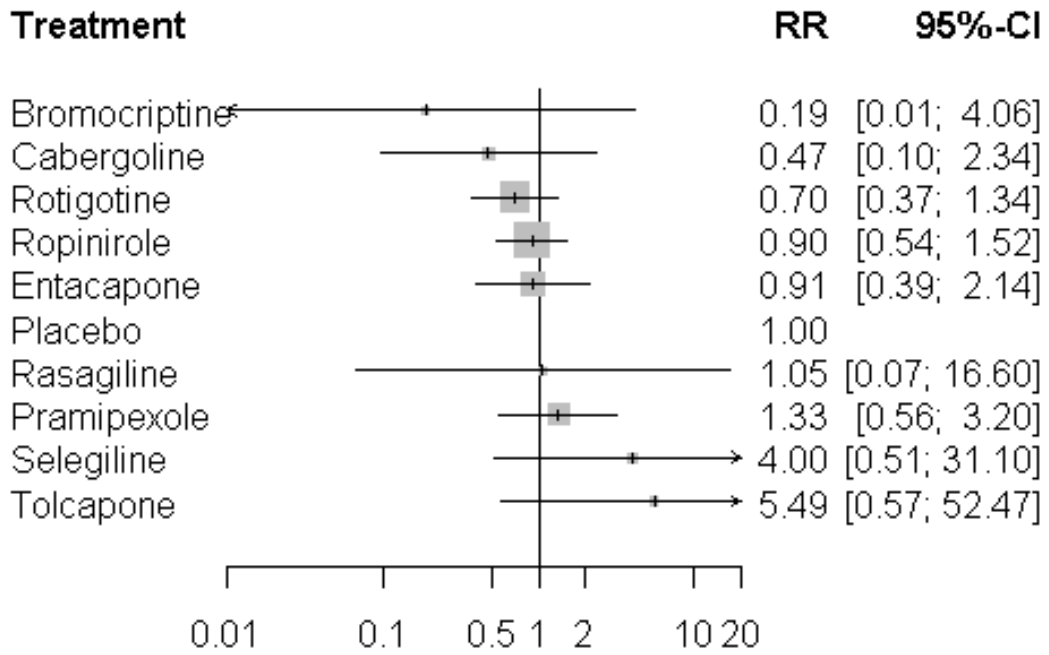
Q d.f. p.value

0 0 <0.0001.

Differences between treatments – relative risk and 95% confidence interval

	Treatment A			
Treatment B		COMTIs	Dopamine agonists	Placebo
	COMTIs	N/A		
	Dopamine agonists	1.15 (0.16, 8.33)	N/A	
	Placebo	2.47 (0.56, 10.92)	2.15 (0.58, 7.98)	N/A

Serious adverse events – FE model

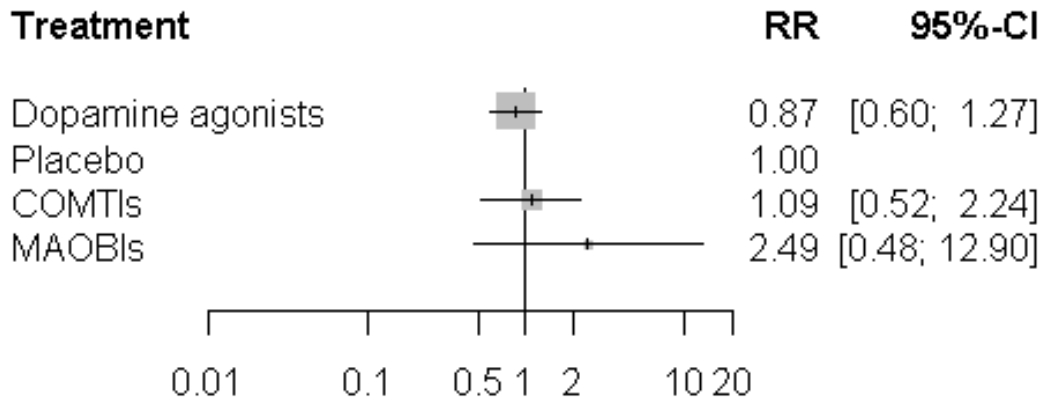


Quantifying heterogeneity/inconsistency:

tau^2 < 0.0001; I^2 = 0%

Test of heterogeneity/inconsistency:

Q	d.f.	p.value
5.75	8	0.675



Quantifying heterogeneity/inconsistency:

tau^2 < 0.0001; I^2 = 0%

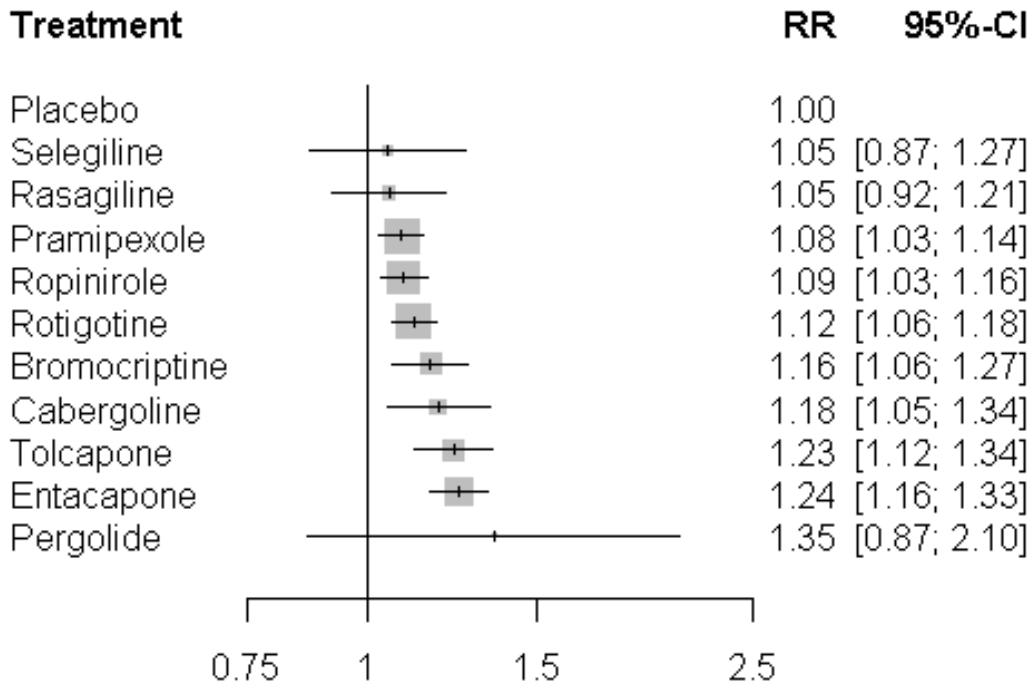
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
8.03	11	0.7104

Differences between treatments – relative risk and 95% confidence interval

	Treatment A				
Treatment B		Dopamine agonists	Placebo	COMTIs	MAOBIs
	Dopamine agonists	N/A			
	Placebo	1.15 (0.78, 1.68)	N/A		
	COMTIs	1.25 (0.58, 2.69)	1.09 (0.52, 2.25)	N/A	
	MAOBIs	2.86 (0.53, 15.47)	2.49 (0.48, 12.90)	2.29 (0.38, 13.85)	N/A

Any adverse event – FE model

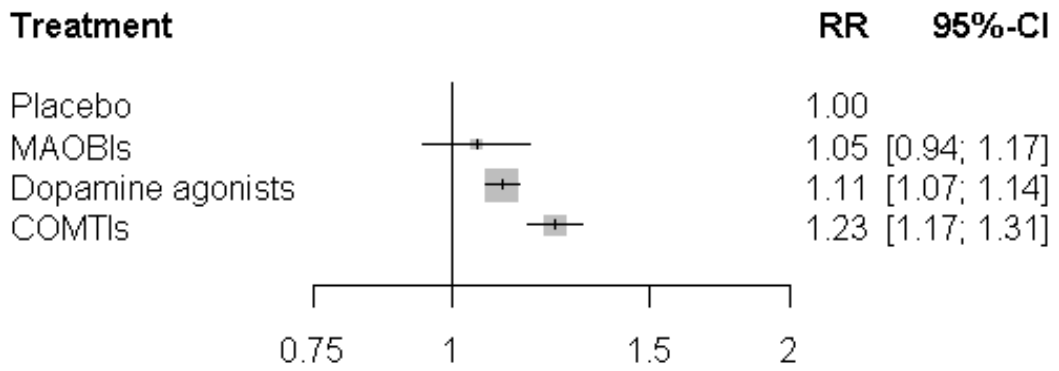


Quantifying heterogeneity/inconsistency:

tau^2 = 0.0028; I^2 = 31.2%

Test of heterogeneity/inconsistency:

Q	d.f.	p.value
26.16	18	0.0961



Quantifying heterogeneity/inconsistency:

tau^2 = 0.0002; I^2 = 3.6%

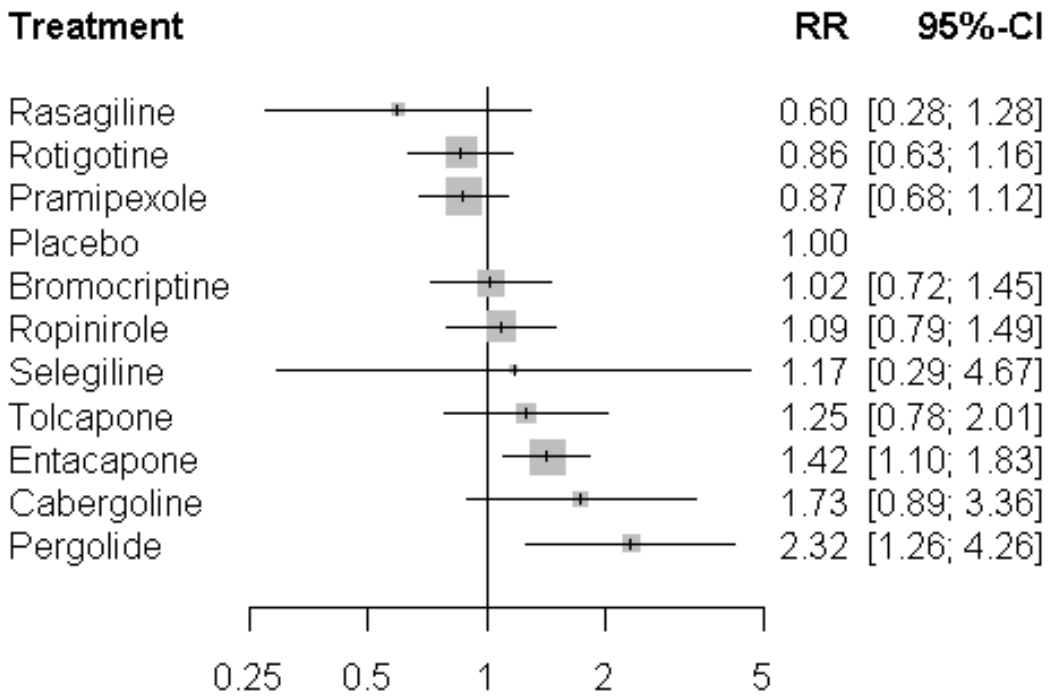
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
20.75	20	0.412

Differences between treatments – relative risk and 95% confidence interval

	Treatment A				
Treatment B		Placebo	MAOBIs	Dopamine agonists	COMTIs
	Placebo	N/A			
	MAOBIs	1.05 (0.94, 1.17)	N/A		
	Dopamine agonists	1.11 (1.07, 1.14)	1.05 (0.94, 1.18)	N/A	
	COMTIs	1.23 (1.17, 1.31)	1.17 (1.04, 1.33)	1.12 (1.05, 1.19)	N/A

Adverse event discontinuations – FE model

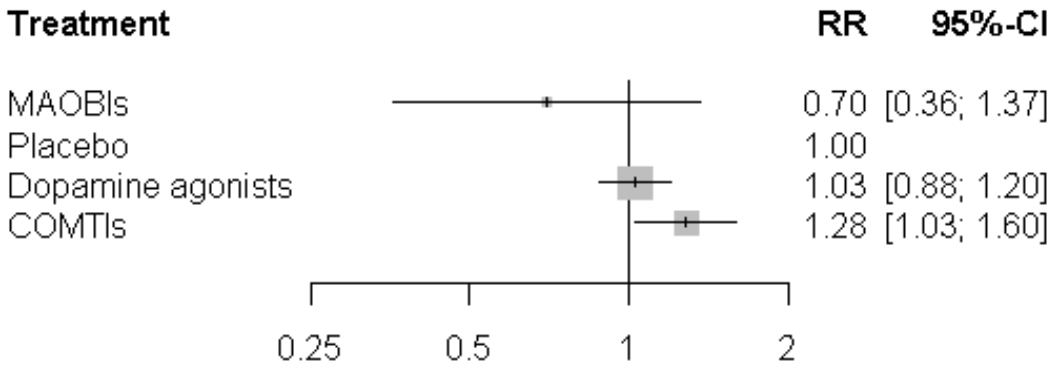


Quantifying heterogeneity/inconsistency:

tau^2 < 0.0001; I^2 = 0%

Test of heterogeneity/inconsistency:

Q	d.f.	p.value
17.85	20	0.597



Quantifying heterogeneity/inconsistency:

tau^2 = 0.0444; I^2 = 27.4%

Test of heterogeneity/inconsistency:

Q	d.f.	p.value
30.3	22	0.1114

Differences between treatments – relative risk and 95% confidence interval

	Treatment A				
Treatment B		MAOBIs	Placebo	Dopamine agonists	COMTIs
	MAOBIs	N/A			
	Placebo	1.43 (0.73, 2.80)	N/A		
	Dopamine agonists	1.47 (0.74, 2.93)	1.03 (0.88, 1.20)	N/A	
	COMTIs	1.84 (0.91, 3.72)	1.28 (1.03, 1.60)	1.25 (0.97, 1.62)	N/A

E.3 Pharmacological management of non-motor symptoms

E.3.1 Daytime hypersomnolence

Effectiveness of modafinil compared to placebo to treat daytime hypersomnolence

Quality assessment						Number of patients		Effect:mean difference (MD)	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	modafinil	placebo	Effect size (95% CI)	Quality
Epworth sleepiness scale (ESS)									
4 studies: Ondo (2008) Lou (2009) Hogl (2003) Adler (2002)	RCT	Serious[1]	Serious[2]	Not serious	Not serious	53	51	MD -2.01 (-3.08, -0.94)	LOW
4 studies: Ondo (2008) Lou (2009) Hogl (2003) Adler (2002)	RCT	Serious[1]	Not serious	Not serious	Serious[3]	45	46	RR 1.55 (0.99, 2.39)	LOW

[1]Serious risk of bias as assessed by NICE RCT quality checklist; [2]Considerable between study heterogeneity (i^2>40%); [3]Non-significant result

E.3.2 Nocturnal akinesia

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Rotigotine	placebo	Mean difference (95% CI)	Quality
Effect of Rotigotine on UPDRS-III motor score									
Trenkwalder 2010	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[4]	166	80	-3.55 (-5.37 to -1.73)	MOD
Effect of Rotigotine on sleep quality (PDSS II total score)									
Trenkwalder 2010	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	166	80	-4.26 (-6.08 to -2.45)	HIGH
Effect of Rotigotine on nocturnal akinesia, dystonia, and cramps (NADCS total score)									
Trenkwalder 2010	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	166	80	-0.41 (-0.79 to -0.04)	HIGH
Effect of Rotigotine on number of nocturias									
Trenkwalder 2010	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[5]	166	80	-0.02 (-0.29 to 0.25)	MOD
Effect of Rotigotine on non-motor symptoms (NMS scale)									
Trenkwalder 2010	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	166	80	-6.65 (-11.99 to -1.31)	HIGH
Effect of Rotigotine on activities of daily life (UPDRS -II)									
Trenkwalder 2010	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious[6]	166	80	-1.49 (-2.32 to -0.65)	HIGH
Effect of Rotigotine on health-related quality of life (PDQ-8)									
Trenkwalder 2010	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious[7]	166	80	-5.74 (-8.74 to -2.75)	HIGH

[1]Low risk of bias as assessed by NICE RCT quality checklist; [2]N/A: Not applicable, only 1 study contributed to this analysis; [3]No serious indirectness, population as was as specified in the review protocol; [4]CI cross MID: between 3.25 (Horváth et al., 2015) and 5 points (Schrag et al., 2006); [5]Non-significant results; [6]CI do not cross MID of 3 points (Schrag et al., 2006); [7]CI do not cross MID of 1.6 points (Peto et al., 2001)

Rotigotine effects on early morning motor function and sleep in Parkinson's disease

Adverse events

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Rotigotine	Placebo	Risk ratio (95%CI)	Quality
Adverse events: Rotigotine vs. placebo									
Trenkwalder 2010	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	166	80	1.27 (1.04 to 1.55)	HIGH

[1]Low risk of bias as assessed by NICE RCT quality checklist; [2]N/A: Not applicable, only 1 study contributed to this analysis; [3]No serious indirectness, population as was as specified in the review protocol

Standard-release compared with controlled-release co-beneldopa

Quality assessment						Effect (number of events)		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Standard Madopar	Madopar CR	Quality
Adverse events								
Madopar Study Group 1989	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	31	32	High

[1]Low risk of bias as assessed by NICE RCT quality checklist; [2]N/A: Not applicable, only 1 study contributed to this analysis; [3]No serious indirectness, population as was as specified in the review protocol

Quality assessment						Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision		
Nocturnal disability							
Madopar Study Group 1989	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[4]	No significant difference[4]	Moderate
Early morning disability							

Quality assessment						Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision		
Madopar Study Group 1989	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[4]	No significant difference[4]	Moderate

[1]Low risk of bias as assessed by NICE RCT quality checklist; [2]N/A: Not applicable, only 1 study contributed to this analysis; [3]No serious indirectness, population as was as specified in the review protocol; [4]Study reported the results to be non-significant. No numerical data was provided to confirm.

E.3.3 Orthostatic hypotension

Droxidopa for Orthostatic Hypotension

Adverse events

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	tmt	control	Odds Ratio (95% CI)	Quality
Total number of adverse events									
2 studies: Hauser 2014 Hauser 2015	RCT	Serious[1]	Serious[2]	Not serious	Serious[3]	111	111	0.99 (0.51, 1.94)	Very low

[1]Serious risk of bias as assessed by NICE RCT quality checklist; [2]Serious inconsistency: I^2 = 40% (Cochrane handbook); [3]Non-significant results

Falls and Fall-related injuries

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	tmt	control	1. Mean Difference (95% CI) 2. Odds Ratio (95% CI)	Quality
Total number of patients experiencing fall related AEs									
2 studies: Hauser 2014 Hauser 2015	RCT	Serious[1]	Not serious	Not serious	Serious[2]	111	111	0.56 (0.29, 1.07)	Low

[1]Serious risk of bias as assessed by NICE RCT quality checklist; [2]Non-significant results

OHQ composite decrease

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	tmt	control	Mean Difference (95% CI)	Quality
Week 1									
2 studies: Hauser 2014 Hauser 2015	RCT	Serious[1]	Not serious	Not serious	Not serious	111	111	-0.88 (-1.65, -0.11)	Moderate
Week 2									
2 studies: Hauser 2014 Hauser 2015	RCT	Serious[1]	Not serious	Not serious	Serious[2]	111	111	-0.52 (-1.09, 0.05)	Low
Week 8									
2 studies: Hauser 2014 Hauser 2015	RCT	Serious[1]	Not serious	Not serious	Serious[2]	111	111	-0.18 (-0.78, 0.42)	Low
[1]Serious risk of bias as assessed by NICE RCT quality checklist; [2]Non-significant results									

Mean change in Standing Systolic BP

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	tmt	control	Mean Difference (95% CI)	Quality
Week 1									
2 studies: Hauser 2014 Hauser 2015	RCT	Serious[1]	Not serious	Not serious	Not serious	111	111	7.34 (2.23, 12.44)	Moderate
Week 8									

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	tmt	control	Mean Difference (95% CI)	Quality
2 studies: Hauser 2014 Hauser 2015	RCT	Serious[1]	Not serious	Not serious	Serious[2]	111	111	3.16 (-1.80, 8.12)	Low

[1]Serious risk of bias as assessed by NICE RCT quality checklist; [2]Non-significant results

Domperidone vs. Fludrocortisone for Orthostatic Hypotension

Adverse events

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	domperidone	fludrocortisone	Odds Ratio (95% CI)	Quality
Patients recording Adverse Events									
1 study: Schoffer 2007	RCT	Very Serious[1]	N/A[2]	Not serious	Serious[3]	13	13	0.73 (0.15, 3.47)	Very Low

[1]Very serious risk of bias as assessed by NICE RCT quality checklist; [2]N/A: only 1 study contributed to the analysis; [3]Non-significant results

Blood pressure

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	domperidone	fludrocortisone	Mean Difference (95% CI)	Quality
Supine blood pressure: mm/Hg									
1 study: Schoffer 2007	RCT	Very Serious[1]	N/A[2]	Not serious	Serious[3]	13	13	-4 (-23.6 to 15.64)	Very Low

[1]Very serious risk of bias as assessed by NICE RCT quality checklist; [2]N/A: only 1 study contributed to the analysis; [3]Non-significant results

Autonomic function

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	domperidone	fludrocortisone	Mean Difference (95% CI)	Quality
COMPASS:OD									
1 study: Schoffer 2007	RCT	Very Serious[1]	N/A[2]	Not serious	Serious[3]	13	13	-1 (-2.96 to 0.96)	Very Low

[1]Very serious risk of bias as assessed by NICE RCT quality checklist; [2]N/A: only 1 study contributed to the analysis; [3]Non-significant results

E.3.4 Psychotic symptoms (hallucinations and delusions)

GRADE profile for network meta-analyses

UPDRS Motor

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in UPDRS motor score					
8	Serious[1]	Not serious	Not serious	Serious[2]	LOW

1 Downgrade 1 level: Limitations in the design or execution of the study
2 Downgrade 1 level: no interventions had a median rank of 1 [1 to ± n/3]

BPRS total

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in hallucination score					
7	Serious[1]	Not serious[2]	Not serious[3]	Not serious	MODERATE

1 Downgrade 1 level: Limitations in the design or execution of the study
2 Assessed based on residual deviance, deviance information criterion and tau2 (tau2<0.5)
3 Considered not serious as population, interventions, comparator and outcomes are as defined in protocol

BPRS Hallucination

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in hallucination score					
3	Serious[1]	Not serious[2]	Not serious[3]	Not serious	MODERATE

1 Downgrade 1 level: Limitations in the design or execution of the study
2 Assessed based on residual deviance, deviance information criterion and tau2 (tau2<0.5)
3 Considered not serious as population, interventions, comparator and outcomes are as defined in protocol

Hallucination – BPRS, NPI, Baylor PD Hallucination, Structured interview for hallucinations in PD

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in hallucination score					
5	Serious[1]	Not serious[2]	Not serious[3]	Serious[4]	LOW

1 Downgrade 1 level: Limitations in the design or execution of the study
2 Assessed based on residual deviance, deviance information criterion and tau2 (tau2<0.5)
3 Considered not serious as population, interventions, comparator and outcomes are as defined in protocol
4 Downgrade 1 level: no interventions had a median rank of 1 [1 to ± n/3]

Positive symptoms – SAPS, Positive PANSS, BPRS Positive

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in positive symptom score					
4	Serious[1]	Not serious[2]	Not serious[3]	Not serious	MODERATE

1 Downgrade 1 level: Limitations in the design or execution of the study
2 Assessed based on residual deviance, deviance information criterion and tau2 (tau2<0.5)
3 Considered not serious as population, interventions, comparator and outcomes are as defined in protocol

Treatment discontinuation due to adverse events

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
The rate of an adverse event occurring					
8	Serious[1]	Not serious[2]	Not serious[3]	Serious[4]	LOW

1 Downgrade 1 level: Limitations in the design or execution of the study
2 Assessed based on residual deviance, deviance information criterion and tau2 (tau2<0.5)
3 Considered not serious as population, interventions, comparator and outcomes are as defined in protocol
4 Downgrade 1 level: no interventions had a median rank of 1 [1 to ± n/3]

Adverse events – Estimate of rate

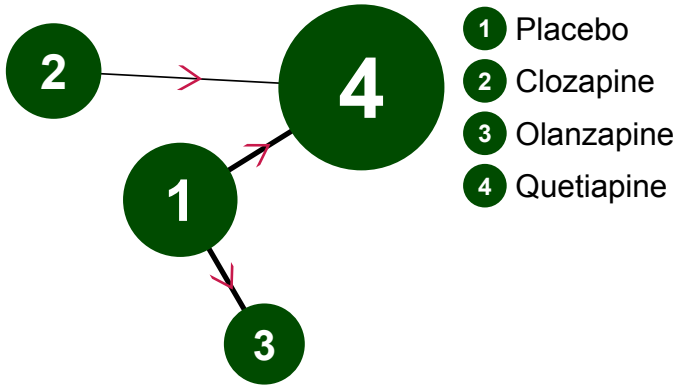
Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Adverse events (Ratio)					

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
5	Serious[1]	Not serious[2]	Not serious[3]	Not serious[4]	LOW

1 Downgrade 1 level: Limitations in the design or execution of the study
2 Assessed based on residual deviance, deviance information criterion and tau2 (tau2<0.5)
3 Considered not serious as population, interventions, comparator and outcomes are as defined in protocol
4 Downgrade 1 level: no interventions had a median rank of 1 [1 to ± n/3]

Network meta-analyses

Adverse events (rate)



Size of nodes is proportional to total number of participants randomised to receive the treatment in question across the evidence-base. Width of connecting lines is proportional to number of trial-level comparisons available. Arrowheads indicate direction of effect in pairwise data (a > b denotes a is more effective than b) – filled arrowheads show comparisons where one option is significantly superior (p<0.05); outlined arrowheads show direction of trend where effect does not reach statistical significance.

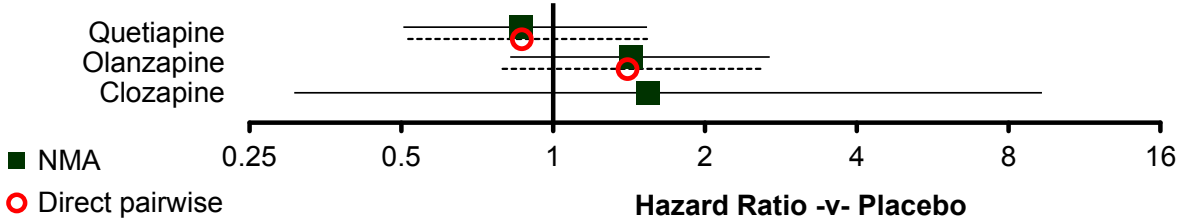
Adverse events (rate) – evidence network

Adverse events (rate) – input data

	Placebo	Clozapine	Olanzapine	Quetiapine
Morgante et al. (2004) - 0.23yr		5/1722		3/1701
Ondo et al. (2002) - 0.17yr	12/735		17/1029	
Fernandez et al. (2009) - 0.19yr	11/538.125			9/430.5
Ondo et al. (2005) - 0.23yr	14/756			23/1596
Nichols et al. (2013) - 0.08yr	5/224		15/280	
Rate data: numerators represent numbers of AEs; denominators are patient-days of exposure				

Adverse events (rate) – relative effectiveness of all pairwise combinations

	Placebo	Clozapine	Olanzapine	Quetiapine
Placebo		N/A	N/A	N/A
Clozapine	1.55 (0.31, 9.31)		N/A	N/A
Olanzapine	1.43 (0.82, 2.69)	0.92 (0.14, 5.29)		N/A
Quetiapine	0.86 (0.50, 1.53)	0.57 (0.10, 2.58)	0.60 (0.27, 1.35)	
Values given are hazard ratios. The segment below and to the left of the shaded diagonal is derived from the network meta-analysis, reflecting direct and indirect evidence of treatment effects (row versus column). The point estimate reflects the mean of the posterior distribution, and numbers in parentheses are 95% credible intervals. Because it is not easily possible to pool dichotomous and rate data and derive analogous estimates of hazard ratios from a single frequentist analysis of direct data only, the segment above and to the right of the shaded diagonal is left blank				

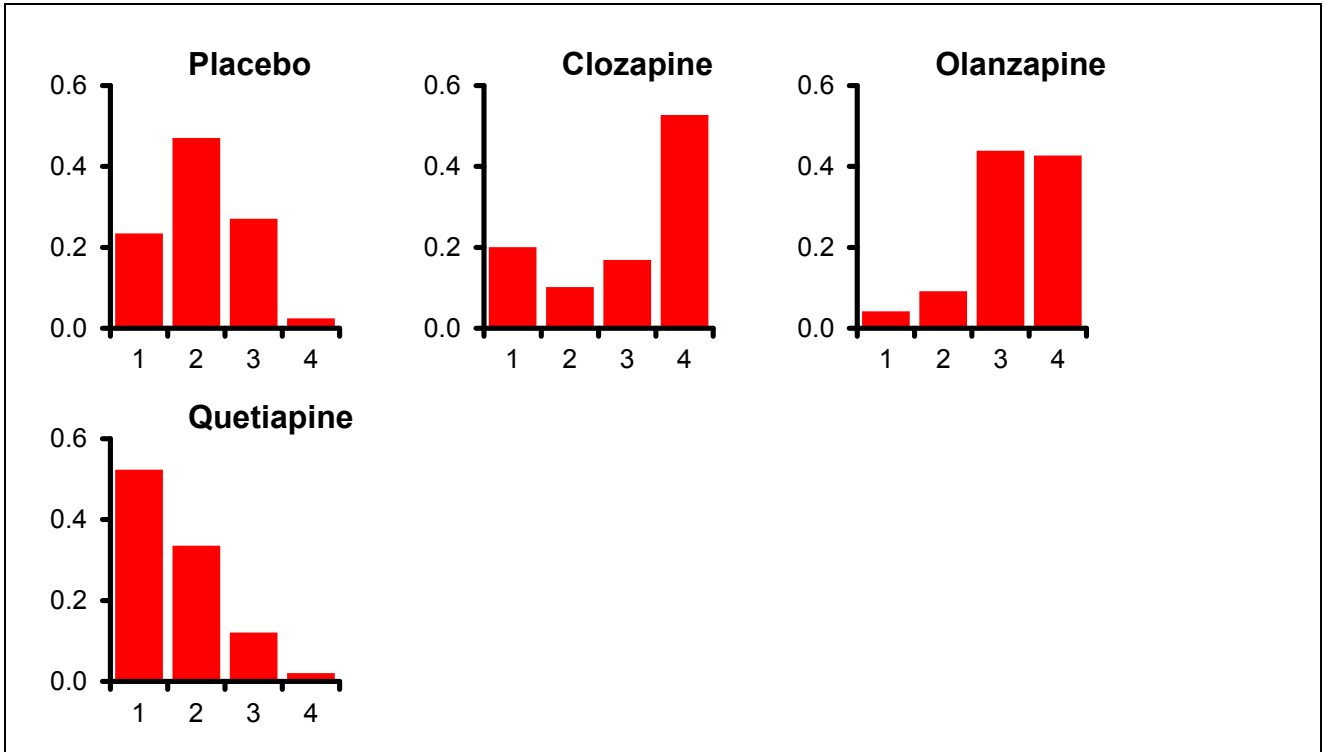


Values less than 1 favour the comparator treatment; values greater than 1 favour placebo. Error bars are 95% credible intervals. Direct pairwise estimates are drawn from inconsistency model.

Adverse events (rate) – relative effect of all options versus common comparator

Adverse events (rate) – rankings for each comparator

	Probability best	Median rank (95%CI)
Placebo	0.234	2 (1, 3)
Clozapine	0.201	4 (1, 4)
Olanzapine	0.042	3 (1, 4)
Quetiapine	0.523	1 (1, 3)



Histograms show probability that each treatment is ranked in each position relative to the other treatments in the network.

Adverse events (rate) – rank probability histograms

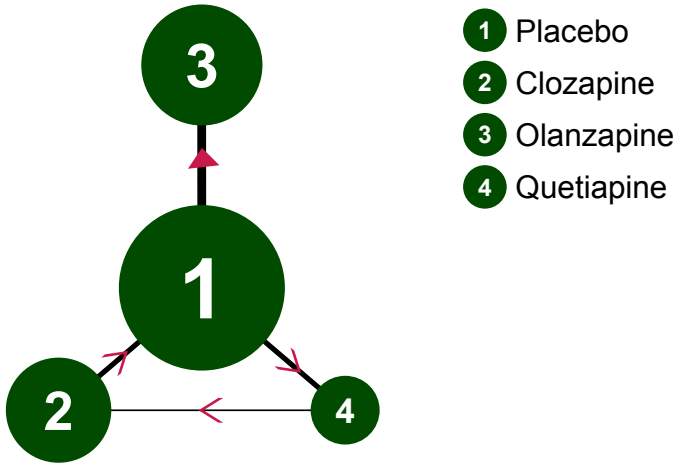
Adverse events (rate) – model fit statistics

Residual deviance	Dbar	Dhat	pD	DIC
10.42 (compared to 10 datapoints)	51.721	43.711	8.01	59.732

Adverse events (rate) – notes

- Count (Poisson; log link); fixed effects
- 50000 burn-ins; 10000 recorded iterations

Treatment discontinuation due to AEs



Size of nodes is proportional to total number of participants randomised to receive the treatment in question across the evidence-base. Width of connecting lines is proportional to number of trial-level comparisons available. Arrowheads indicate direction of effect in pairwise data (a > b denotes a is more effective than b) – filled arrowheads show comparisons where one option is significantly superior (p<0.05); outlined arrowheads show direction of trend where effect does not reach statistical significance.

Treatment discontinuation due to AEs – evidence network

Treatment discontinuation due to AEs – input data

	Placebo	Clozapine	Olanzapine	Quetiapine
Morgante et al. (2004)		3/23		2/22

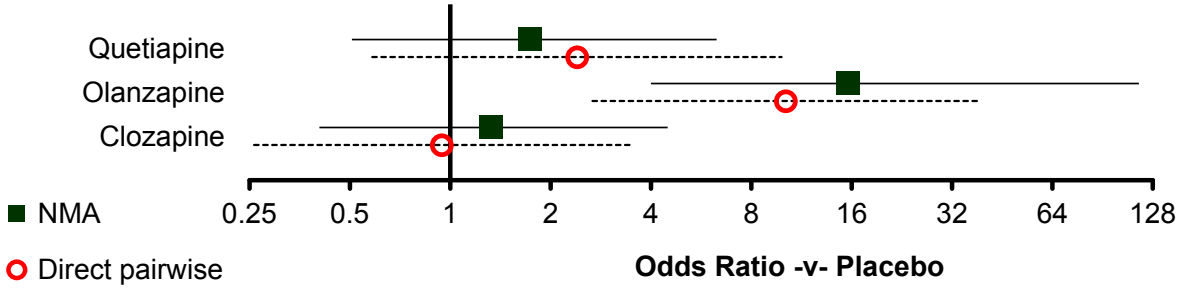
	Placebo	Clozapine	Olanzapine	Quetiapine
Friedman (1999)	3/30	3/30		
Pollak et al. (2004)	2/28	2/32		
Fernandez et al. (2009)	1/8			4/8
Breier et al. (2002) – Europe	1/28		8/49	
Breier et al. (2002) – USA	1/42		10/41	
Nichols et al. (2013)	0/9		7/14	
Shotbolt et al. (2009)	3/13			3/11

Treatment discontinuation due to AEs – relative effectiveness of all pairwise combinations

	Placebo	Clozapine	Olanzapine	Quetiapine
Placebo		0.94 (0.26, 3.45)	10.14 (2.67, 38.50)	2.40 (0.58, 9.87)
Clozapine	1.33 (0.41, 4.49)		-	0.67 (0.10, 4.43)
Olanzapine	15.70 (4.01, 116.30)	12.25 (1.86, 116.70)		-
Quetiapine	1.74 (0.51, 6.29)	1.32 (0.33, 5.52)	0.11 (0.01, 0.73)	

Values given are odds ratios.

The segment below and to the left of the shaded cells is derived from the network meta-analysis, reflecting direct and indirect evidence of treatment effects (row versus column). The point estimate reflects the mean of the posterior distribution, and numbers in parentheses are 95% credible intervals. The segment above and to the right of the shaded cells gives pooled direct evidence (random-effects pairwise meta-analysis), where available (column versus row). Numbers in parentheses are 95% confidence intervals.

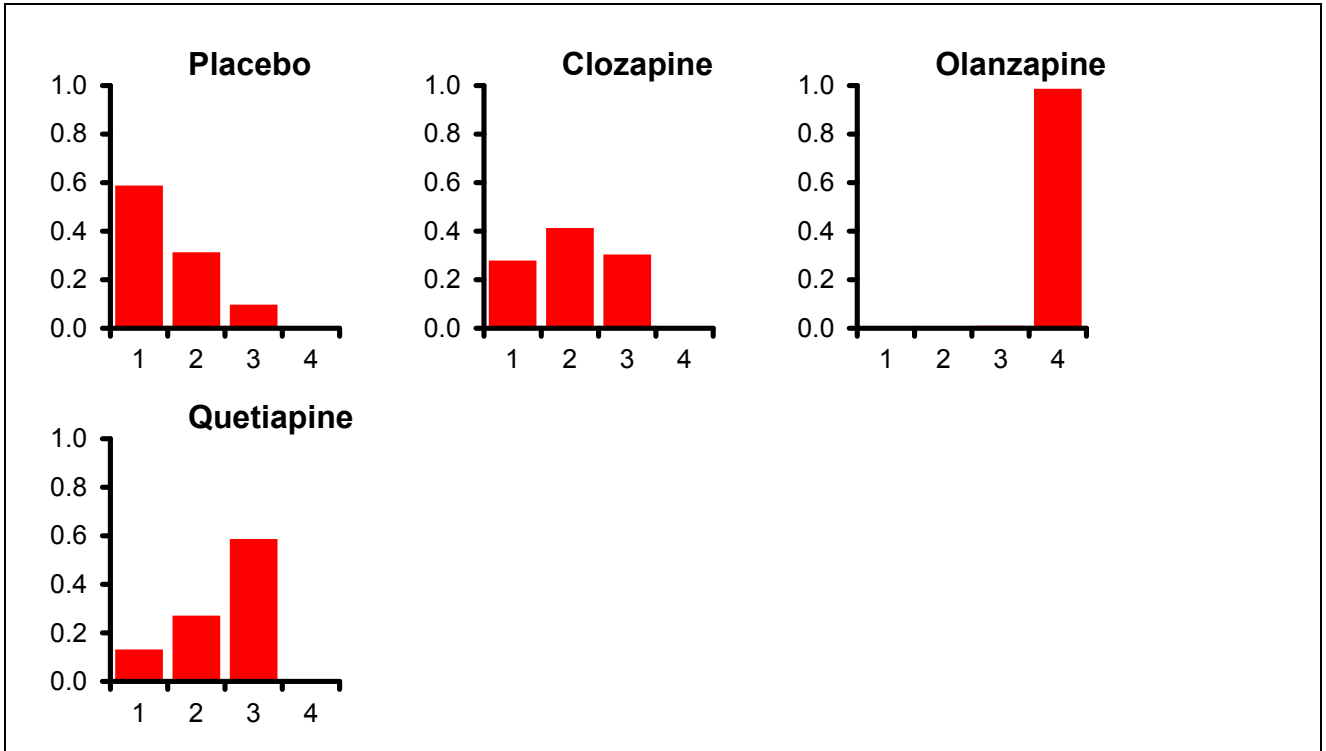


Values less than 1 favour placebo; values greater than 1 favour the comparator treatment. Solid error bars are 95% credible intervals; dashed error bars are 95% confidence interval.

Treatment discontinuation due to AEs – relative effect of all options versus common comparator

Treatment discontinuation due to AEs – rankings for each comparator

	Probability best	Median rank (95%CI)
Placebo	0.589	1 (1, 3)
Clozapine	0.280	2 (1, 3)
Olanzapine	0.000	4 (4, 4)
Quetiapine	0.132	3 (1, 3)



Histograms show probability that each treatment is ranked in each position relative to the other treatments in the network.

Treatment discontinuation due to AEs – rank probability histograms

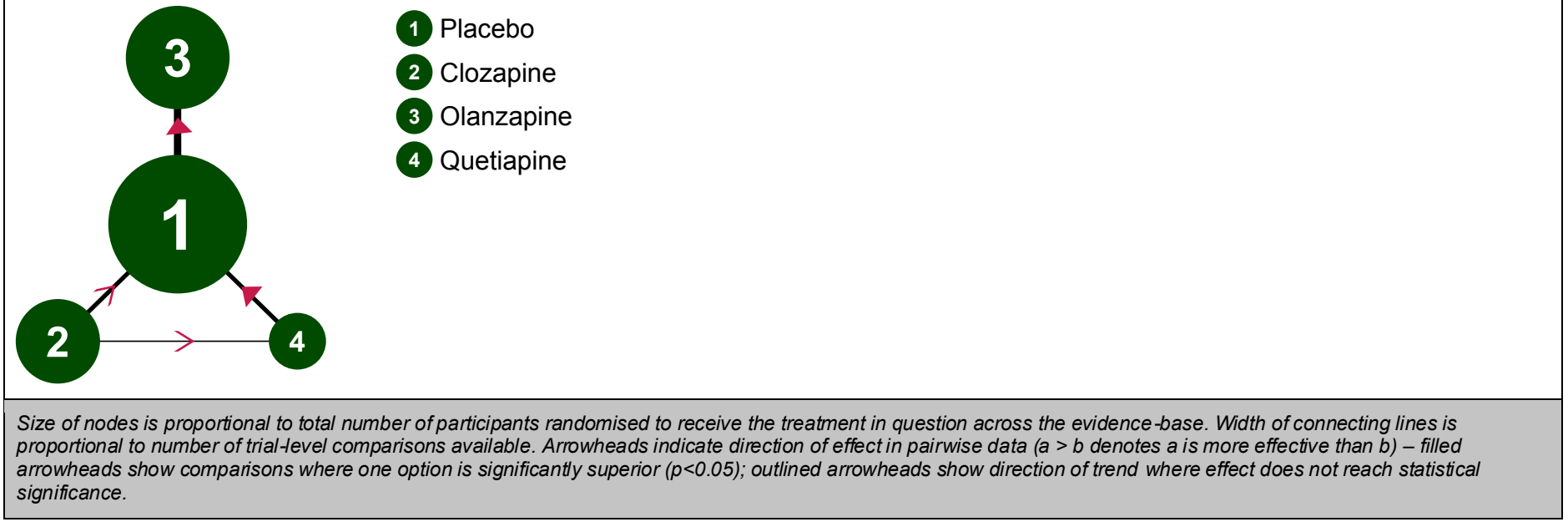
Treatment discontinuation due to AEs – model fit statistics

Residual deviance	Dbar	Dhat	pD	DIC
15.52 (compared to 16 datapoints)	56.334	45.307	11.028	67.362

Treatment discontinuation due to AEs – notes

- Dichotomous synchronic (binomial; logit link); fixed effects
- 50000 burn-ins; 10000 recorded iterations

UPDRS III (motor) score



Size of nodes is proportional to total number of participants randomised to receive the treatment in question across the evidence-base. Width of connecting lines is proportional to number of trial-level comparisons available. Arrowheads indicate direction of effect in pairwise data (a > b denotes a is more effective than b) – filled arrowheads show comparisons where one option is significantly superior (p<0.05); outlined arrowheads show direction of trend where effect does not reach statistical significance.

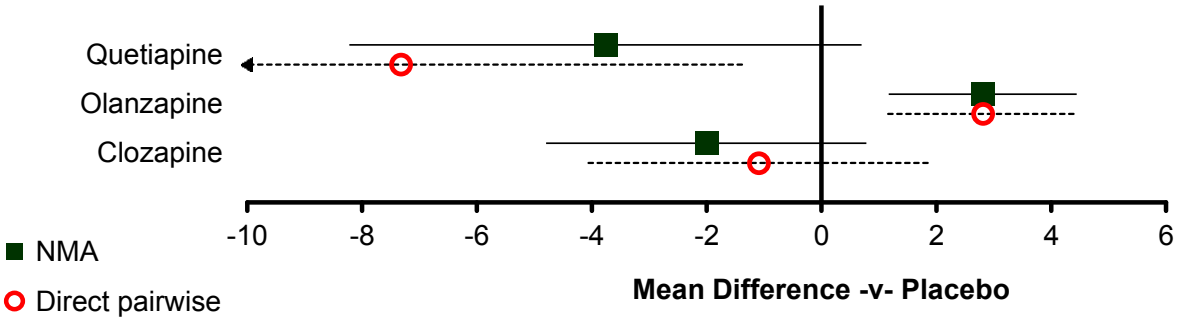
UPDRS III (motor) score – evidence network

UPDRS III (motor) score – input data

	Placebo	Clozapine	Olanzapine	Quetiapine
Morgante et al. (2004)		-1.30 (9.30)		1.00 (11.00)
Friedman (1999)	-1.80 (6.00)	-3.60 (9.50)		
Pollak et al. (2004)	-3.00 (8.10)	-3.50 (7.70)		
Fernandez et al. (2009)	2.83 (7.46)			-5.74 (6.84)
Breier et al. (2002) – Europe	-0.30 (5.00)		2.70 (6.00)	
Breier et al. (2002) – USA	-0.20 (4.30)		2.60 (6.00)	
Nichols et al. (2013)	1.00 (12.18)		0.80 (12.86)	
Shotbolt et al. (2009)	1.10 (14.69)			-3.00 (13.47)
Values are mean change from baseline to follow up (SD)				

UPDRS III (motor) score – relative effectiveness of all pairwise combinations

	Placebo	Clozapine	Olanzapine	Quetiapine
Placebo		-1.09 (-4.06, 1.88)	2.81 (1.16, 4.46)	-7.32 (-13.28, -1.37)
Clozapine	-1.98 (-4.80, 0.78)		-	2.30 (-4.01, 8.61)
Olanzapine	2.82 (1.17, 4.44)	4.80 (1.62, 8.07)		-
Quetiapine	-3.75 (-8.22, 0.70)	-1.75 (-6.29, 2.74)	-6.58 (-11.32, -1.83)	
Values given are weighted mean differences. The segment below and to the left of the shaded cells is derived from the network meta-analysis, reflecting direct and indirect evidence of treatment effects (row versus column). The point estimate reflects the mean of the posterior distribution, and numbers in parentheses are 95% credible intervals. The segment above and to the right of the shaded cells gives pooled direct evidence (random-effects pairwise meta-analysis), where available (column versus row). Numbers in parentheses are 95% confidence intervals.				

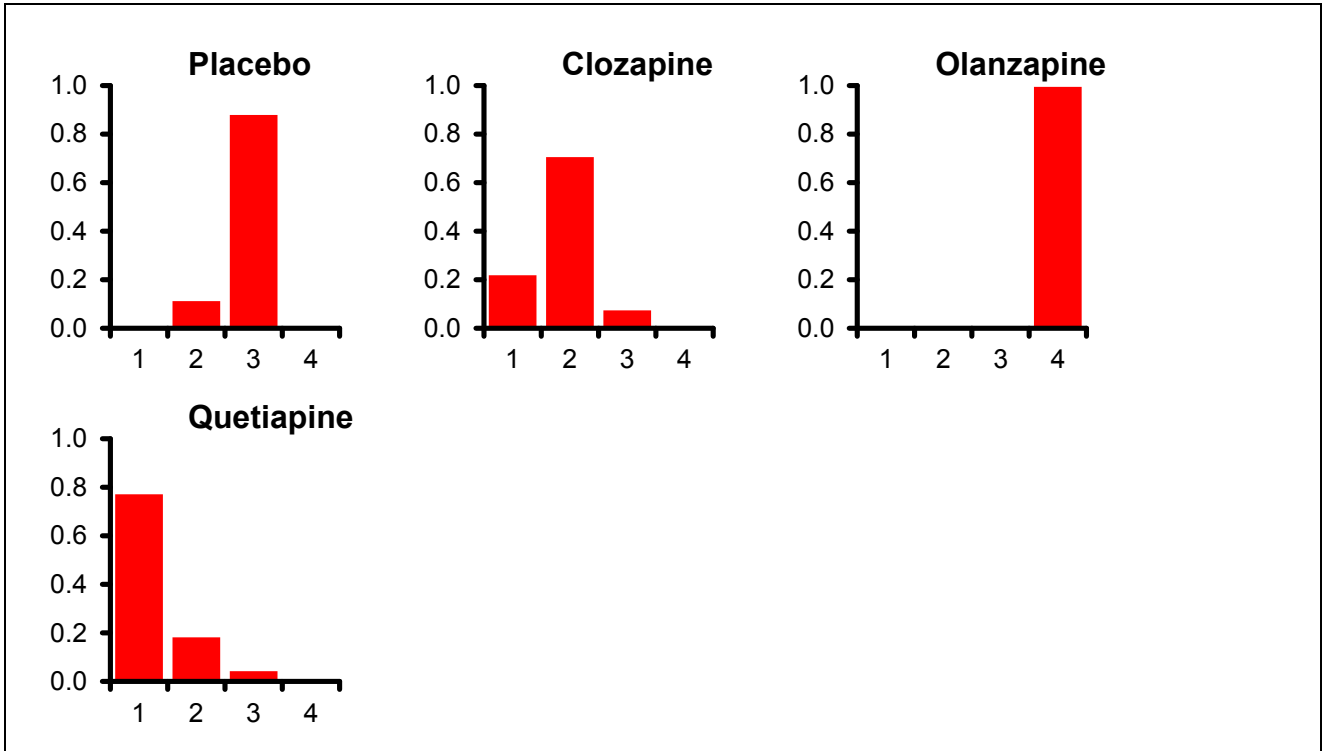


Values greater than 0 favour placebo; values less than 0 favour the comparator treatment. Solid error bars are 95% credible intervals; dashed error bars are 95% confidence interval.

UPDRS III (motor) score – relative effect of all options versus common comparator

UPDRS III (motor) score – rankings for each comparator

	Probability best	Median rank (95%CI)
Placebo	0.009	3 (2, 3)
Clozapine	0.219	2 (1, 3)
Olanzapine	0.000	4 (4, 4)
Quetiapine	0.772	1 (1, 3)



Histograms show probability that each treatment is ranked in each position relative to the other treatments in the network.

UPDRS III (motor) score – rank probability histograms

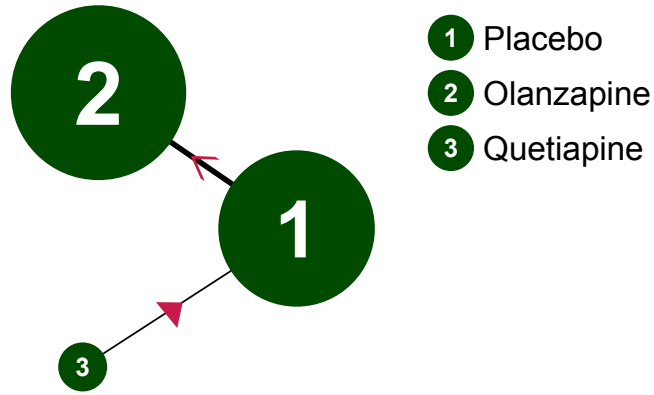
UPDRS III (motor) score – model fit statistics

Residual deviance	Dbar	Dhat	pD	DIC
15.25 (compared to 16 datapoints)	64.259	53.29	10.969	75.228

UPDRS III (motor) score – notes

- Continuous (normal; identity link); fixed effects
- 50000 burn-ins; 10000 recorded iterations

BPRS hallucinations



Size of nodes is proportional to total number of participants randomised to receive the treatment in question across the evidence-base. Width of connecting lines is proportional to number of trial-level comparisons available. Arrowheads indicate direction of effect in pairwise data (a > b denotes a is more effective than b) – filled arrowheads show comparisons where one option is significantly superior (p<0.05); outlined arrowheads show direction of trend where effect does not reach statistical significance.

BPRS hallucinations – evidence network

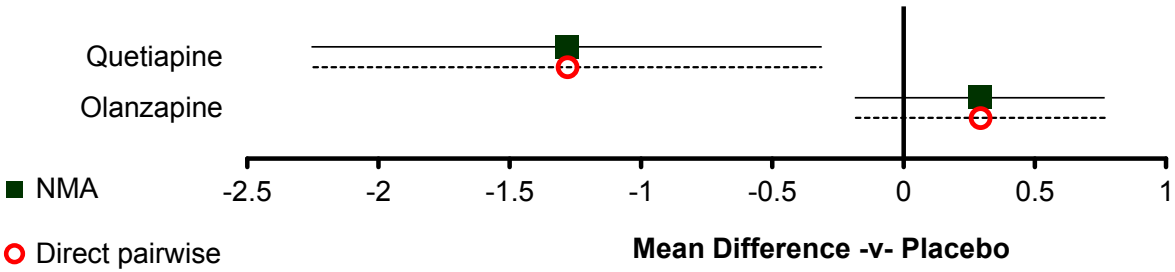
BPRS hallucinations – input data

	Placebo	Olanzapine	Quetiapine
Fernandez et al. (2009)	-0.04 (0.82)		-1.32 (1.13)
Breier et al. (2002) – Europe	-1.40 (1.50)	-1.00 (1.50)	

	Placebo	Olanzapine	Quetiapine
Breier et al. (2002) – USA	-0.90 (1.40)	-0.70 (1.60)	
Values are mean change from baseline to follow up (SD)			

BPRS hallucinations – relative effectiveness of all pairwise combinations

	Placebo	Olanzapine	Quetiapine
Placebo		0.29 (-0.18, 0.77)	-1.28 (-2.25, -0.31)
Olanzapine	0.29 (-0.19, 0.77)		-
Quetiapine	-1.28 (-2.26, -0.31)	-1.58 (-2.65, -0.48)	
Values given are weighted mean differences. The segment below and to the left of the shaded cells is derived from the network meta-analysis, reflecting direct and indirect evidence of treatment effects (row versus column). The point estimate reflects the mean of the posterior distribution, and numbers in parentheses are 95% credible intervals. The segment above and to the right of the shaded cells gives pooled direct evidence (random-effects pairwise meta-analysis), where available (column versus row). Numbers in parentheses are 95% confidence intervals.			

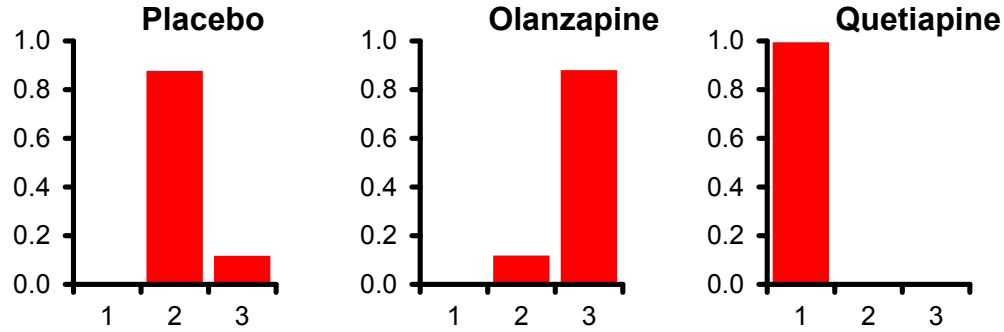


Values greater than 0 favour placebo; values less than 0 favour the comparator treatment. Solid error bars are 95% credible intervals; dashed error bars are 95% confidence interval.

BPRS hallucinations – relative effect of all options versus common comparator

BPRS hallucinations – rankings for each comparator

	Probability best	Median rank (95%CI)
Placebo	0.005	2 (2, 3)
Olanzapine	0.001	3 (2, 3)
Quetiapine	0.994	1 (1, 1)



Histograms show probability that each treatment is ranked in each position relative to the other treatments in the network.

BPRS hallucinations – rank probability histograms

BPRS hallucinations – model fit statistics

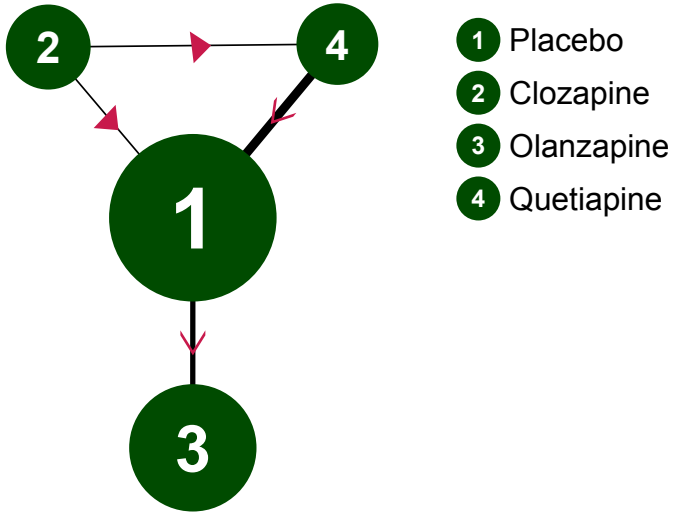
Residual deviance	Dbar	Dhat	pD	DIC
5.17 (compared to 6 datapoints)	0.446	-4.555	5	5.446

BPRS hallucinations – notes

- Continuous (normal; identity link); fixed effects

- 50000 burn-ins; 10000 recorded iterations

BPRS total



Size of nodes is proportional to total number of participants randomised to receive the treatment in question across the evidence-base. Width of connecting lines is proportional to number of trial-level comparisons available. Arrowheads indicate direction of effect in pairwise data (a > b denotes a is more effective than b) – filled arrowheads show comparisons where one option is significantly superior (p<0.05); outlined arrowheads show direction of trend where effect does not reach statistical significance.

BPRS total – evidence network

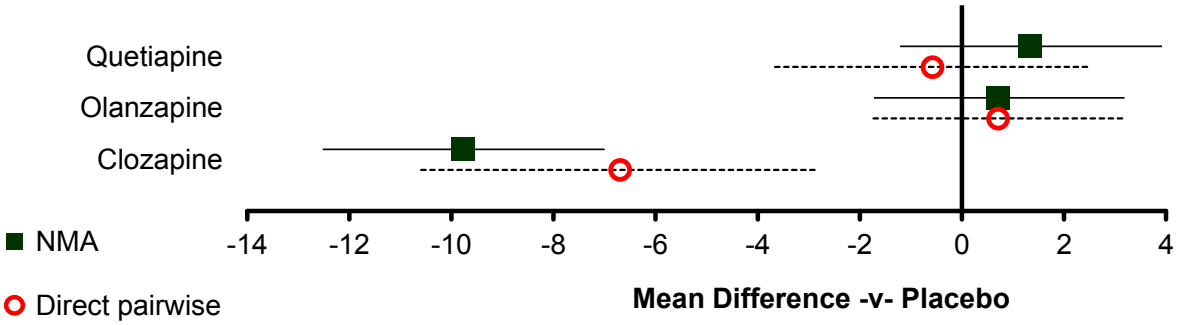
BPRS total – input data

	Placebo	Clozapine	Olanzapine	Quetiapine
Morgante et al. (2004)		-10.70 (3.60)		1.60 (4.20)
Friedman (1999)	-2.60 (6.75)	-9.30 (7.79)		
Fernandez et al. (2009)	-0.28 (7.63)			-1.00 (6.97)
Breier et al. (2002)	-5.50 (8.30)		-4.30 (8.30)	
Breier et al. (2002)	-3.10 (5.90)		-2.70 (8.30)	
Shotbolt et al. (2009)	-2.50 (6.40)			-4.20 (6.10)
Rabey (2007)	-4.50 (2.90)			-4.20 (6.00)

BPRS total – relative effectiveness of all pairwise combinations

	Placebo	Clozapine	Olanzapine	Quetiapine
Placebo		-6.70 (-10.59, -2.81)	0.71 (-1.73, 3.15)	-0.58 (-3.66, 2.50)
Clozapine	-9.78 (-12.52, -7.00)		-	12.30 (9.88, 14.72)
Olanzapine	0.71 (-1.72, 3.19)	10.49 (6.79, 14.19)		-
Quetiapine	1.33 (-1.21, 3.92)	11.12 (8.91, 13.30)	0.62 (-2.90, 4.19)	

Values given are hazard ratios.
The segment below and to the left of the shaded diagonal is derived from the network meta-analysis, reflecting direct and indirect evidence of treatment effects (row versus column). The point estimate reflects the mean of the posterior distribution, and numbers in parentheses are 95% credible intervals. Because it is not easily possible to pool dichotomous and rate data and derive analogous estimates of hazard ratios from a single frequentist analysis of direct data only, the segment above and to the right of the shaded diagonal is left blank

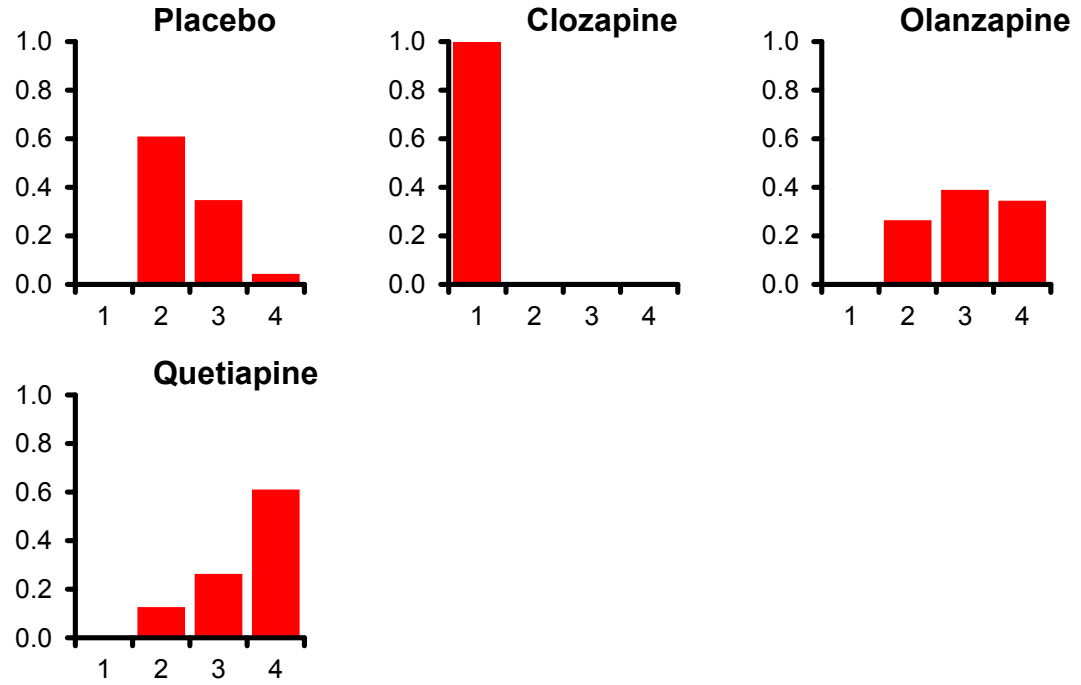


Values less than 1 favour the comparator treatment; values greater than 1 favour placebo. Error bars are 95% credible intervals. Direct pairwise estimates are drawn from inconsistency model.

BPRS total – relative effect of all options versus common comparator

BPRS total – rankings for each comparator

	Probability best	Median rank (95%CI)
Placebo	0.000	2 (2, 4)
Clozapine	1.000	1 (1, 1)
Olanzapine	0.000	3 (2, 4)
Quetiapine	0.000	4 (2, 4)



Histograms show probability that each treatment is ranked in each position relative to the other treatments in the network.

BPRS total – rank probability histograms

BPRS total – model fit statistics

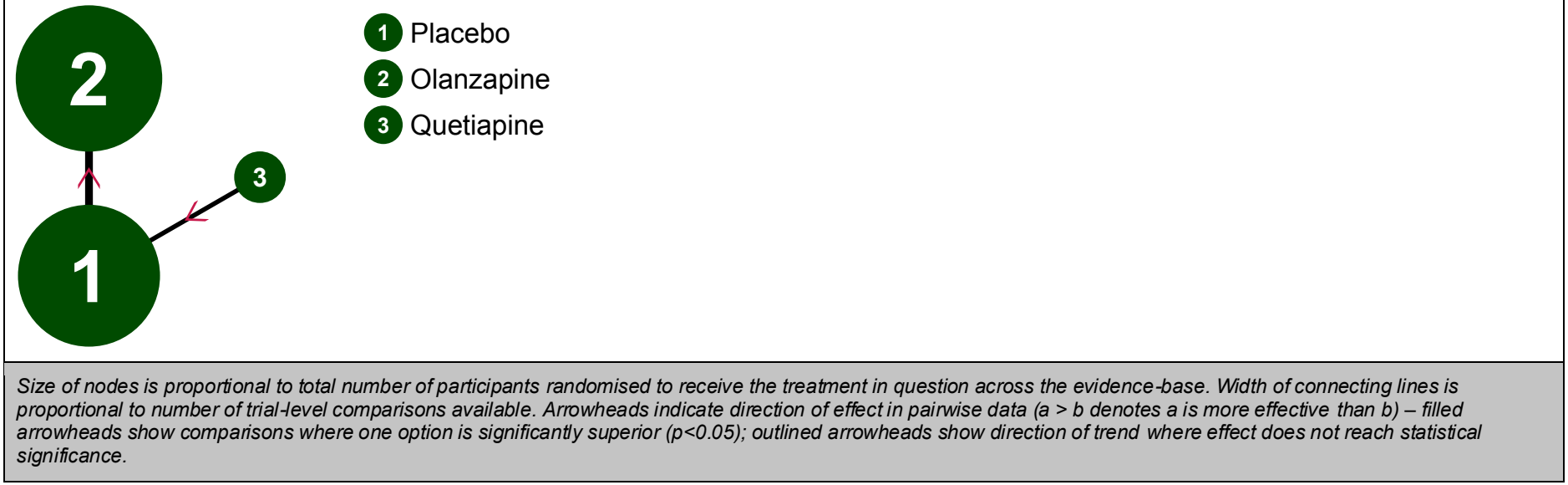
Residual deviance	Dbar	Dhat	pD	DIC
15.3 (compared to 14 datapoints)	51.777	41.735	10.042	61.819

BPRS total – notes

- Continuous (normal; identity link); fixed effects
- 50000 burn-ins; 10000 recorded iterations

Network meta-analyses (pooling across outcomes)

Hallucinations



Size of nodes is proportional to total number of participants randomised to receive the treatment in question across the evidence-base. Width of connecting lines is proportional to number of trial-level comparisons available. Arrowheads indicate direction of effect in pairwise data (a > b denotes a is more effective than b) – filled arrowheads show comparisons where one option is significantly superior (p<0.05); outlined arrowheads show direction of trend where effect does not reach statistical significance.

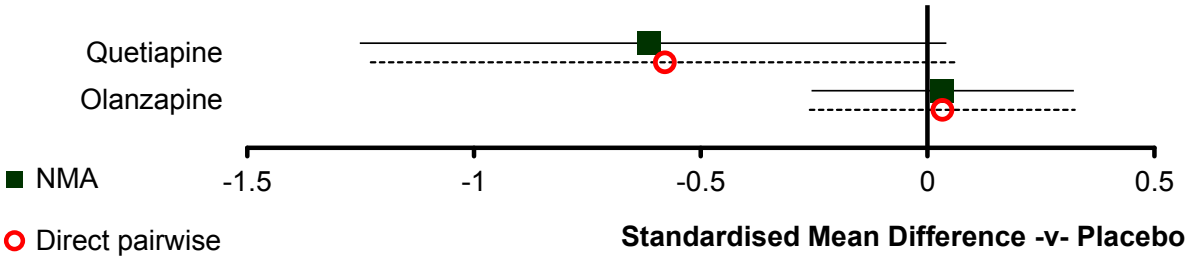
Hallucinations (multiple scales pooled) – evidence network

Hallucinations (multiple scales pooled) – input data

Study	Scale	Placebo	Olanzapine	Quetiapine
Ondo et al. (2002)	Bespoke interview	-2.80 (4.18)	-3.50 (5.94)	
Fernandez et al. (2009)	BPRS hallucination	-0.04 (0.82)		-1.32 (1.13)
Breier et al. (2002) – Europe	NPS hallucination	-2.70 (3.60)	-2.70 (3.30)	
Breier et al. (2002) – USA	NPS hallucination	-2.50 (2.70)	-2.10 (4.30)	
Shotbolt et al. (2009)	Baylor PD hallucination	-2.50 (5.11)		-3.30 (2.81)
Values are mean change from baseline to follow up (SD)				

Hallucinations (multiple scales pooled) – relative effectiveness of all pairwise combinations

	Placebo	Olanzapine	Quetiapine
Placebo		0.03 (-0.26, 0.32)	-0.58 (-1.23, 0.07)
Olanzapine	0.03 (-0.26, 0.32)		-
Quetiapine	-0.61 (-1.25, 0.04)	-0.65 (-1.34, 0.07)	
Values given are standardised mean differences. The segment below and to the left of the shaded cells is derived from the network meta-analysis, reflecting direct and indirect evidence of treatment effects (row versus column). The point estimate reflects the mean of the posterior distribution, and numbers in parentheses are 95% credible intervals. The segment above and to the right of the shaded cells gives pooled direct evidence (random-effects pairwise meta-analysis), where available (column versus row). Numbers in parentheses are 95% confidence intervals.			

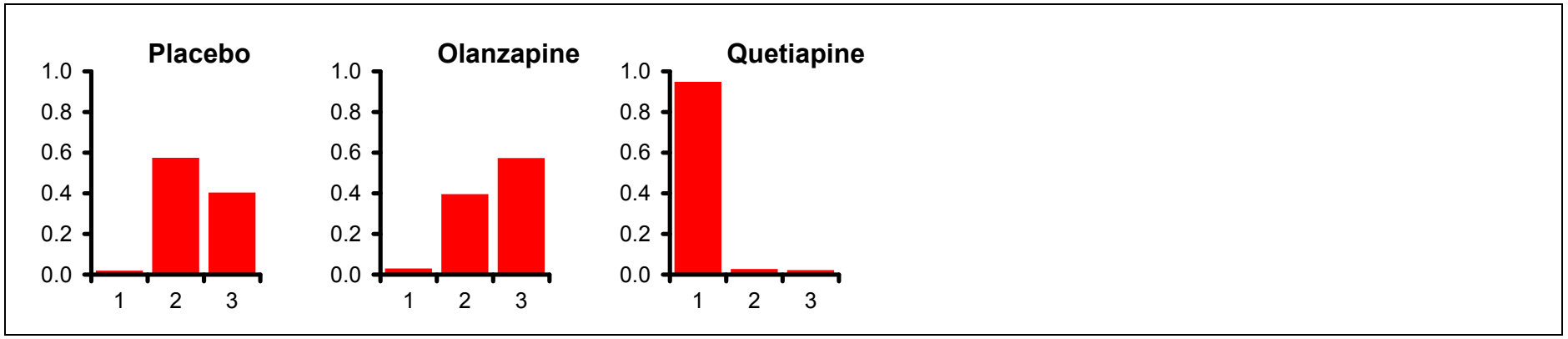


Values greater than 0 favour placebo; values less than 0 favour the comparator treatment. Solid error bars are 95% credible intervals; dashed error bars are 95% confidence interval.

Hallucinations (multiple scales pooled) – relative effect of all options versus common comparator

Hallucinations (multiple scales pooled) – rankings for each comparator

	Probability best	Median rank (95%CI)
Placebo	0.021	2 (2, 3)
Olanzapine	0.030	3 (1, 3)
Quetiapine	0.949	1 (1, 2)



Hallucinations (multiple scales pooled) – rank probability histograms

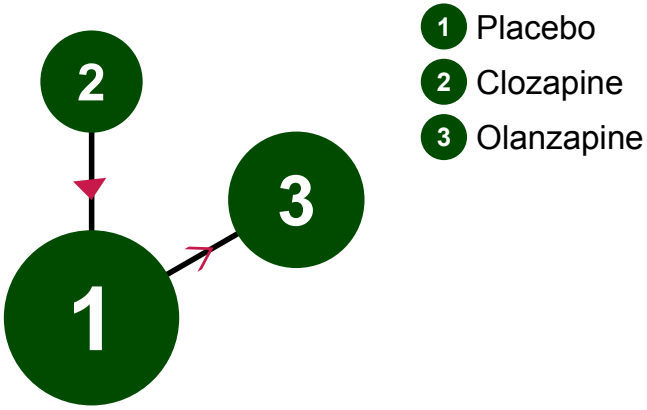
Hallucinations (multiple scales pooled) – model fit statistics

Residual deviance	Dbar	Dhat	pD	DIC
5.22 (compared to 5 datapoints)	3.703	1.721	1.981	5.684

Hallucinations (multiple scales pooled) – notes

- Continuous SMD (normal; identity link); fixed effects
- 50000 burn-ins; 10000 recorded iterations

Positive symptoms



Size of nodes is proportional to total number of participants randomised to receive the treatment in question across the evidence-base. Width of connecting lines is proportional to number of trial-level comparisons available. Arrowheads indicate direction of effect in pairwise data (a > b denotes a is more effective than b) – filled arrowheads show comparisons where one option is significantly superior (p<0.05); outlined arrowheads show direction of trend where effect does not reach statistical significance.

Positive symptoms (multiple scales pooled) – evidence network

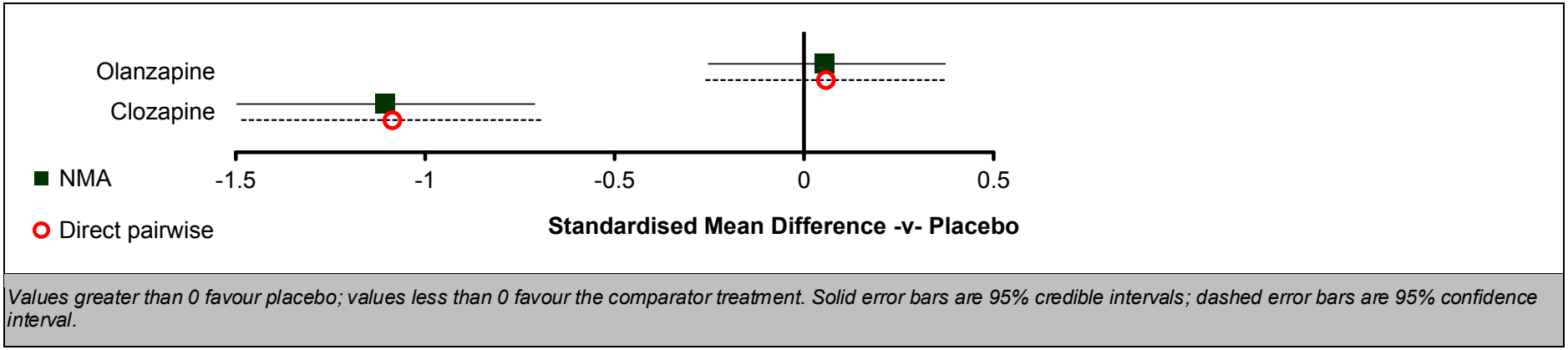
Positive symptoms (multiple scales pooled) – input data

Study	Scale	Placebo	Clozapine	Olanzapine
Friedman (1999)	SAPS	-3.80 (9.87)	-11.80 (10.39)	
Pollak et al. (2004)	Positive PANSS	-0.80 (2.80)	-5.60 (3.90)	

Study	Scale	Placebo	Clozapine	Olanzapine
Breier et al. (2002) – Europe	BPRS Positive	-2.90 (3.40)		-2.30 (4.10)
Breier et al. (2002) – USA	BPRS Positive	-1.60 (3.90)		-1.70 (3.50)
Values are mean change from baseline to follow up (SD)				

Positive symptoms (multiple scales pooled) – relative effectiveness of all pairwise combinations

	Placebo	Clozapine	Olanzapine
Placebo		-1.09 (-1.48, -0.69)	0.06 (-0.26, 0.37)
Clozapine	-1.11 (-1.50, -0.71)		-
Olanzapine	0.06 (-0.25, 0.37)	1.16 (0.66, 1.67)	
Values given are standardised mean differences. The segment below and to the left of the shaded cells is derived from the network meta-analysis, reflecting direct and indirect evidence of treatment effects (row versus column). The point estimate reflects the mean of the posterior distribution, and numbers in parentheses are 95% credible intervals. The segment above and to the right of the shaded cells gives pooled direct evidence (random-effects pairwise meta-analysis), where available (column versus row). Numbers in parentheses are 95% confidence intervals.			

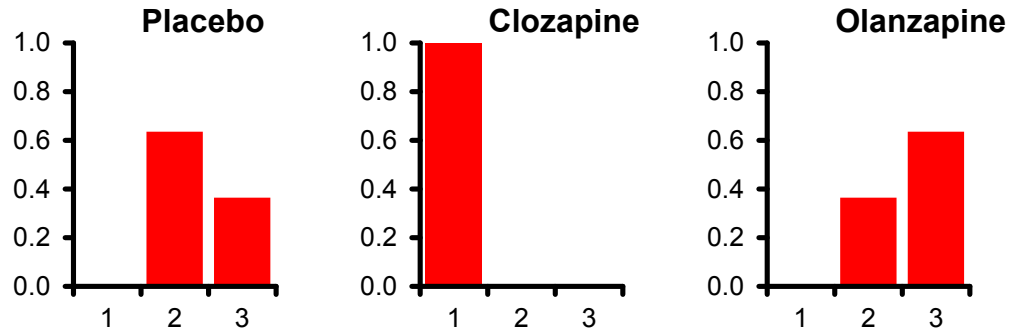


Values greater than 0 favour placebo; values less than 0 favour the comparator treatment. Solid error bars are 95% credible intervals; dashed error bars are 95% confidence interval.

Positive symptoms (multiple scales pooled) – relative effect of all options versus common comparator

Positive symptoms (multiple scales pooled) – rankings for each comparator

	Probability best	Median rank (95%CI)
Placebo	0.000	2 (2, 3)
Clozapine	1.000	1 (1, 1)
Olanzapine	0.000	3 (2, 3)



Positive symptoms (multiple scales pooled) – rank probability histograms

Positive symptoms (multiple scales pooled) – model fit statistics

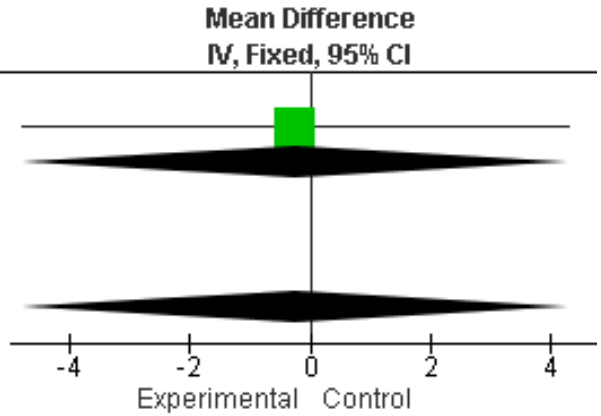
Residual deviance	Dbar	Dhat	pD	DIC
4.624 (compared to 4 datapoints)	1.071	-0.91	1.981	3.053

Positive symptoms (multiple scales pooled) – notes

- Continuous SMD (normal; identity link); fixed effects
- 50000 burn-ins; 10000 recorded iterations

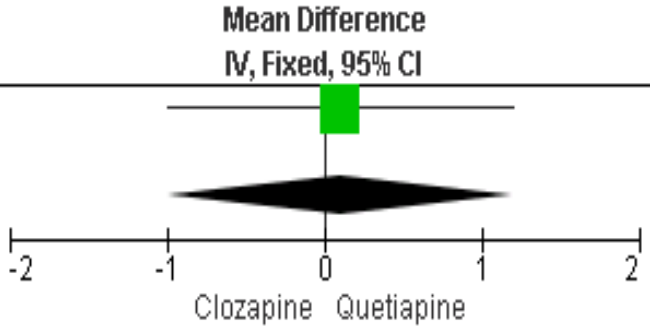
BPRS psychosis

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
20.1.1 Olanzapine								
Nichols MJ et al., 2013	7.75	4.97	9	8	4.9	9	100.0%	-0.25 [-4.81, 4.31]
Subtotal (95% CI)			**9**			**9**	**100.0%**	**-0.25 [-4.81, 4.31]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.11 (P = 0.91)								
Total (95% CI)			**9**			**9**	**100.0%**	**-0.25 [-4.81, 4.31]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.11 (P = 0.91)								
Test for subgroup differences: Not applicable								



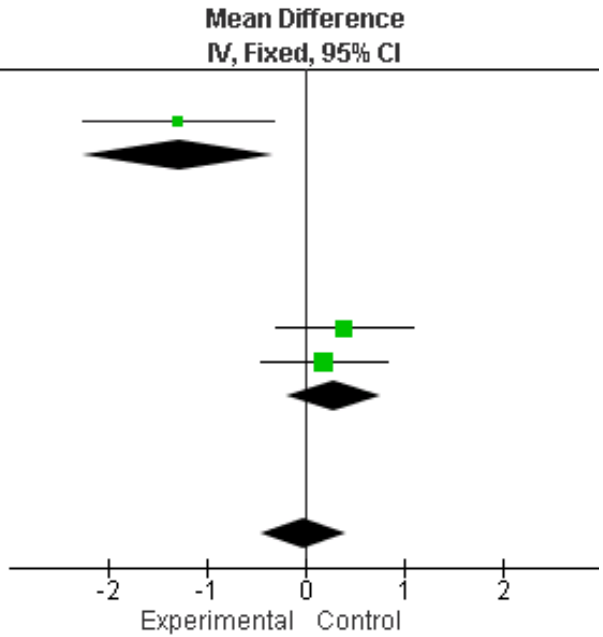
BPRS psychosis – Clozapine vs. Quetiapine

Study or Subgroup	Clozapine Mean	SD	Total	Quetiapine Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Morgante L et al., 2004	8.5	2	20	8.4	1.5	20	100.0%	0.10 [-1.00, 1.20]
Total (95% CI)			**20**			**20**	**100.0%**	**0.10 [-1.00, 1.20]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.18 (P = 0.86)								



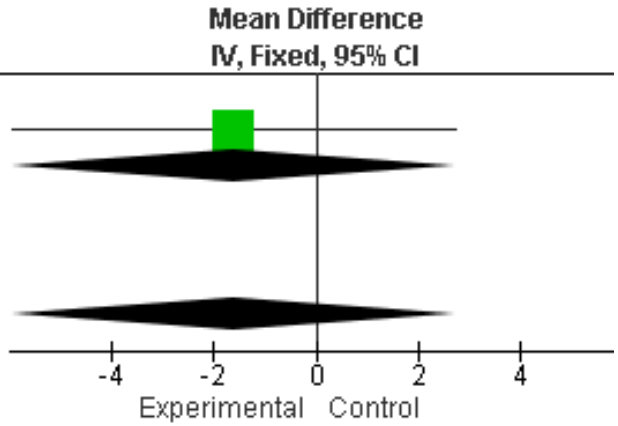
BPRS hallucination

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
20.1.1 Quetiapine								
Fernandez HH et al., 2009	-1.32	1.13	8	-0.04	0.82	8	19.4%	-1.28 [-2.25, -0.31]
Subtotal (95% CI)			**8**			**8**	**19.4%**	**-1.28 [-2.25, -0.31]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.59 (P = 0.010)								
20.1.2 Olanzapine								
Breier A et al., 2002 EU	-1	1.5	49	-1.4	1.5	28	37.4%	0.40 [-0.30, 1.10]
Breier A et al., 2002 US	-0.7	1.6	41	-0.9	1.4	42	43.3%	0.20 [-0.45, 0.85]
Subtotal (95% CI)			**90**			**70**	**80.6%**	**0.29 [-0.18, 0.77]**
Heterogeneity: $Chi^2 = 0.17$, df = 1 (P = 0.68); $I^2 = 0\%$								
Test for overall effect: Z = 1.21 (P = 0.23)								
Total (95% CI)			**98**			**78**	**100.0%**	**-0.01 [-0.44, 0.41]**
Heterogeneity: $Chi^2 = 8.35$, df = 2 (P = 0.02); $I^2 = 76\%$								
Test for overall effect: Z = 0.05 (P = 0.96)								
Test for subgroup differences: $Chi^2 = 8.18$, df = 1 (P = 0.004), $I^2 = 87.8\%$								



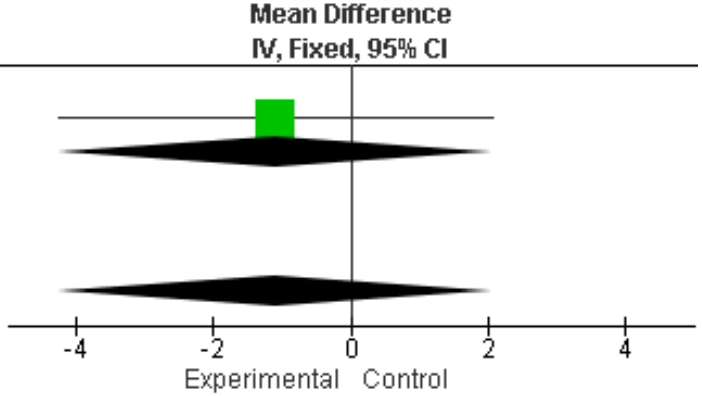
Structured interview for hallucinations in PD

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
20.1.1 Olanzapine								
Ondo WG et al., 2002	9.5	6.8	16	11.1	4.7	11	100.0%	-1.60 [-5.94, 2.74]
Subtotal (95% CI)			**16**			**11**	**100.0%**	**-1.60 [-5.94, 2.74]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.72 (P = 0.47)								
Total (95% CI)			**16**			**11**	**100.0%**	**-1.60 [-5.94, 2.74]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.72 (P = 0.47)								
Test for subgroup differences: Not applicable								



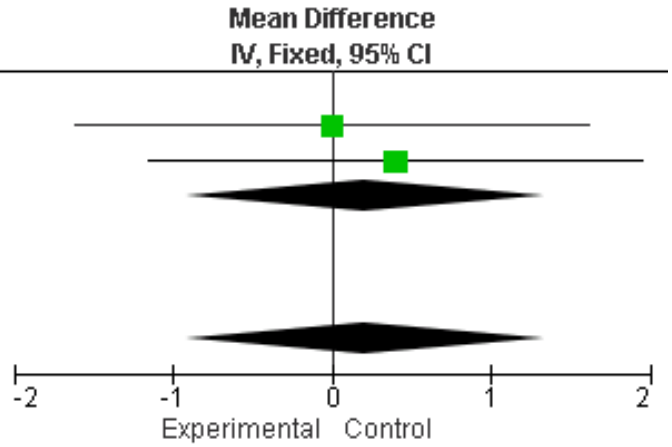
Baylor PD hallucination

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
20.1.1 Quetiapine								
Shotbolt P et al., 2009	8.3	2.9	11	9.4	4.9	13	100.0%	-1.10 [-4.27, 2.07]
Subtotal (95% CI)			**11**			**13**	**100.0%**	**-1.10 [-4.27, 2.07]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.68 (P = 0.50)								
Total (95% CI)			**11**			**13**	**100.0%**	**-1.10 [-4.27, 2.07]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.68 (P = 0.50)								
Test for subgroup differences: Not applicable								



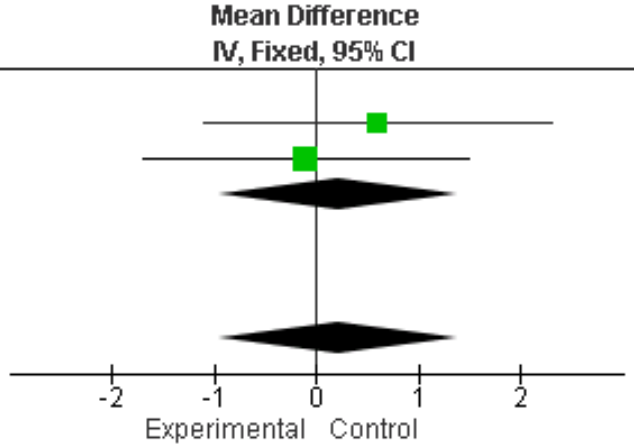
NPI hallucination

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
20.1.1 Olanzapine								
Breier A et al., 2002 EU	-2.7	3.3	49	-2.7	3.6	28	47.7%	0.00 [-1.62, 1.62]
Breier A et al., 2002 US	-2.1	4.3	41	-2.5	2.7	42	52.3%	0.40 [-1.15, 1.95]
Subtotal (95% CI)			**90**			**70**	**100.0%**	**0.21 [-0.91, 1.33]**
Heterogeneity: Chi² = 0.12, df = 1 (P = 0.73); I² = 0%								
Test for overall effect: Z = 0.37 (P = 0.71)								
Total (95% CI)			**90**			**70**	**100.0%**	**0.21 [-0.91, 1.33]**
Heterogeneity: Chi² = 0.12, df = 1 (P = 0.73); I² = 0%								
Test for overall effect: Z = 0.37 (P = 0.71)								
Test for subgroup differences: Not applicable								



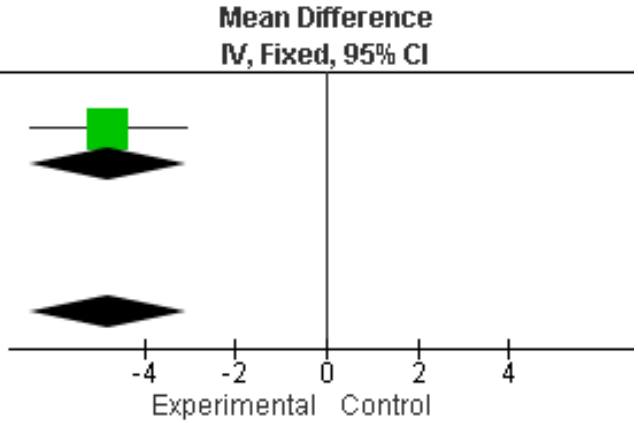
BPRS positive

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
20.1.1 Olanzapine								
Breier A et al., 2002 EU	-2.3	4.1	49	-2.9	3.4	28	46.6%	0.60 [-1.10, 2.30]
Breier A et al., 2002 US	-1.7	3.5	41	-1.6	3.9	42	53.4%	-0.10 [-1.69, 1.49]
Subtotal (95% CI)			**90**			**70**	**100.0%**	**0.23 [-0.94, 1.39]**
Heterogeneity: Chi² = 0.35, df = 1 (P = 0.56); I² = 0%								
Test for overall effect: Z = 0.38 (P = 0.70)								
Total (95% CI)			**90**			**70**	**100.0%**	**0.23 [-0.94, 1.39]**
Heterogeneity: Chi² = 0.35, df = 1 (P = 0.56); I² = 0%								
Test for overall effect: Z = 0.38 (P = 0.70)								
Test for subgroup differences: Not applicable								



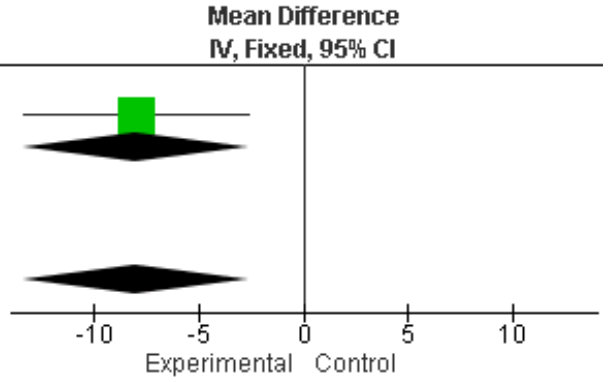
PANSS positive

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
20.1.1 Clozapine								
Pollak P et al., 2004	-5.6	3.9	32	-0.8	2.8	28	100.0%	-4.80 [-6.50, -3.10]
Subtotal (95% CI)			**32**			**28**	**100.0%**	**-4.80 [-6.50, -3.10]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 5.52 (P < 0.00001)								
Total (95% CI)			**32**			**28**	**100.0%**	**-4.80 [-6.50, -3.10]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 5.52 (P < 0.00001)								
Test for subgroup differences: Not applicable								



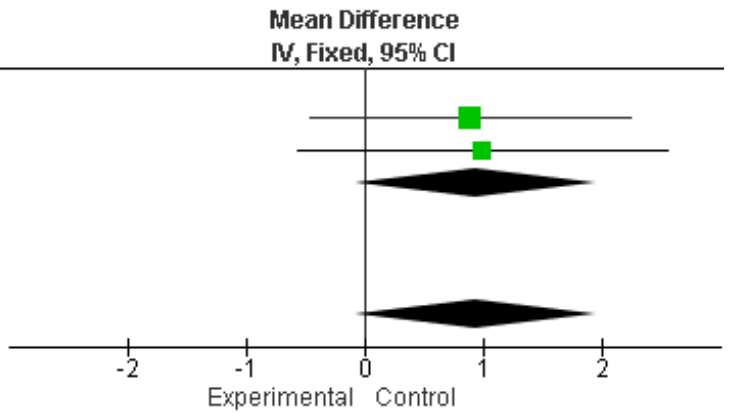
SAPS

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
20.1.1 Clozapine								
Friedman et al., 1999	-11.8	10.3923	27	-3.8	9.87269	27	100.0%	-8.00 [-13.41, -2.59]
Subtotal (95% CI)			**27**			**27**	**100.0%**	**-8.00 [-13.41, -2.59]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.90 (P = 0.004)								
Total (95% CI)			**27**			**27**	**100.0%**	**-8.00 [-13.41, -2.59]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.90 (P = 0.004)								
Test for subgroup differences: Not applicable								

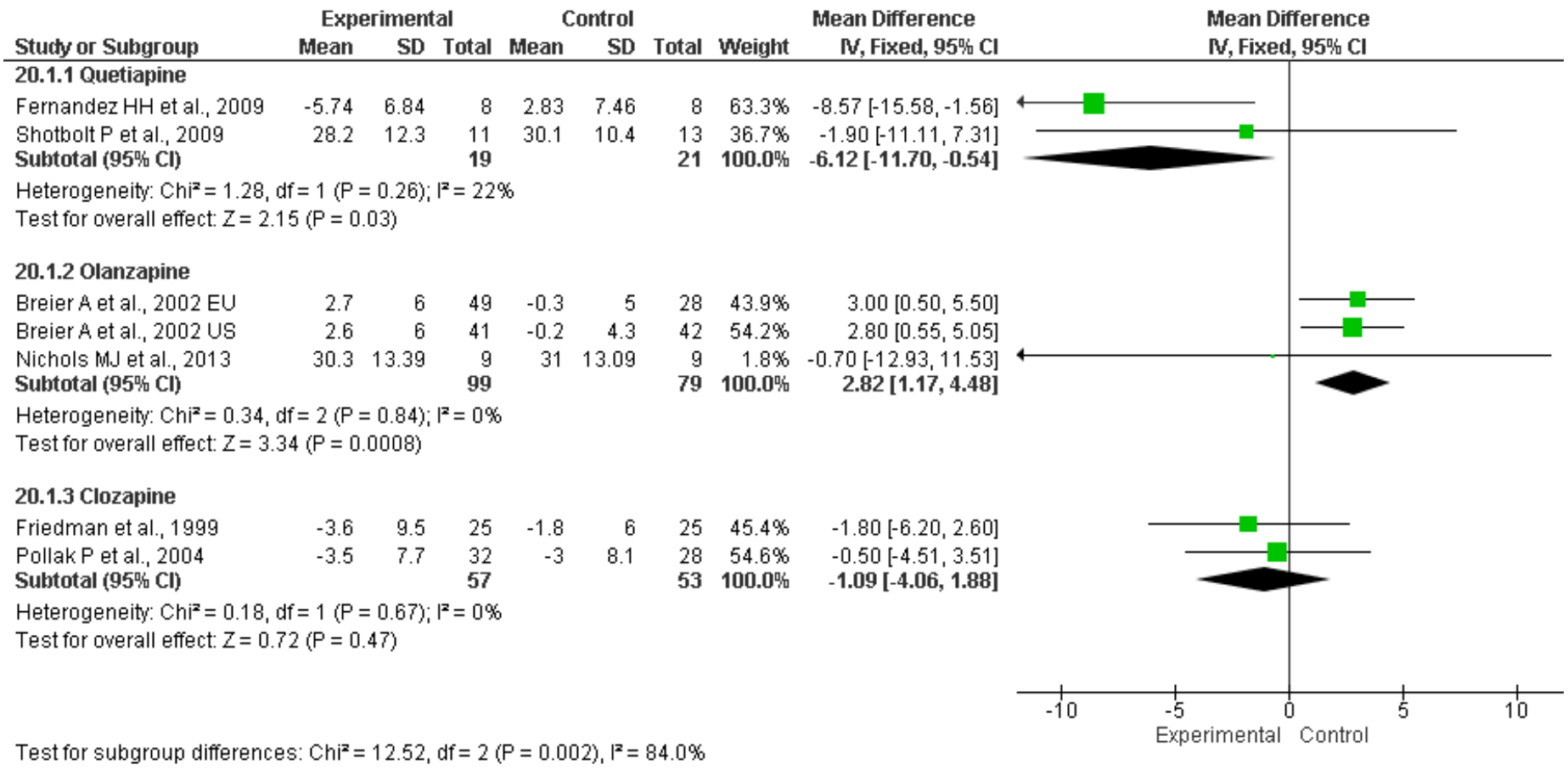


NPI delusions

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
20.1.1 Olanzapine								
Breier A et al., 2002 EU	-1.1	3.4	49	-2	2.6	28	56.8%	0.90 [-0.45, 2.25]
Breier A et al., 2002 US	-0.7	3.3	41	-1.7	3.9	42	43.2%	1.00 [-0.55, 2.55]
Subtotal (95% CI)			**90**			**70**	**100.0%**	**0.94 [-0.08, 1.96]**
Heterogeneity: Chi² = 0.01, df = 1 (P = 0.92); I² = 0%								
Test for overall effect: Z = 1.81 (P = 0.07)								
Total (95% CI)			**90**			**70**	**100.0%**	**0.94 [-0.08, 1.96]**
Heterogeneity: Chi² = 0.01, df = 1 (P = 0.92); I² = 0%								
Test for overall effect: Z = 1.81 (P = 0.07)								
Test for subgroup differences: Not applicable								

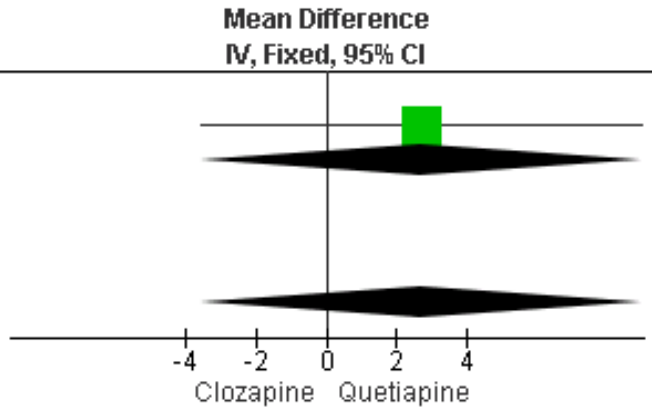


UPDRS motor


Experimental
Control
Mean Difference
Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Fixed, 95% CI
IV, Fixed, 95% CI
20.1.1 Quetiapine
Fernandez HH et al., 2009 -5.74 6.84 8 2.83 7.46 8 63.3% -8.57 [-15.58, -1.56]
Shotbolt P et al., 2009 28.2 12.3 11 30.1 10.4 13 36.7% -1.90 [-11.11, 7.31]
Subtotal (95% CI) 19 21 100.0% -6.12 [-11.70, -0.54]
Heterogeneity: Chi² = 1.28, df = 1 (P = 0.26); I² = 22%
Test for overall effect: Z = 2.15 (P = 0.03)
20.1.2 Olanzapine
Breier A et al., 2002 EU 2.7 6 49 -0.3 5 28 43.9% 3.00 [0.50, 5.50]
Breier A et al., 2002 US 2.6 6 41 -0.2 4.3 42 54.2% 2.80 [0.55, 5.05]
Nichols MJ et al., 2013 30.3 13.39 9 31 13.09 9 1.8% -0.70 [-12.93, 11.53]
Subtotal (95% CI) 99 79 100.0% 2.82 [1.17, 4.48]
Heterogeneity: Chi² = 0.34, df = 2 (P = 0.84); I² = 0%
Test for overall effect: Z = 3.34 (P = 0.0008)
20.1.3 Clozapine
Friedman et al., 1999 -3.6 9.5 25 -1.8 6 25 45.4% -1.80 [-6.20, 2.60]
Pollak P et al., 2004 -3.5 7.7 32 -3 8.1 28 54.6% -0.50 [-4.51, 3.51]
Subtotal (95% CI) 57 53 100.0% -1.09 [-4.06, 1.88]
Heterogeneity: Chi² = 0.18, df = 1 (P = 0.67); I² = 0%
Test for overall effect: Z = 0.72 (P = 0.47)
Test for subgroup differences: Chi² = 12.52, df = 2 (P = 0.002), I² = 84.0%
-10
-5
0
5
10
Experimental
Control

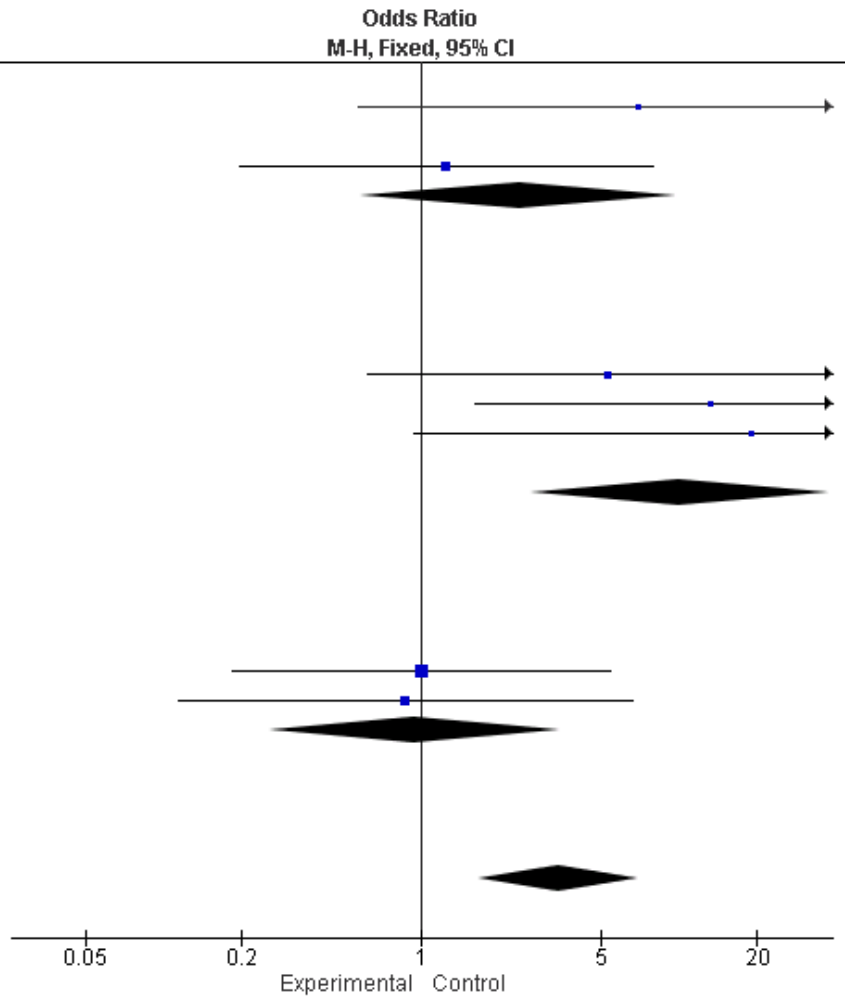
UPDRS motor – Clozapine vs. Quetiapine

Study or Subgroup	Clozapine Mean	SD	Total	Quetiapine Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
20.4.4 Clozapine vs. Quetiapine								
Morgante L et al., 2004	56.7	9.2	20	54	11	20	100.0%	2.70 [-3.58, 8.98]
Subtotal (95% CI)			**20**			**20**	**100.0%**	**2.70 [-3.58, 8.98]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.84 (P = 0.40)								
Total (95% CI)			**20**			**20**	**100.0%**	**2.70 [-3.58, 8.98]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.84 (P = 0.40)								
Test for subgroup differences: Not applicable								



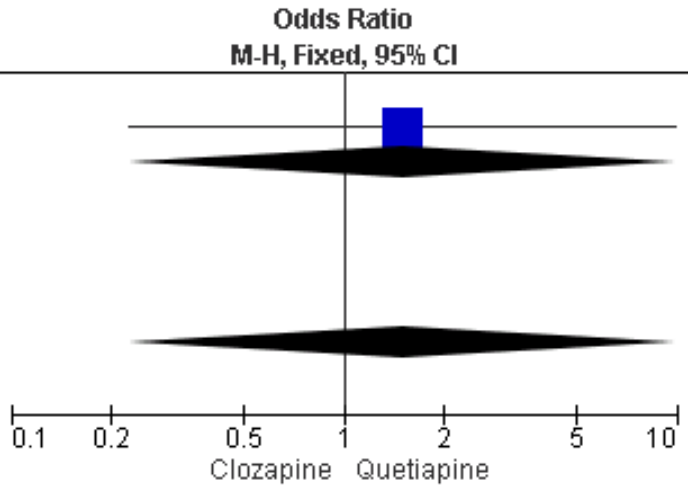
Treatment discontinuation due to AEs

Study or Subgroup	Experimental Events	Experimental Total	Control Events	Control Total	Weight	Odds Ratio M-H, Fixed, 95% CI
20.1.1 Quetiapine						
Fernandez HH et al., 2009	4	8	1	8	5.4%	7.00 [0.57, 86.32]
Ondo WG et al., 2005	0	21	0	10		Not estimable
Shotbolt P et al., 2009	3	11	3	13	21.5%	1.25 [0.20, 7.96]
Subtotal (95% CI)		**40**		**31**	**26.8%**	**2.40 [0.58, 9.87]**
Total events	7		4			
Heterogeneity: Chi² = 1.17, df = 1 (P = 0.28); I² = 15%						
Test for overall effect: Z = 1.21 (P = 0.22)						
20.1.2 Olanzapine						
Breier A et al., 2002 EU	8	49	1	28	11.4%	5.27 [0.62, 44.55]
Breier A et al., 2002 US	10	41	1	42	8.0%	13.23 [1.61, 108.86]
Nichols MJ et al., 2013	7	14	0	9	3.2%	19.00 [0.93, 388.77]
Ondo WG et al., 2002	0	18	0	12		Not estimable
Subtotal (95% CI)		**122**		**91**	**22.7%**	**10.03 [2.64, 38.13]**
Total events	25		2			
Heterogeneity: Chi² = 0.59, df = 2 (P = 0.75); I² = 0%						
Test for overall effect: Z = 3.39 (P = 0.0007)						
20.1.3 Clozapine						
Friedman et al., 1999	3	30	3	30	29.0%	1.00 [0.19, 5.40]
Pollak P et al., 2004	2	32	2	28	21.5%	0.87 [0.11, 6.59]
Subtotal (95% CI)		**62**		**58**	**50.5%**	**0.94 [0.26, 3.45]**
Total events	5		5			
Heterogeneity: Chi² = 0.01, df = 1 (P = 0.92); I² = 0%						
Test for overall effect: Z = 0.09 (P = 0.93)						
Total (95% CI)		**224**		**180**	**100.0%**	**3.40 [1.67, 6.91]**
Total events	37		11			
Heterogeneity: Chi² = 8.21, df = 6 (P = 0.22); I² = 27%						
Test for overall effect: Z = 3.38 (P = 0.0007)						
Test for subgroup differences: Chi² = 6.26, df = 2 (P = 0.04), I² = 68.0%						



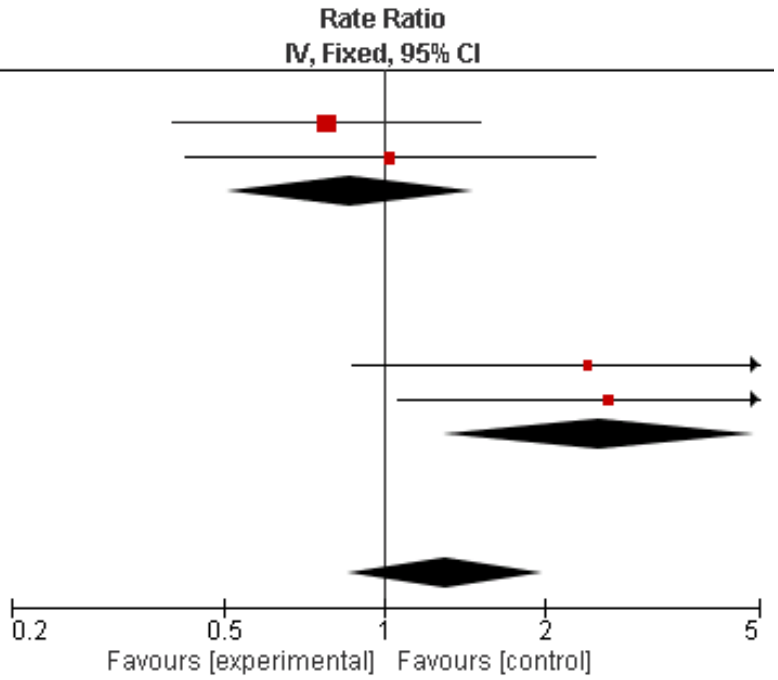
Treatment discontinuation due to AEs – Clozapine vs. Quetiapine

Study or Subgroup	Experimental Events	Experimental Total	Control Events	Control Total	Weight	Odds Ratio M-H, Fixed, 95% CI
20.5.4 Clozapine vs. Quetiapine						
Morgante L et al., 2004	3	23	2	22	100.0%	1.50 [0.23, 9.96]
Subtotal (95% CI)		**23**		**22**	**100.0%**	**1.50 [0.23, 9.96]**
Total events	3		2			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.42 (P = 0.67)						
Total (95% CI)		**23**		**22**	**100.0%**	**1.50 [0.23, 9.96]**
Total events	3		2			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.42 (P = 0.67)						
Test for subgroup differences: Not applicable						



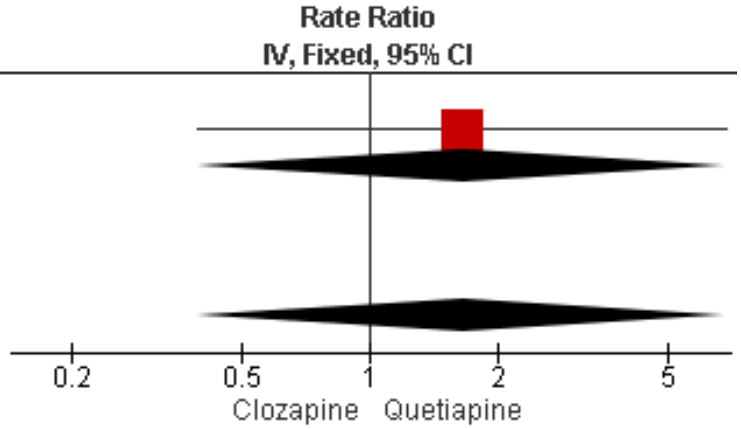
Adverse events (rate)

Study or Subgroup	log[Rate Ratio]	SE	Weight	Rate Ratio IV, Fixed, 95% CI
20.2.1 Quetiapine				
Ondo WG et al., 2005	-0.25078	0.338979	39.3%	0.78 [0.40, 1.51]
Fernandez HH et al., 2009	0.022473	0.449467	22.4%	1.02 [0.42, 2.47]
Subtotal (95% CI)			**61.7%**	**0.86 [0.51, 1.46]**
Heterogeneity: Chi² = 0.24, df = 1 (P = 0.63); I² = 0%				
Test for overall effect: Z = 0.56 (P = 0.58)				
20.2.2 Olanzapine				
Nichols MJ et al., 2013	0.875469	0.516398	17.0%	2.40 [0.87, 6.60]
Ondo WG et al., 2002	0.962811	0.460566	21.3%	2.62 [1.06, 6.46]
Subtotal (95% CI)			**38.3%**	**2.52 [1.28, 4.94]**
Heterogeneity: Chi² = 0.02, df = 1 (P = 0.90); I² = 0%				
Test for overall effect: Z = 2.69 (P = 0.007)				
Total (95% CI)			**100.0%**	**1.30 [0.85, 1.97]**
Heterogeneity: Chi² = 6.30, df = 3 (P = 0.10); I² = 52%				
Test for overall effect: Z = 1.22 (P = 0.22)				
Test for subgroup differences: Chi² = 6.05, df = 1 (P = 0.01), I² = 83.5%				



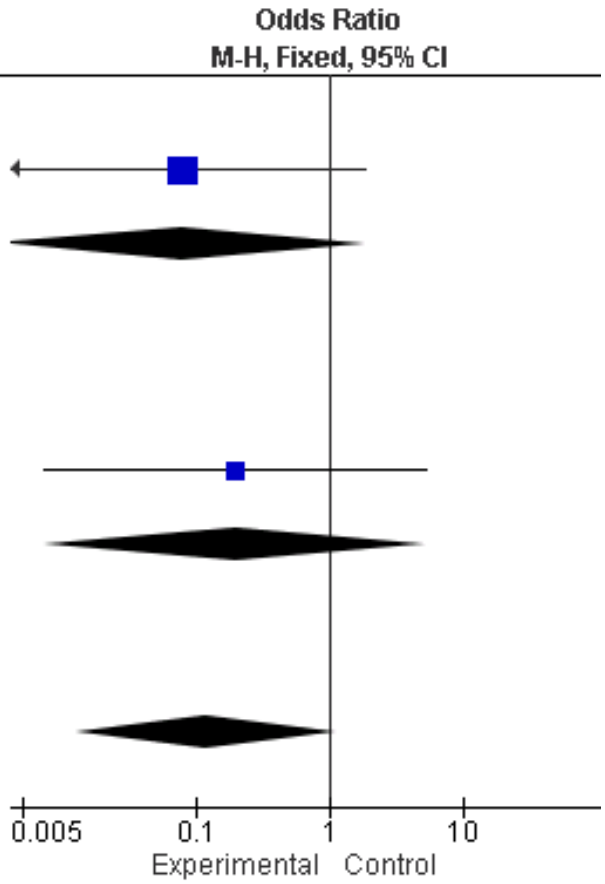
Adverse events (rate) – Clozapine vs. Quetiapine

Study or Subgroup	log[Rate Ratio]	SE	Weight	Rate Ratio IV, Fixed, 95% CI
20.6.4 Clozapine vs. Quetiapine				
Morgante L et al., 2004	0.498556	0.730297	100.0%	1.65 [0.39, 6.89]
Subtotal (95% CI)			**100.0%**	**1.65 [0.39, 6.89]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 0.68 (P = 0.49)				
Total (95% CI)			**100.0%**	**1.65 [0.39, 6.89]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 0.68 (P = 0.49)				
Test for subgroup differences: Not applicable				



Mortality

Study or Subgroup	Experimental Events	Experimental Total	Control Events	Control Total	Weight	Odds Ratio M-H, Fixed, 95% CI
20.1.1 Quetiapine						
Fernandez HH et al., 2009	0	8	0	8		Not estimable
Ondo WG et al., 2005	0	21	2	10	65.2%	0.08 [0.00, 1.82]
Shotbolt P et al., 2009	0	11	0	13		Not estimable
Subtotal (95% CI)		**40**		**31**	**65.2%**	**0.08 [0.00, 1.82]**
Total events	0		2			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.58 (P = 0.11)						
20.1.2 Olanzapine						
Nichols MJ et al., 2013	0	14	1	9	34.8%	0.20 [0.01, 5.35]
Ondo WG et al., 2002	0	18	0	12		Not estimable
Subtotal (95% CI)		**32**		**21**	**34.8%**	**0.20 [0.01, 5.35]**
Total events	0		1			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.97 (P = 0.33)						
Total (95% CI)		**72**		**52**	**100.0%**	**0.12 [0.01, 1.14]**
Total events	0		3			
Heterogeneity: Chi² = 0.15, df = 1 (P = 0.70); I² = 0%						
Test for overall effect: Z = 1.85 (P = 0.06)						
Test for subgroup differences: Chi² = 0.15, df = 1 (P = 0.70), I² = 0%						



BPRS Psychosis - Olanzapine

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
Average CI score change									
1 study: Nichols et al., 2013	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[4]	9	9	-0.25 (-4.81, 4.31)	LOW

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

BPRS Psychosis - Clozapine vs. Quetiapine

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
Average CI score change									
1 study: Morgante et al., 2004	RCT	Serious[1]	N/A[2]	Not serious[3]	serious[4]	20	20	0.1 (-1, 1.2)	LOW

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

BPRS Hallucination – Quetiapine

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
Average CI score change									
1 study:	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[6]	8	8	-1.28 (-2.25, -0.31)	LOW

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	Quality
Fernandez et al., 2009									

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results
[5] Serious imprecision: CI cross MID between 3.25 (Horvath et al., 2015) and 5 points (Schrag et al., 2006)
[6] Very small sample size

BPRS Hallucination – Olanzapine

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	Quality
Average CI score change									
2 studies: Breier et al., 2002 – EU study Breier et al., 2002 – US study	RCT	Serious[1]	Not serious	Not serious[3]	Serious[4]	90	70	0.29 (-0.18, 0.77)	LOW

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

NPI hallucination – Olanzapine

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
Average CI score change									

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
2 studies: Breier et al., 2002 – EU study Breier et al., 2002 – US study	RCT	Serious[1]	Not serious	Not serious[3]	Serious[4]	90	70	0.21 (-0.91, 1.33)	LOW

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

Baylor PD Hallucination – Quetiapine

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
Average CI score change									
1 study: Shotbolt et al., 2009	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[4]	11	13	-1.1 (-4.27, 2.07)	LOW

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

Structured interview for hallucinations in PD – Olanzapine

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
Average CI score change									
1 study:	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[4]	16	11	-1.6 (-5.94, 2.74)	LOW

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
Ondo et al., 2002									

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

BPRS Positive – Olanzapine

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
Average CI score change									
2 studies: Breier et al., 2002 – EU study Breier et al., 2002 – US study	RCT	Serious[1]	Not serious	Not serious[3]	Serious[4]	90	70	0.23 (-0.94, 1.39)	LOW

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

Positive PANSS – Clozapine

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
Average CI score change									
1 study: Pollak et al., 2004	RCT	Serious[1]	N/A[2]	Not serious[3]	Not serious	32	28	-4.8 (-6.5, -3.1)	MODERATE

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol

SAPS – Clozapine

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
Average CI score change									
1 study: Friedman et al., 1999	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[4]	27	27	-8 (-13.41, - 2.59)	LOW

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

NPI Delusions – Olanzapine

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
Average CI score change									
2 studies: Breier et al., 2002 – EU study Breier et al., 2002 – US study	RCT	Serious[1]	Not serious	Not serious[3]	Serious[4]	90	70	0.94 (-0.08, 1.96)	LOW

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

UPDRS Motor – Quetiapine

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
UPDRS Motor - Quetiapine (Better indicated by lower values)									
2 studies: Fernandez et al., 2009 Shotbolt et al., 2009	RCT	Serious[1]	Not serious	Not serious[3]	Serious[5]	19	21	-6.12 (-11.7, - 0.54)	LOW

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results
[5] Serious imprecision: CI cross MID between 3.25 (Horvath et al., 2015) and 5 points (Schrag et al., 2006)

UPDRS Motor - Olanzapine

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
Average CI score change									
3 studies: Breier A et al., 2002 - EU study Breier A et al., 2002 – US study Nichols et al., 2013	RCT	Serious[1]	Not serious	Not serious[3]	Serious[5]	99	79	2.82 (1.17, 4.48)	LOW

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results
[5] Serious imprecision: CI cross MID between 3.25 (Horvath et al., 2015) and 5 points (Schrag et al., 2006)

UPDRS Motor – Clozapine

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
Average CI score change									
2 studies: Friedman et al., 1999 Pollak et al., 2004	RCT	Serious[1]	Not serious	Not serious[3]	Serious[5]	57	53	-1.09 (-4.06, 1.88)	LOW

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results
[5] Serious imprecision: CI cross MID between 3.25 (Horvath et al., 2015) and 5 points (Schrag et al., 2006)

UPDRS Motor - Clozapine vs. Quetiapine

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Mean Difference (95% CI)	
Average CI score change									
1 study: Morgante et al., 2004	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[5]	20	20	2.7 (-3.58, 8.98)	LOW

Dropouts due to AEs – Quetiapine

Quality assessment						No of events/ Total no of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Odds Ratio (95% CI)	
Dropouts due to AEs									
3 studies: Fernandez et al., 2009 Ondo et al., 2005 Shotbolt et al., 2005	RCT	Serious[1]	Not serious	Not serious[3]	Serious[4]	7/40	4/31	2.4 (0.58, 9.87)	LOW

Quality assessment						No of events/ Total no of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Odds Ratio (95% CI)	Quality

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

Dropouts due to AEs – Olanzapine

Quality assessment						No of events/ Total no of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Odds Ratio (95% CI)	Quality
Dropouts due to AEs									
4 studies: Breier et al., 2002 – EU Breier et al., 2002 – US Nichols et al., 2013 Ondo et al., 2002	RCT	Serious[1]	Not serious	Not serious[3]	Not serious	25/122	2/91	10.03 (2.64, 38.13)	MODERATE

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol

Dropouts due to AEs – Clozapine

Quality assessment						No of events/ Total no of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Odds Ratio (95% CI)	Quality
Dropouts due to AEs									

Quality assessment						No of events/ Total no of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Odds Ratio (95% CI)	
2 studies: Friedman et al., 1999 Pollak et al., 2014	RCT	Serious[1]	Not serious	Not serious[3]	Serious[4]	5/62	5/58	0.94 (0.26 to 3.45)	LOW

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

Dropouts due to AEs - Clozapine vs. Quetiapine

Quality assessment						No of events/ Total no of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Odds Ratio (95% CI)	
Dropouts due to AEs									
1 study: Morgante et al., 2004	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[4]	3/23	2/22	1.5 (0.23, 9.96)	LOW

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

Adverse event - Estimate of rate – Quetiapine

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Rate Ratio (95% CI)	
The rate of an adverse event occurring									

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Rate Ratio (95% CI)	
2 studies: Fernandez et al., 2009 Ondo et al., 2005	RCT	Serious[1]	Not serious	Not serious[3]	Serious[4]	29	18	0.86 (0.51, 1.46)	LOW

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

Adverse event - Estimate of rate – Olanzapine

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Rate Ratio (95% CI)	
The rate of an adverse event occurring									
2 studies: Nichols et al., 2013 Ondo et al., 2002	RCT	Serious[1]	Not serious[5]	Not serious[3]	Not serious	31	21	2.52 (1.28, 4.94)	MODERATE

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

Adverse event - Estimate of rate - Clozapine vs. Quetiapine

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Rate Ratio (95% CI)	
The rate of an adverse event occurring									
1 study:	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[4]	23	22	1.65 (0.39, 6.89)	LOW

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Rate Ratio (95% CI)	Quality
Morgante et al., 2004									

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

Mortality - Quetiapine

Quality assessment						No of events/ Total no of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Odds Ratio (95% CI)	Quality
Mortality									
3 studies: Fernandez et al., 2009 Ondo et al., 2005 Shotbolt et al., 2009	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[4]	0/40	2/31	OR 0.08 (0, 1.82)	LOW

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

Mortality – Olanzapine

Quality assessment						No of events/ Total no of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Odds Ratio (95% CI)	Quality
Mortality - Olanzapine									
2 studies:	RCT	Serious[1]	N/A[2]	Not serious[3]	Serious[4]	0/32	1/21	OR 0.2 (0.01, 5.35)	LOW

Quality assessment						No of events/ Total no of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Odds Ratio (95% CI)	Quality
Nichols et al., 2013 Ondo et al., 2002									

[1] Serious risk of bias as assessed by NICE RCT quality checklist
[2] N/A; Not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, population was as specified in review protocol
[4] Non-significant results

E.3.5 REM sleep disorder behaviour

Rivastigmine effects on RBD sleep disorder in Parkinson's disease

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Rivastigmine	placebo	Median difference (25th - 75th %ile)	Quality
Frequency of RBD episodes									
Di Giacomo 2012	RCT	Serious[1]	NA[2]	Not serious	Serious[3]	12	12	2.5 (0.0 to 4.5)	LOW

[1]Very serious risk of bias as assessed by NICE RCT quality checklist; [2]N/A: only 1 study contributed to the analysis; [3]Study number is very small

Rivastigmine for the treatment of RBD sleep disorder: Serious adverse events

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Rivastigmine	placebo	Number of adverse events leading to discontinuation	Quality
Adverse events leading to study discontinuation in rivastigmine group									
Di Giacomo 2012	RCT	Serious[1]	NA[2]	Not serious	Serious[3]	12	12	2	LOW
Adverse events leading to study discontinuation in placebo group									

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Rivastigmine	placebo	Number of adverse events leading to discontinuation	Quality
Di Giacomo 2012	RCT	Serious[1]	NA[2]	Not serious	Serious[3]	12	12	0	LOW

[1]Very serious risk of bias as assessed by NICE RCT quality checklist; [2]N/A: only 1 study contributed to the analysis; [3]Study number is very small

E.3.6 Thermoregulatory dysfunction

None

E.4 Pharmacological management of dementia associated with Parkinson's disease

Parkinson's disease dementia – cholinesterase inhibitors

PDD – cholinesterase inhibitor vs. placebo: adverse events

Quality assessment						No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	ChI	Placebo	Relative (95% CI)	Absolute (95% CI)	
Any adverse events – cholinesterase inhibitors (probability of experiencing ≥1; follow-up 10 to 24 weeks; lower is better); see Figure 1 for forest plot										
4[1–4]	RCT	not serious	not serious	not serious	not serious	609/774 (78.7%)	268/384 (69.8%)	RR 1.12 (1.04 to 1.21)	84 more per 1000 (from 28 more to 147 more)	⊕⊕⊕⊕ HIGH
Any adverse events – donepezil (probability of experiencing ≥1; follow-up 10 to 24 weeks; lower is better)										
3[1,2,4]	RCT	not serious	not serious	not serious	serious[5]	306/412 (74.3%)	141/205 (68.8%)	RR 1.07 (0.96 to 1.19)	48 more per 1000 (from 28 fewer to 131 more)	⊕⊕⊕O MODERATE
Any adverse events – rivastigmine (probability of experiencing ≥1; follow-up 24 weeks; lower is better)										
1[3]	RCT	not serious	N/A	not serious	not serious	303/362 (83.7%)	127/179 (70.9%)	RR 1.18 (1.06 to 1.31)	128 more per 1000 (from 43 more to 220 more)	⊕⊕⊕⊕ HIGH
Serious adverse events – cholinesterase inhibitors (probability of experiencing ≥1; follow-up 24 weeks; lower is better); see Figure 2 for forest plot										
2[2,3]	RCT	not serious	serious[6]	not serious	serious[5]	114/739 (15.4%)	48/352 (13.6%)	RR 1.13 (0.82 to 1.54)	18 more per 1000 (from 25 fewer to 74 more)	⊕⊕OO LOW
Serious adverse events – donepezil (probability of experiencing ≥1; follow-up 24 weeks; lower is better)										
1[2]	RCT	not serious	N/A	not serious	serious[5]	67/377 (17.8%)	22/173 (12.7%)	RR 1.4 (0.89 to 2.18)	51 more per 1000 (from 14 fewer to 150 more)	⊕⊕⊕O MODERATE
Serious adverse events – rivastigmine (probability of experiencing ≥1; follow-up 24 weeks; lower is better)										
1[3]	RCT	not serious	N/A	not serious	serious[5]	47/362 (13%)	26/179 (14.5%)	RR 0.89 (0.57 to 1.39)	16 fewer per 1000 (from 62 fewer to 57 more)	⊕⊕⊕O MODERATE
Adverse events requiring treatment withdrawal – cholinesterase inhibitors (probability of experiencing; follow-up 24 weeks; lower is better); see Figure 3 for forest plot										
3[1–3]	RCT	not serious	not serious	not serious	not serious	122/753 (16.2%)	33/364 (9.1%)	RR 1.76 (1.23 to 2.53)	69 more per 1000 (from 21 more to 139 more)	⊕⊕⊕⊕ HIGH
Adverse events requiring treatment withdrawal – donepezil (probability of experiencing; follow-up 24 weeks)										
2[1,2]	RCT	not serious	not serious	not serious	serious[5]	60/391 (15.3%)	19/185 (10.3%)	RR 1.46 (0.91 to 2.35)	47 more per 1000 (from 9 fewer to 139 more)	⊕⊕⊕O MODERATE
Adverse events requiring treatment withdrawal – rivastigmine (probability of experiencing; follow-up 24 weeks)										
1[3]	RCT	not serious	N/A	not serious	not serious	62/362 (17.1%)	14/179 (7.8%)	RR 2.19 (1.26 to 3.8)	93 more per 1000 (from 20 more to 219 more)	⊕⊕⊕⊕ HIGH
Hallucinations – cholinesterase inhibitors (probability of experiencing; follow-up 24 weeks; lower is better); see Figure 4 for forest plot										
2[2,3]	RCT	not serious	not serious	not serious	not serious	35/739 (4.7%)	31/352 (8.8%)	RR 0.54 (0.34 to 0.86)	41 fewer per 1000 (from 12 fewer to 58 fewer)	⊕⊕⊕⊕ HIGH
Hallucinations – donepezil (probability of experiencing; follow-up 24 weeks; lower is better)										

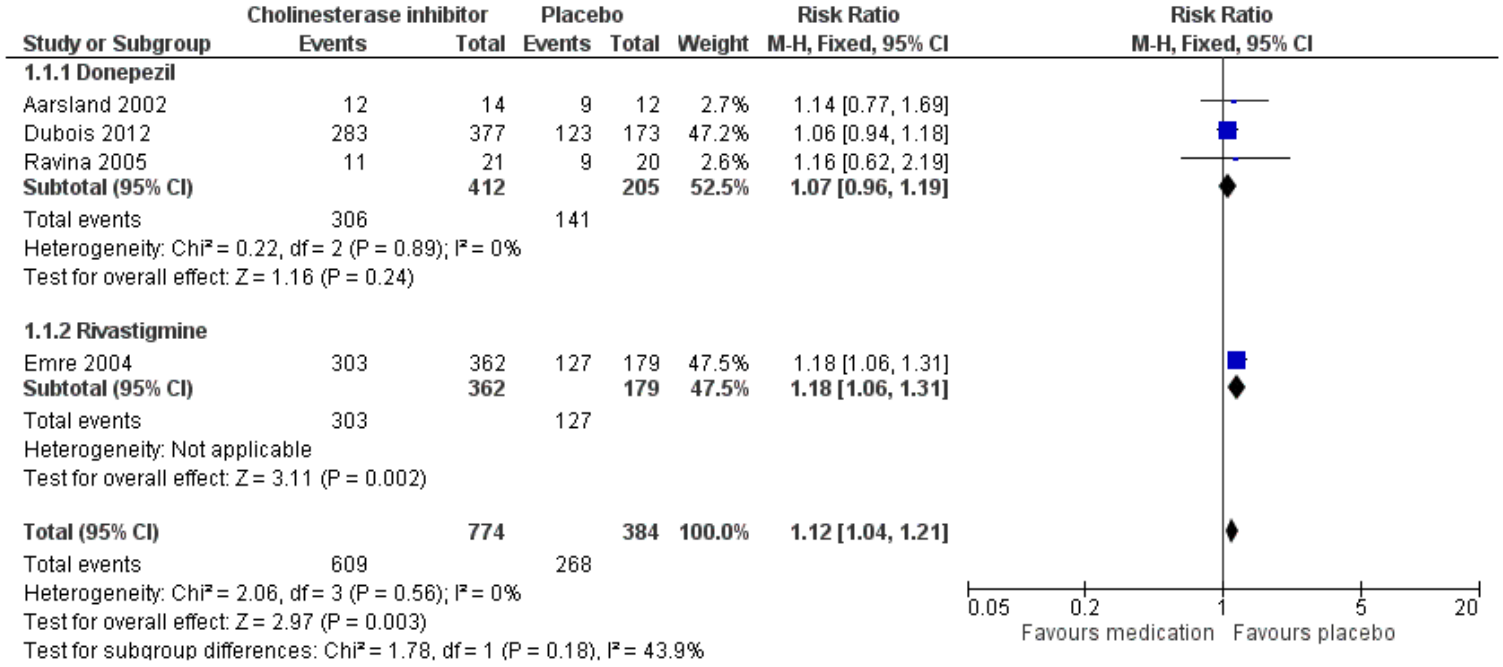
Quality assessment						No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Chl	Placebo	Relative (95% CI)	Absolute (95% CI)	
1[2]	RCT	not serious	N/A	not serious	serious[5]	18/377 (4.8%)	14/173 (8.1%)	RR 0.59 (0.3 to 1.16)	33 fewer per 1000 (from 57 fewer to 13 more)	⊕⊕⊕O MODERATE
Hallucinations – rivastigmine (probability of experiencing; follow-up 24 weeks; lower is better)										
1[3]	RCT	not serious	N/A	not serious	not serious	17/362 (4.7%)	17/179 (9.5%)	RR 0.49 (0.26 to 0.95)	48 fewer per 1000 (from 5 fewer to 70 fewer)	⊕⊕⊕⊕ HIGH

[1] Aarsland 2002
[2] Dubois 2012; data for 2 active treatment groups were combined (donepezil 5mg and 10mg)
[3] Emre 2004
[4] Ravina 2005
[5] At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit or no difference
[6] i^2 > 40% between studies

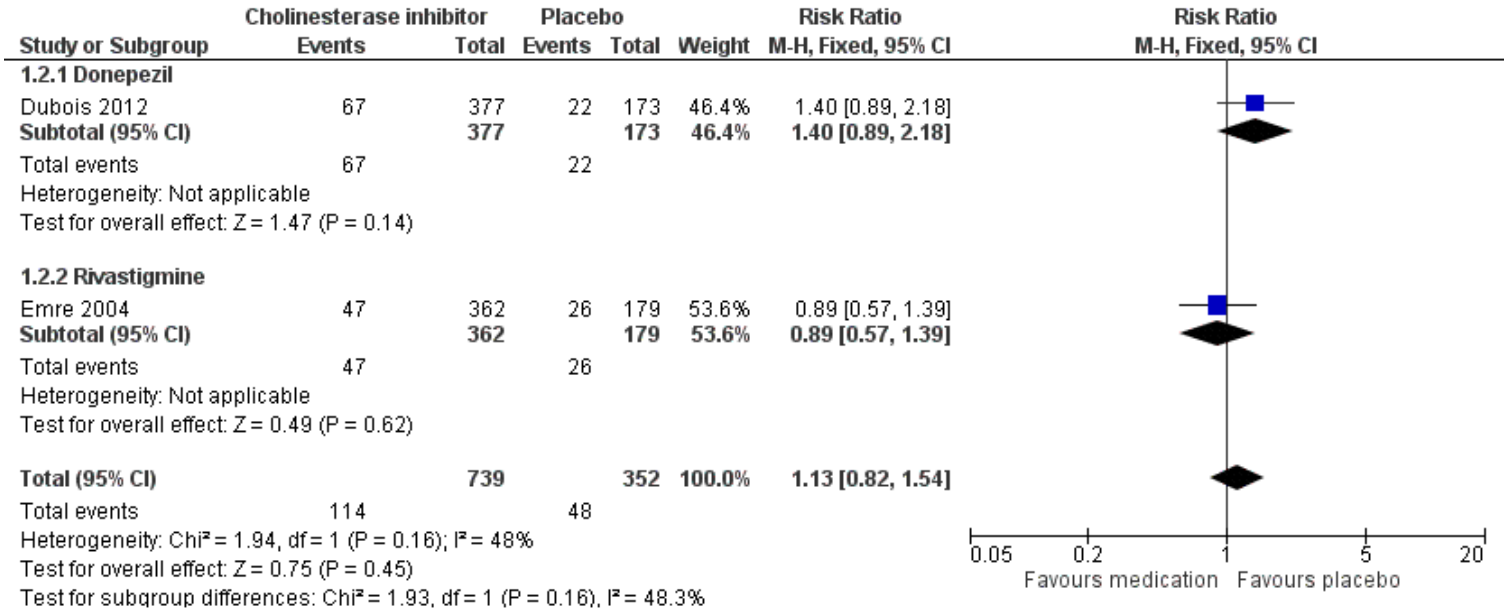
PDD – rivastigmine patches vs. rivastigmine capsules: adverse events

Quality assessment						No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Rivastigmine patches	Rivastigmine capsules	Relative (95% CI)	Absolute (95%CI)	
Any adverse events (probability of experiencing ≥1; follow-up 76 weeks; lower is better)										
1[1]	RCT	serious[2]	N/A	not serious	not serious	263/288 (91.3%)	274/294 (93.2%)	RR 0.98 (0.93 to 1.03)	19 fewer per 1000 (from 65 fewer to 28 more)	⊕⊕OO LOW
Serious adverse events (probability of experiencing ≥1; follow-up 76 weeks; lower is better)										
1[1]	RCT	serious[2]	N/A	not serious	serious[3]	83/288 (28.8%)	87/294 (29.6%)	RR 0.97 (0.76 to 1.25)	9 fewer per 1000 (from 71 fewer to 74 more)	⊕⊕OO LOW
Adverse events requiring treatment withdrawal (probability of experiencing; follow-up 76 weeks; lower is better)										
1[1]	RCT	serious[2]	N/A	not serious	serious[3]	71/288 (24.7%)	80/294 (27.2%)	RR 0.91 (0.69 to 1.19)	24 fewer per 1000 (from 84 fewer to 52 more)	⊕⊕OO LOW
Hallucinations (probability of experiencing ; follow-up 76 weeks)										
1[1]	RCT	serious[2]	N/A	not serious	serious[3]	25/288 (8.7%)	20/294 (6.8%)	RR 1.28 (0.73 to 2.25)	19 more per 1000 (from 18 fewer to 85 more)	⊕⊕OO LOW

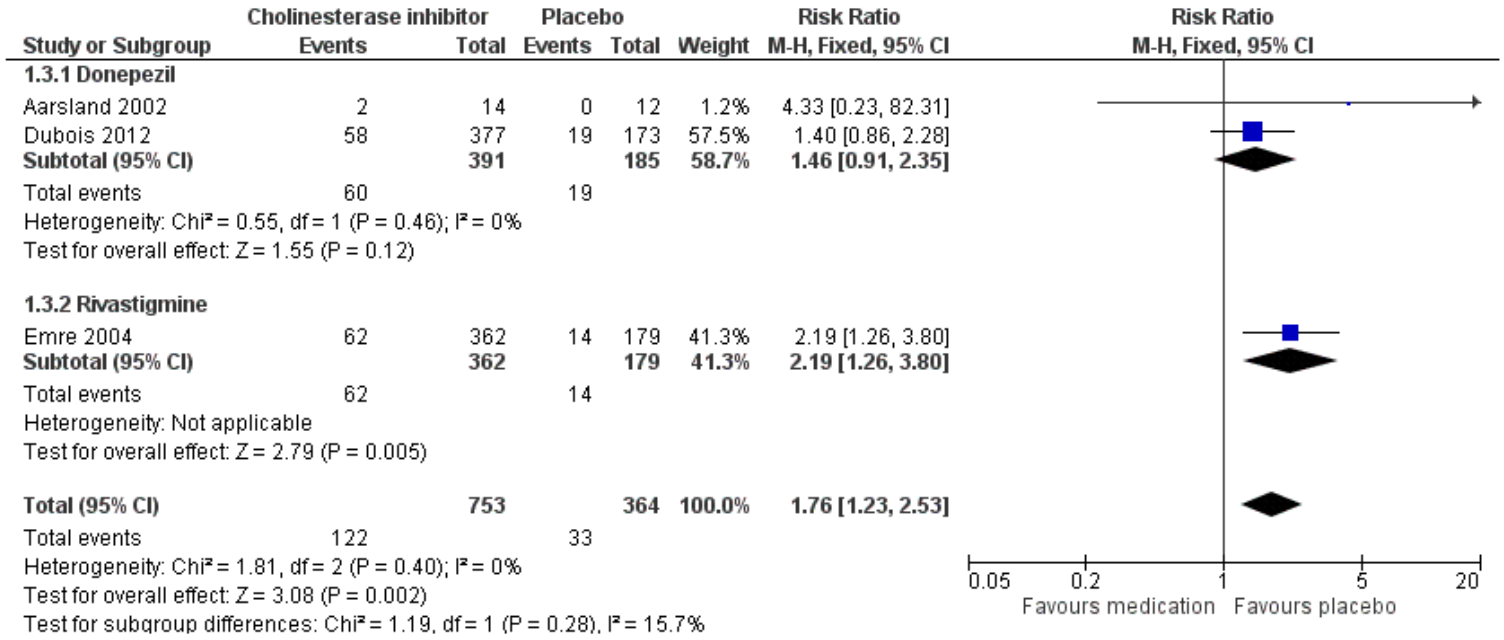
[1] Emre 2014
[2] Open-label study
[3] Data are consistent with appreciable harm, appreciable benefit or no difference



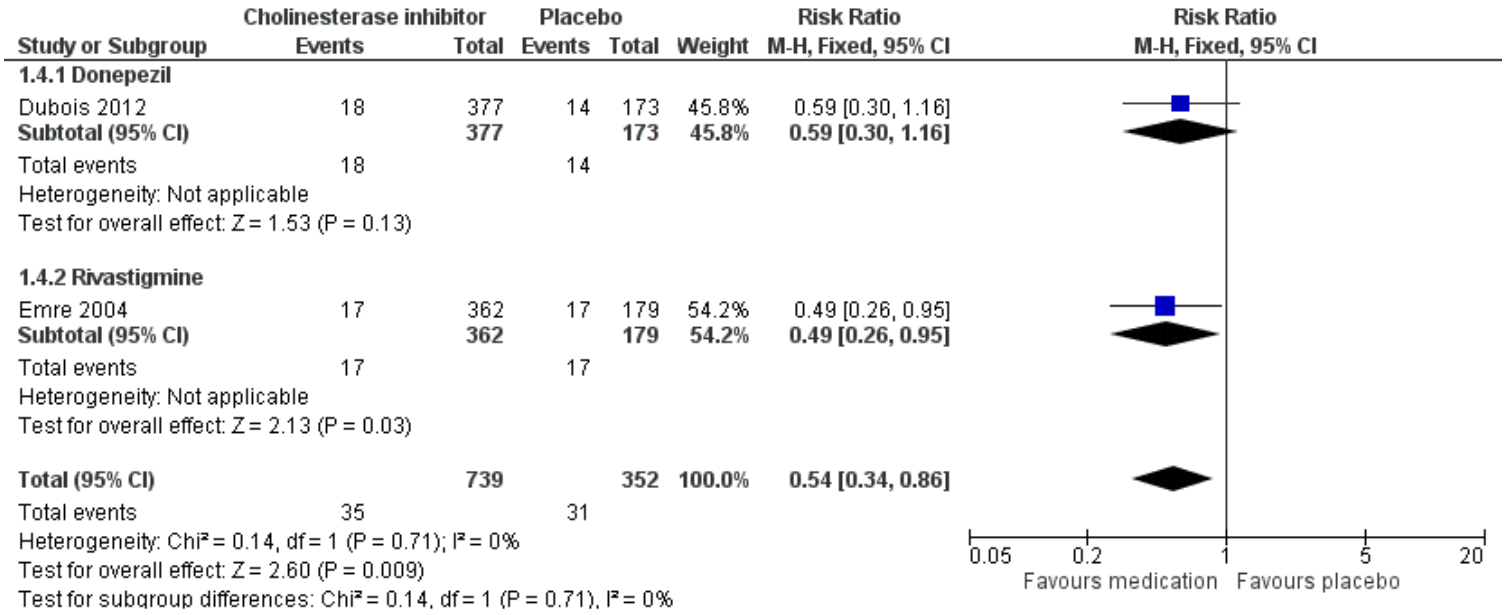
PDD – cholinesterase inhibitor vs placebo: any adverse events (proportion of participants experiencing ≥1) – forest plot



PDD – cholinesterase inhibitor vs placebo: serious adverse events (proportion of participants experiencing ≥1) – forest plot



PDD – cholinesterase inhibitor vs placebo: adverse events requiring treatment withdrawal (proportion of participants experiencing) forest plot



PDD – cholinesterase inhibitor vs placebo: hallucinations (proportion of participants experiencing) – forest plot

PDD – cholinesterase inhibitor vs. placebo: cognitive function

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Chl	Placebo	Mean difference (95% CI)	
MMSE – cholinesterase inhibitors (follow-up 10 to 24 weeks; range of scores: 0-30; higher is better); see Figure 5 for forest plot									
4[1–4]	RCT	not serious	not serious	not serious	not serious	752	367	1.36 higher (0.95 to 1.77 higher)	⊕⊕⊕⊕ HIGH
MMSE – donepezil (follow-up 10 to 24 weeks; range of scores: 0-30; higher is better)									
3[1,2,4]	RCT	not serious	not serious	not serious	not serious	417	201	1.58 higher (1.06 to 2.1 higher)	⊕⊕⊕⊕ HIGH
MMSE – rivastigmine (follow-up 24 weeks; range of scores: 0-30; higher is better)									
1[3]	RCT	not serious	N/A	not serious	not serious	335	166	1 higher (0.33 to 1.67 higher)	⊕⊕⊕⊕ HIGH
ADAS-cog – cholinesterase inhibitors (follow-up 10 to 24 weeks; range of scores: 0-70; lower is better); see Figure 6 for forest plot									
3[1,2,4]	RCT	not serious	not serious	not serious	not serious	689	346	2.28 lower (3.40 to 1.15 lower)	⊕⊕⊕⊕ HIGH
ADAS-cog – donepezil (follow-up 10 to 24 weeks; range of scores: 0-70; lower is better)									
2[2,4]	RCT	not serious	not serious	not serious	serious[5]	360	185	1.5 lower (3.28 lower to 0.27 higher)	⊕⊕⊕O MODERATE
ADAS-cog – rivastigmine (follow-up 24 weeks; range of scores: 0-70; lower is better)									
1[3]	RCT	not serious	N/A	not serious	not serious	329	161	2.8 lower (4.26 to 1.34 lower)	⊕⊕⊕⊕ HIGH
MDRS (total score) – cholinesterase inhibitors (follow-up 10 to 24 weeks; range of scores: 0-144; higher is better)[6] see Figure 7 for forest plot									
2[3,4]	RCT	not serious	not serious	not serious	very serious[5,7]	35	31	3.39 higher (4.06 lower to 10.84 higher)	⊕⊕OO LOW
MDRS (total score) – donepezil (follow-up 10 weeks; range of scores: 0-144; higher is better)									
1[4]	RCT	not serious	N/A	not serious	very serious[5,7]	19	19	0.2 lower (11.44 lower to 11.04 higher)	⊕⊕OO LOW
MDRS (total score) – rivastigmine (follow-up 24 weeks; range of scores: 0-144; higher is better)[6]									
1[3]	RCT	serious[7]	N/A	not serious	serious[5]	16	12	6.21 higher (3.75 lower to 16.17 higher)	⊕⊕OO LOW
Clock drawing test – rivastigmine (follow-up 24 weeks; range of scores: 0-10; higher is better)									
1[3]	RCT	serious[7]	N/A	not serious	serious[5]	49	30	1.1 higher (0.01 lower to 2.21 higher)	⊕⊕OO LOW
D-KEFS verbal fluency test (total score) – rivastigmine (follow-up 24 weeks; measured by number of correct responses; higher is better)									
1[3]	RCT	not serious	N/A	not serious	not serious	258	144	2.8 higher (1.47 to 4.13 higher)	⊕⊕⊕⊕ HIGH
D-KEFS verbal fluency test (letter fluency) – donepezil (follow-up 24 weeks; higher is better)									
1[2]	RCT	not serious	N/A	not serious	not serious	307	152	2.83 higher (0.95 to 4.71 higher)	⊕⊕⊕⊕ HIGH
D-KEFS verbal fluency test (category fluency) – donepezil (follow-up 24 weeks; higher is better)									

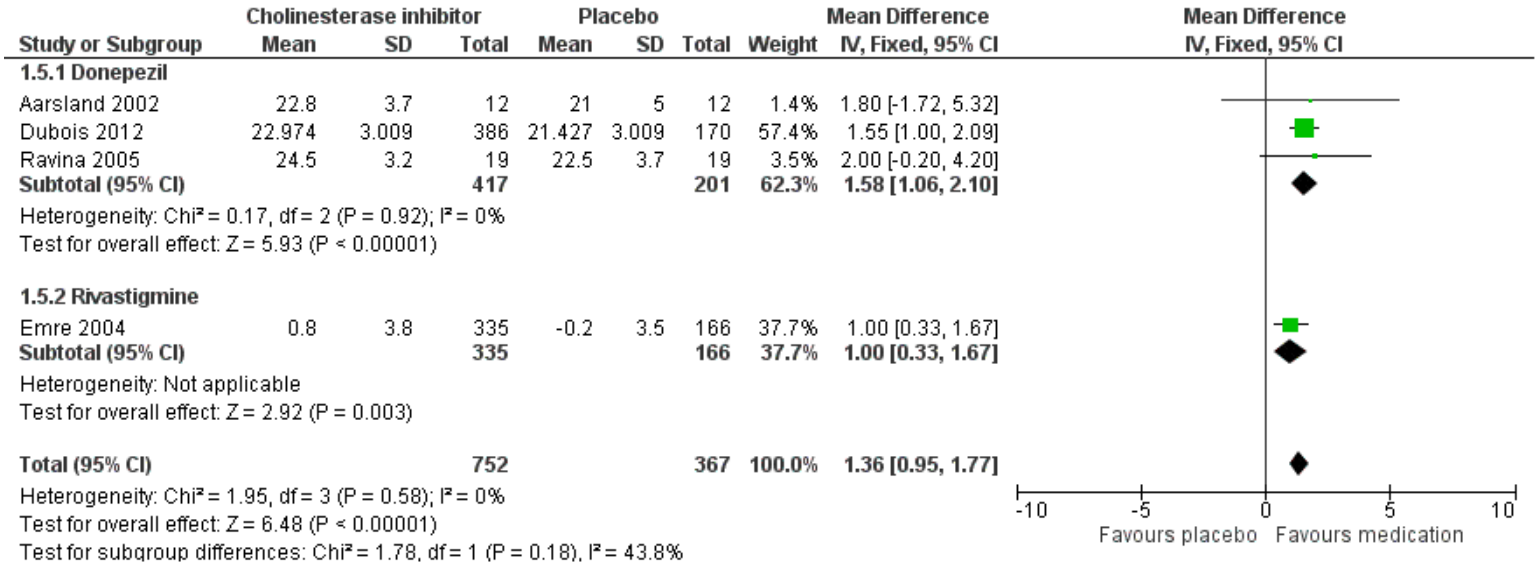
Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Chl	Placebo	Mean difference (95% CI)	
1[2]	RCT	not serious	N/A	not serious	not serious	307	152	3.93 higher (2.05 to 5.81 higher)	⊕⊕⊕⊕ HIGH
D-KEFS verbal fluency test (category switching) – donepezil (follow-up 24 weeks; higher is better)									
1[2]	RCT	not serious	N/A	not serious	serious[5]	307	152	1.09 higher (0.79 lower to 2.97 higher)	⊕⊕⊕O MODERATE
CDR – rivastigmine (follow-up 24 weeks; measured with: milliseconds; lower is better)									
1[3]	RCT	not serious	N/A	not serious	serious[5]	328	158	173.7 lower (471.23 lower to 123.83 higher)	⊕⊕⊕O MODERATE
BTA – donepezil (follow-up 24 weeks; range of scores: 0-20; higher is better)									
1[2]	RCT	serious[8]	N/A	not serious	not serious	221	111	0.88 higher (0.4 to 1.37 higher)	⊕⊕⊕O MODERATE

[1] Aarsland 2002
[2] Dubois 2012; data for 2 active treatment groups were combined (donepezil 5mg and 10mg). Mean and standard deviation calculated from data reported in paper
[3] Emre 2004
[4] Ravina 2005
[5] At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit or no difference
[6] Data from Emre 2004 reported in a secondary publication (Dujardin 2006)
[7] Small numbers of participants in the analysis
[8] Data available for only a small proportion of all participants for this outcome

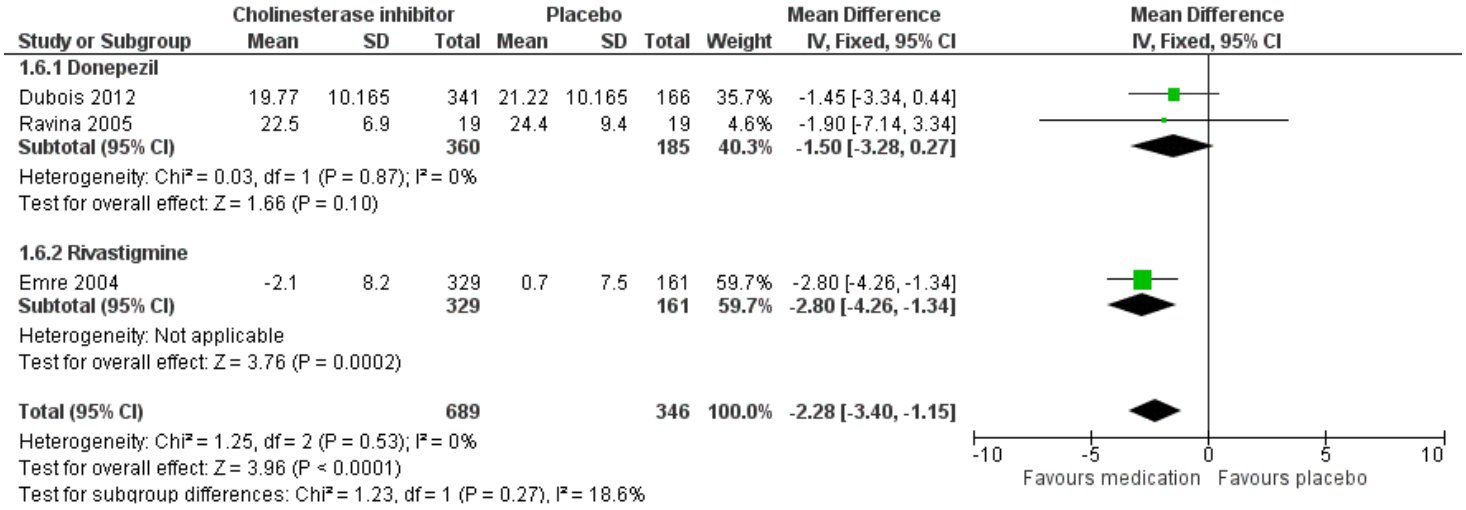
PDD – rivastigmine patches vs. rivastigmine capsules: cognitive outcomes

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Rivastigmine patches	Rivastigmine capsules	Mean difference (95% CI)	
MDRS (total score) (follow-up 24 weeks; range of scores 0-144; higher is better)									
1[1]	RCT	serious[2]	N/A	not serious	serious[3]	273	273	2.1 lower (4.27 lower to 0.07 higher)	⊕⊕OO LOW
MDRS (total score) (follow-up 76 weeks; range of scores 0-144; higher is better)									
1[1]	RCT	serious[2]	N/A	not serious	not serious	273	273	5.3 lower (8.17 to 2.43 lower)	⊕⊕⊕O MODERATE

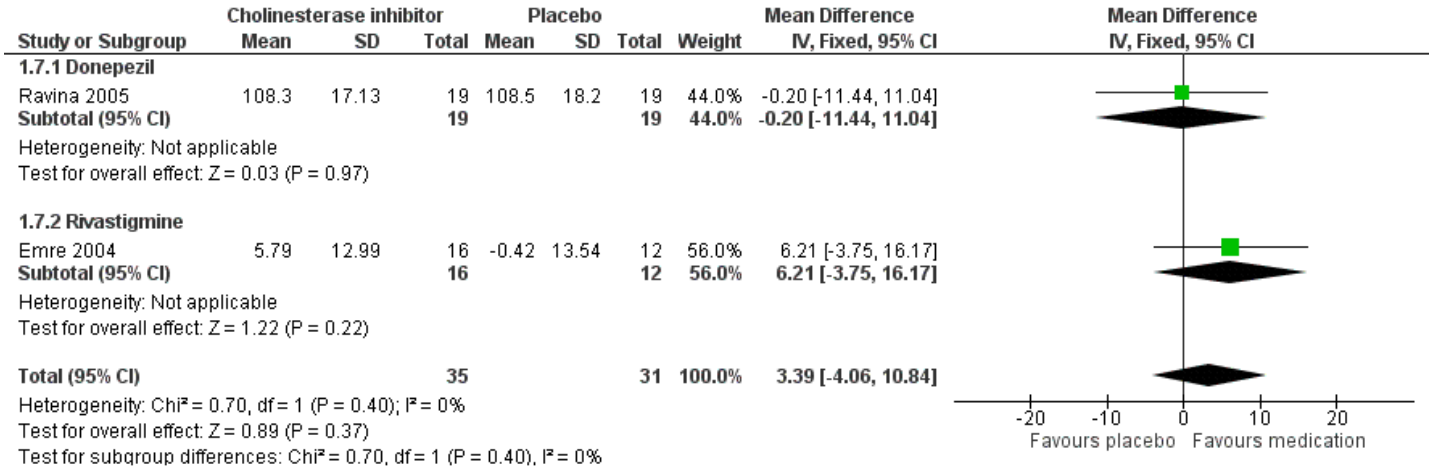
[1] *Emre 2014*
[2] *Open-label study*
[3] *At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit or no difference*



PDD – cholinesterase inhibitor vs placebo: MMSE – forest plot



PDD – cholinesterase inhibitor vs placebo: ADAS-cog – forest plot

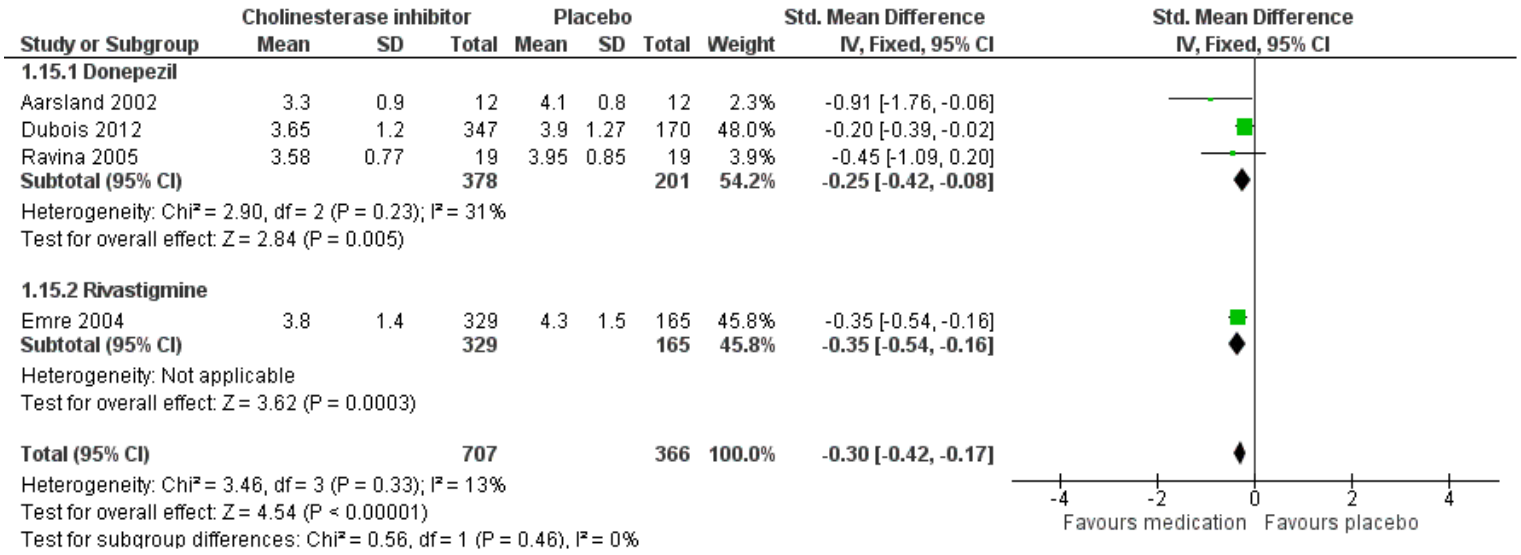


PDD – cholinesterase inhibitor vs placebo: MDRS (total score) – forest plot

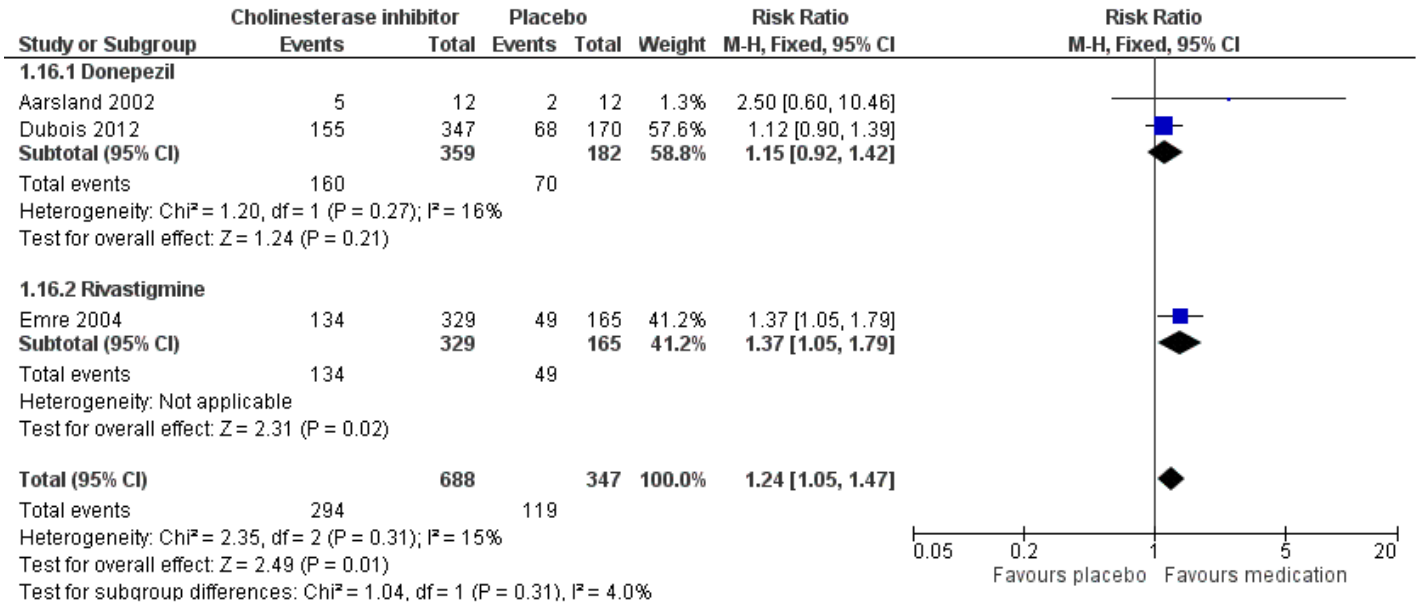
PDD – cholinesterase inhibitor vs. placebo: global assessment

Quality assessment						No of patients		Effect (95%CI)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Chl	Placebo		
Global function – cholinesterase inhibitors (follow-up 10 to 24 weeks; measured with: CIBIC+, ADCS-CGIC or CGIC; range of scores: 1-7; lower is better); see Figure 8 for forest plot									
4[1–4]	RCT	not serious	not serious	not serious	not serious	707	366	SMD 0.3 lower (0.42 to 0.17 lower)	⊕⊕⊕⊕ HIGH
Global response – cholinesterase inhibitors (at least minimal improvement; follow-up 10 to 24 weeks; measured with: CIBIC+ or ADCS-CGIC; higher is better); see Figure 9 for forest plot									
3[1–3]	RCT	not serious	not serious	not serious	not serious	294/688 (42.7%)	119/347 (34.3%)	RR 1.24 (1.05 to 1.47) 82 more per 1000 (from 17 more to 161 more)	⊕⊕⊕⊕ HIGH
Global response – donepezil (at least minimal improvement; follow-up 10 to 24 weeks; measured with: CIBIC+; higher is better)									
2[1,2]	RCT	not serious	not serious	not serious	serious[5]	160/359 (44.6%)	70/182 (38.5%)	RR 1.15 (0.92 to 1.42) 58 more per 1000 (from 31 fewer to 162 more)	⊕⊕⊕O MODERATE
Global response – rivastigmine (at least minimal improvement; follow-up 24 weeks; measured with: ADCS-CGIC; higher is better)									
1[3]	RCT	not serious	N/A	not serious	not serious	134/329 (40.7%)	49/165 (29.7%)	RR 1.37 (1.05 to 1.79) 110 more per 1000 (from 15 more to 235 more)	⊕⊕⊕⊕ HIGH
CIBIC+ – donepezil (follow-up 10 to 24 weeks; range of scores: 1-7; lower is better); see Figure 10 for forest plot									
2[1,2]	RCT	not serious	serious[6]	not serious	serious[5]	359	182	MD 0.43 lower (0.93 lower to 0.08 higher)	⊕⊕OO LOW
CGIC – donepezil (follow-up 10 weeks; range of scores: 1-7; lower is better)									
1[4]	RCT	not serious	N/A	not serious	very serious[5,7]	19	19	MD 0.37 lower (0.89 lower to 0.15 higher)	⊕⊕OO LOW
UPDRS (total score) – donepezil (follow-up 10 weeks; range of scores: 0-199; lower is better)									
1[4]	RCT	not serious	N/A	not serious	very serious[5,7,8]	21	20	MD 2.3 lower (15.77 lower to 11.17 higher)	⊕⊕OO LOW
ADCS-CGIC – rivastigmine (follow-up 24 weeks; range of scores: 1-7; lower is better)									
1[3]	RCT	not serious	N/A	not serious	not serious	329	165	MD 0.5 lower (0.77 to 0.23 lower)	⊕⊕⊕⊕ HIGH

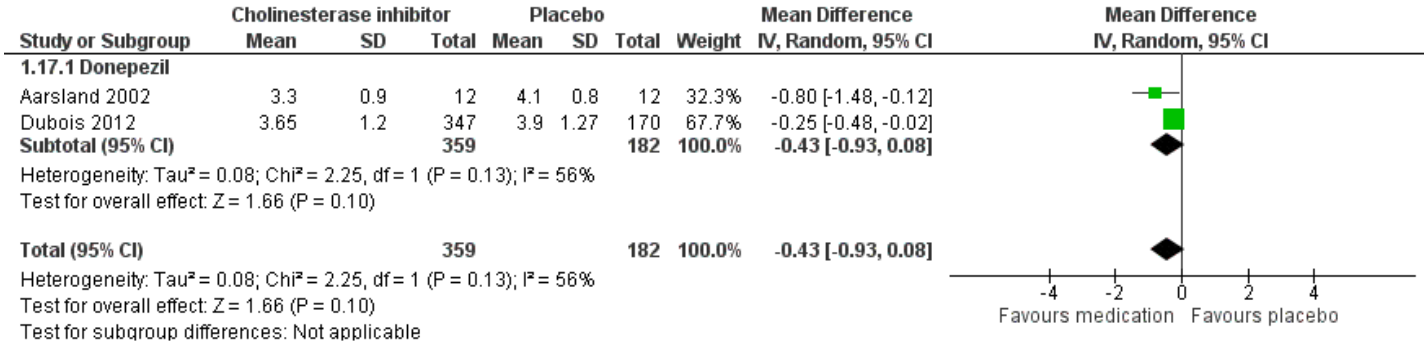
[1] Aarsland 2002
[2] Dubois 2012; data for 2 active treatment groups were combined (donepezil 5mg and 10mg). Mean and standard deviation calculated from data reported in paper
[3] Emre 2004
[4] Ravina 2005
[5] At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit or no difference
[6] i^2 > 40% between studies
[7] Data from a single very small study
[8] CI cross MID of 7.3 points (Schrag et al., 2006)



PDD – cholinesterase inhibitor vs placebo: global function (different measures)



PDD – cholinesterase inhibitor vs placebo: global response (at least minimal improvement) – forest plot



PDD – cholinesterase inhibitor (donepezil) vs placebo: CIBIC+ – forest plot

PDD – cholinesterase inhibitor vs. placebo: activities of daily living

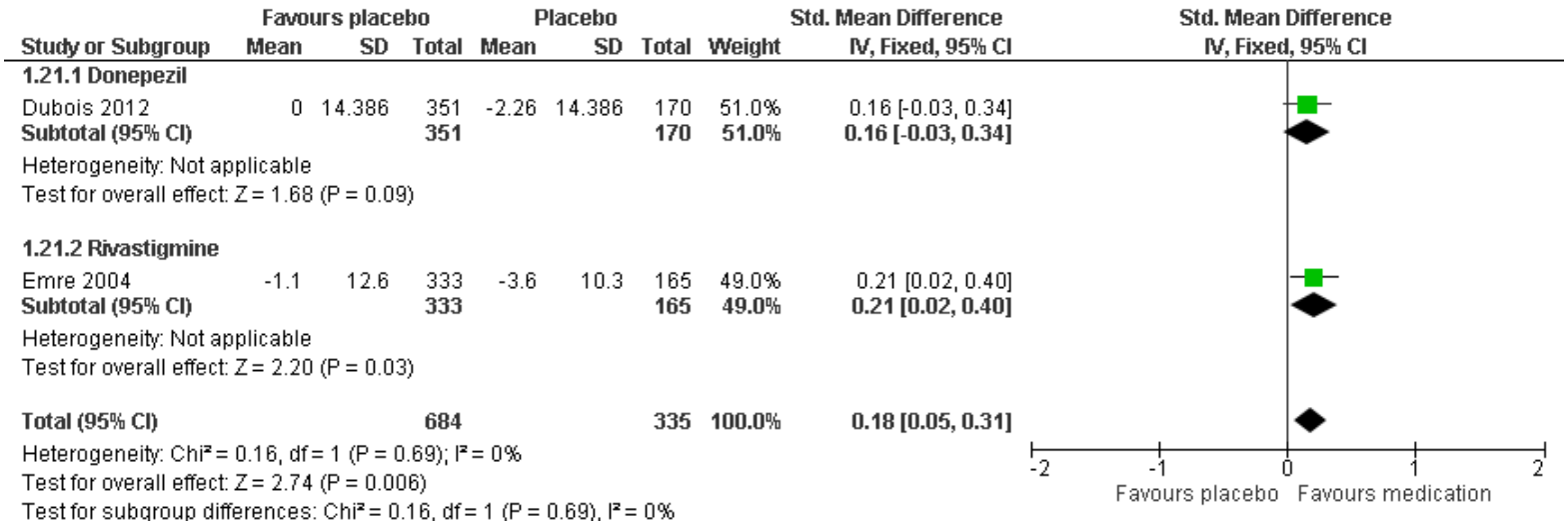
Quality assessment						No of patients		Effect (95% CI)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	ChI	Placebo		
ADL – cholinesterase inhibitors (follow-up 24 weeks; measured with: ADCS-ADL or DAD; higher is better); see Figure 11 for forest plot									
2[1,2]	RCT	not serious	not serious	not serious	not serious	684	335	SMD 0.18 higher (0.05 to 0.31 higher)	⊕⊕⊕⊕ HIGH
DAD – donepezil (follow-up 24 weeks; range of scores 0-100; higher is better)									
1[1]	RCT	not serious	N/A	not serious	serious[3]	351	170	MD 2.26 higher (0.38 lower to 4.89 higher)	⊕⊕⊕O MODERATE
ADCS-ADL – rivastigmine (follow-up 24 weeks; range of scores: 0-78; higher is better)									
1[2]	RCT	not serious	N/A	not serious	not serious	333	165	MD 2.5 higher (0.43 to 4.57 higher)	⊕⊕⊕⊕ HIGH

[1] Dubois 2012; data for 2 active treatment groups were combined (donepezil 5mg and 10mg). Mean and standard deviation calculated from data reported in paper
[2] Emre 2004
[3] At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit or no difference

PDD – rivastigmine patches vs. rivastigmine capsules: activities of daily living

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Rivastigmine patches	Rivastigmine capsules	Mean difference (95% CI)	
ADCS-ADL (follow-up 24 weeks; range of scores: 0-78; higher is better)									
1[1]	RCT	serious[2]	N/A	not serious	serious[3]	270	273	0.9 lower (2.67 lower to 0.87 higher)	⊕⊕OO LOW
ADCS-ADL (follow-up 76 weeks; range of scores: 0-78; higher is better)									
1[1]	RCT	serious[2]	N/A	not serious	not serious	270	273	3.4 lower (5.84 to 0.96 lower)	⊕⊕⊕O MODERATE

[1] Emre 2014
[2] Open-label study
[3] At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit or no difference



PDD – cholinesterase inhibitor vs placebo: ADL (different measures) – forest plot

PDD – cholinesterase inhibitor vs. placebo: other non-cognitive outcomes

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	ChI	Placebo	Mean difference (95% CI)	
NPI-10 item – cholinesterase inhibitors (follow-up 24 weeks; range of scores: 0-120; lower is better); see Figure 12 for forest plot									
2[1,2]	RCT	not serious[3]	not serious	not serious	not serious	688	336	1.67 lower (3.01 to 0.32 lower)	⊕⊕⊕⊕ HIGH
NPI-10 item – donepezil (follow-up 24 weeks; range of scores: 0-120; lower is better)									
1[1]	RCT	not serious[3]	N/A	not serious	serious[4]	354	170	1.34 lower (3.23 lower to 0.54 higher)	⊕⊕⊕O MODERATE
NPI-10 item – rivastigmine (follow-up 24 weeks; range of scores: 0-120; lower is better)									
1[2]	RCT	not serious	N/A	not serious	not serious	334	166	2.00 lower (3.91 to 0.09 lower)	⊕⊕⊕⊕ HIGH
UPDRS III – donepezil (follow-up 10 weeks; lower is better); see Figure 13 for forest plot									
2[5,6]	RCT	serious[7]	not serious	not serious	serious[4,8]	33	32	1.5 lower (7.87 lower to 4.87 higher)	⊕⊕OO LOW

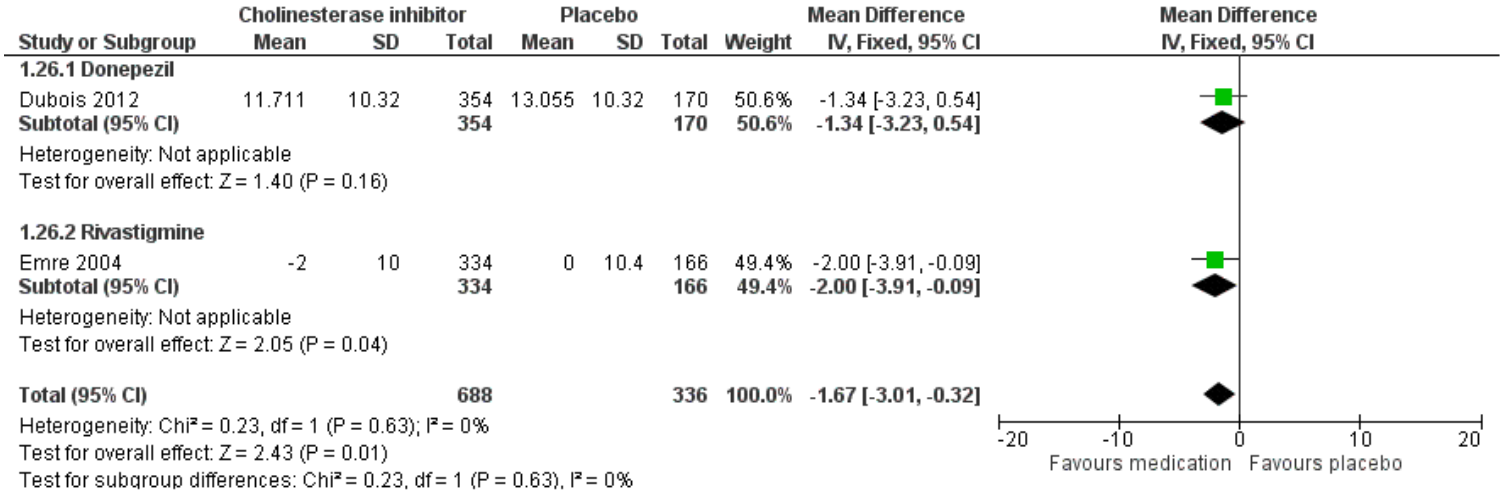
[1] Dubois 2012; data for 2 active treatment groups were combined (donepezil 5mg and 10mg). Mean and standard deviation calculated from data reported in paper
[2] Emre 2004
[3] Data for this outcome not reported in Aarsland 2002. This represents a very small proportion of the total participants in the analysis, therefore quality assessment not downgraded
[4] At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit or no difference
[5] Aarsland 2002
[6] Ravina 2005
[7]Data for this outcome not reported in 2 large RCTs (Dubois 2012 and Emre 2004). Papers stated no significant difference between groups
[8]CI cross MID between 3.25 (Horvath et al., 2015) and 5 points (Schrag et al., 2006)

PDD – rivastigmine patches vs. rivastigmine capsules: other non-cognitive outcomes

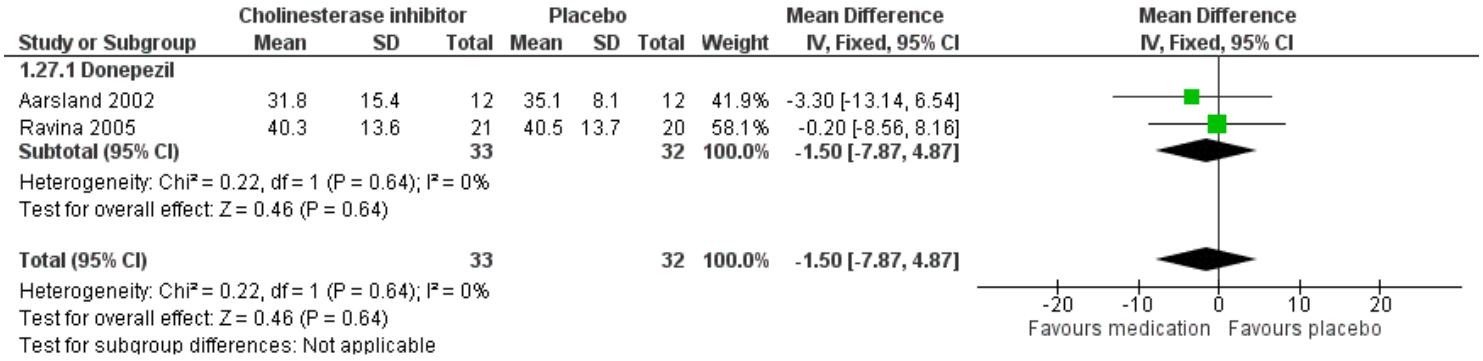
Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Rivastigmine patches	Rivastigmine capsules	Mean difference (95% CI)	
NPI-10 item (follow-up 24 weeks; range of scores: 0-120; lower is better)									
1[1]	RCT	serious[2]	N/A	not serious	serious[3]	273	273	1.6 higher (0.13 lower to 3.33 higher)	⊕⊕OO LOW
NPI-10 item (follow-up 76 weeks; range of scores: 0-120; lower is better)									
1[1]	RCT	serious[2]	N/A	not serious	not serious	273	273	2.3 lower (4.3 to 0.3 lower)	⊕⊕⊕O MODERATE
UPDRS III (follow-up 76 weeks; lower is better)									
1[1]	RCT	serious[2]	N/A	not serious	not serious[4]	175	183	0 higher (2.04 lower to 2.04 higher)	⊕⊕⊕O MODERATE

[1] Emre 2014
[2] Open-label study
[3] At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit or no difference

[4]CI do not cross MID between 3.25 (Horvath et al., 2015) and 5 points (Schrag et al., 2006)



PDD – cholinesterase inhibitor vs placebo: NPI-10 item – forest plot



PDD – cholinesterase inhibitor vs placebo: UPDRS III – forest plot

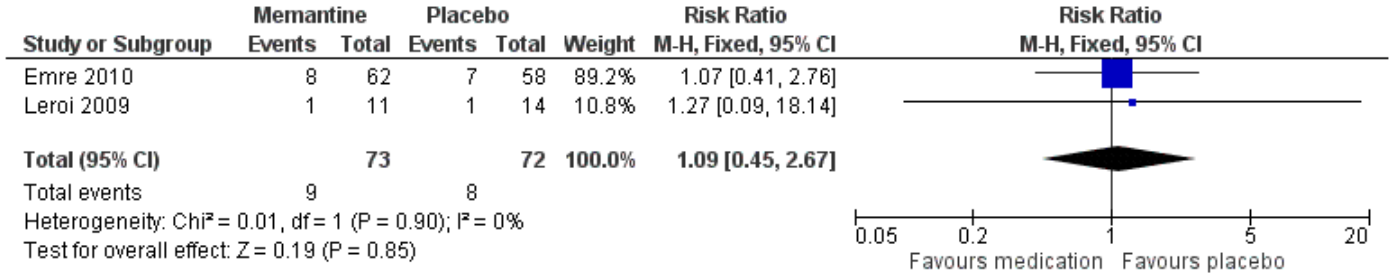
Parkinsons disease dementia – memantine

PDD – memantine vs. placebo: adverse events

Quality assessment						No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo	Relative (95% CI)	Absolute (95% CI)	
Any adverse events (probability of experiencing ≥1; follow-up 16 to 24 weeks, lower is better); see Figure 14 for forest plot										
2[1,2]	RCT	not serious	not serious	not serious	serious[3]	34/73 (46.6%)	35/72 (48.6%)	RR 0.97 (0.69 to 1.37)	15 fewer per 1000 (from 151 fewer to 180 more)	⊕⊕⊕O MODERATE
Serious adverse events (probability of experiencing ≥1; follow-up 16 to 24 weeks, lower is better); see Figure 15 for forest plot										
2[1,2]	RCT	not serious	not serious	not serious	very serious[3,4]	9/73 (12.3%)	8/72 (11.1%)	RR 1.09 (0.45 to 2.67)	10 more per 1000 (from 61 fewer to 186 more)	⊕⊕OO LOW
Adverse events requiring treatment withdrawal (probability of experiencing; follow-up 24 weeks, lower is better)										
1[1]	RCT	not serious	N/A	not serious	very serious[3,4]	6/62 (9.7%)	5/58 (8.6%)	RR 1.12 (0.36 to 3.48)	10 more per 1000 (from 55 fewer to 214 more)	⊕⊕OO LOW

[1] Emre 2010; data reported for PDD population only; study also included people with DLB
[2] Leroi 2009; not clear if adverse event data reported at end of active treatment (16 weeks) or end of drug withdrawal phase (22 weeks)
[3] At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit or no difference
[4] Very small numbers of events

PDD – memantine vs placebo: any adverse events (proportion of participants experiencing ≥1) – forest plot



PDD – memantine vs placebo: serious adverse events (proportion of participants experiencing ≥1) – forest plot

PDD – memantine vs. placebo: cognitive function

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo	Mean difference (95% CI)	
MMSE (follow-up 16 weeks; range of scores: 0-30; higher is better)									
1[1]	RCT	not serious	N/A	not serious	very serious[2,3]	10	14	1 lower (6.01 lower to 4.01 higher)	⊕⊕OO LOW
Clock drawing test (follow-up 24 weeks; range of scores: 0-10; higher is better)									
1[4]	RCT	not serious	N/A	not serious	serious[2]	57	56	3.1 higher (6.94 lower to 13.14 higher)	⊕⊕⊕O MODERATE

[1] Leroi 2009; data reported for end of drug treatment phase (16 weeks)
[2] At a 95% confidence level, data are consistent with appreciable benefit, appreciable harm or no difference
[3] Very small numbers of participants in the study
[4] Emre 2010; data reported for PDD population only; study also included people with DLB

PDD – memantine vs. placebo: global assessment

Quality assessment						No of patients		Effect (95% CI)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo		
ADCS-CGIC (follow-up 24 weeks; range of scores: 1-7; lower is better)									
1[1]	RCT	not serious	N/A	not serious	serious[2]	60	56	MD 0.2 lower (0.69 lower to 0.29 higher)	⊕⊕⊕O MODERATE
CIBIC+ (at least minimal improvement; follow-up 16 weeks; higher is better)									
1[3]	RCT	not serious	N/A	not serious	very serious[2,4]	6/10 (60%)	6/14 (42.9%)	RR 1.4 (0.64 to 3.08) 171 more per 1000 (from 154 fewer to 891 more)	⊕⊕OO LOW

[1] Emre 2010; data reported for PDD population only; study also included people with DLB
[2] At a 95% confidence level, data are consistent with appreciable benefit, appreciable harm or no difference
[3] Leroi 2009; data reported for end of drug treatment phase (16 weeks)
[4] Data from a single very small study

PDD – memantine vs. placebo: activities of daily living

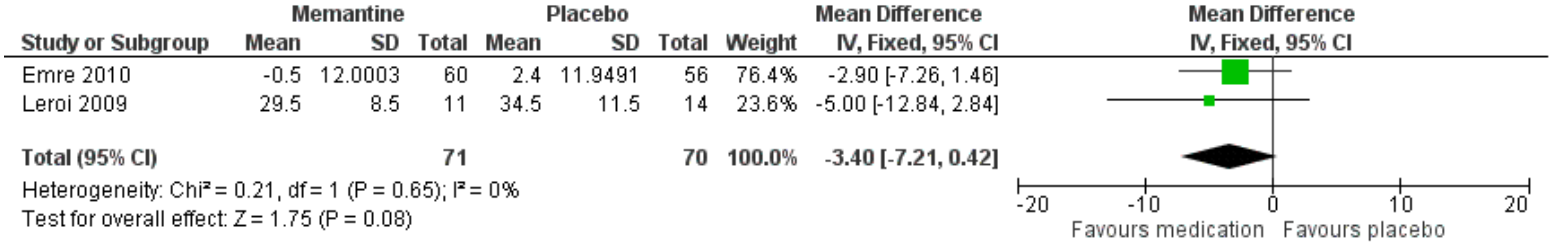
Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo	Mean difference (95% CI)	
ADCS-ADL (follow-up 24 weeks; measured with: 23-item score; higher is better)									
1[1]	RCT	not serious	N/A	not serious	serious[2]	60	56	0.8 higher (3.22 lower to 4.82 higher)	⊕⊕⊕O MODERATE

[1] Emre 2010; data reported for PDD population only; study also included people with DLB
[2] At a 95% confidence level, data are consistent with appreciable benefit, appreciable harm or no difference

PDD – memantine vs. placebo: carer-reported outcomes

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo	Mean difference (95% CI)	
ZBI (follow-up 16 to 24 weeks; lower is better)[1]; see Figure 16 for forest plot									
2[2,3]	RCT	not serious	not serious	not serious	serious[4]	71	70	3.4 lower (7.21 lower to 0.42 higher)	⊕⊕⊕O MODERATE

[1] Data from Leroi 2009 reported in a secondary publication (Leroi 2014)
[2] Leroi 2009; data reported for end of drug treatment phase (16 weeks)
[3] Emre 2010; data reported for PDD population only; study also included people with DLB
[4] At a 95% confidence level, data are consistent with appreciable benefit, appreciable harm or no difference

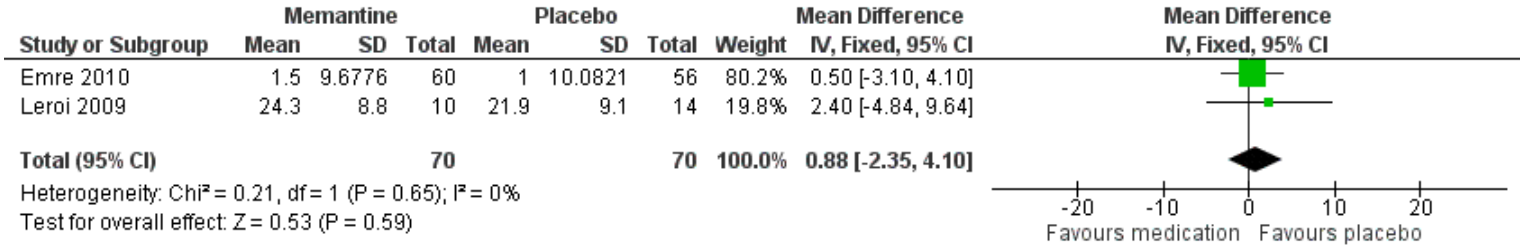


PDD – memantine vs placebo: ZBI – forest plot

PDD – memantine vs. placebo: other non-cognitive outcomes

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo	Mean difference (95% CI)	
NPI 12-item (follow-up 24 weeks; range of scores: 0-144; lower is better)									
1[1]	RCT	not serious	N/A	not serious	serious[3]	60	56	MD 1.50 lower (6.35 lower to 3.35 higher)	⊕⊕⊕O MODERATE
NPI 10-item (follow-up 16 weeks; range of scores: 0-120; lower is better)									
1[2]	RCT	not serious	N/A	not serious	very serious[3,4]	10	14	MD 2.00 lower (11.64 lower to 7.64 higher)	⊕⊕OO LOW
UPDRS III (follow-up 16 to 24 weeks; lower is better); see Figure 17 for forest plot									
2[1,2]	RCT	not serious	not serious	not serious	serious[3,5]	70	70	MD 0.88 higher (2.35 lower to 4.1 higher)	⊕⊕⊕O MODERATE

[1] Emre 2010; data reported for PDD population only; study also included people with DLB
[2] Leroi 2009; data reported for end of drug treatment phase (16 weeks)
[3] At a 95% confidence level, data are consistent with appreciable benefit, appreciable harm or no difference
[4] Data from a single very small study
[5]CI cross MID between 3.25 (Horvath et al 2015) and 5 points (Schrag et al., 2006)



PDD – memantine vs placebo: UPDRS III – forest plot

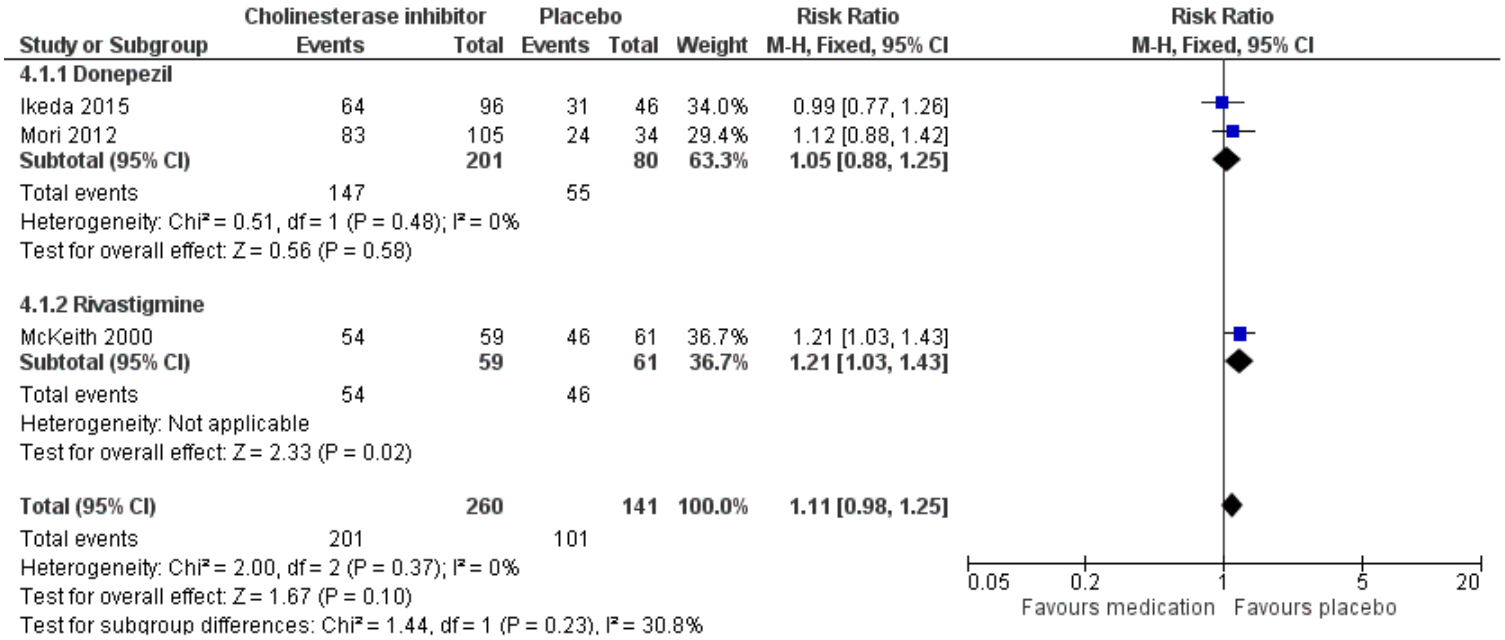
Dementia with Lewy bodies – cholinesterase inhibitors

DLB – cholinesterase inhibitor vs. placebo: adverse events

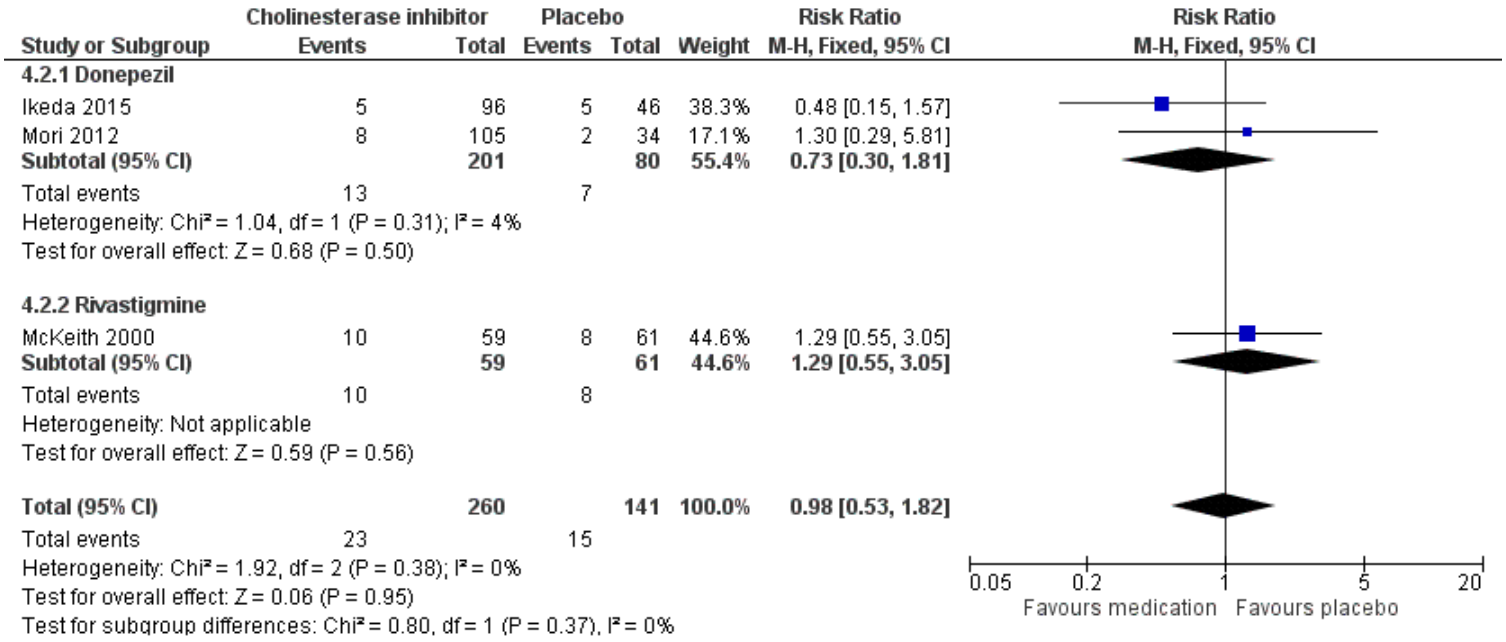
Quality assessment						No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Chl	Placebo	Relative (95% CI)	Absolute (95% CI)	
Any adverse events – cholinesterase inhibitors (probability of experiencing ≥1; follow-up 12 to 20 weeks);); see Figure 18 for forest plot										

Quality assessment						No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Chl	Placebo	Relative (95% CI)	Absolute (95% CI)	
3[1–3]	RCT	not serious	not serious	not serious	serious[4]	201/260 (77.3%)	101/141 (71.6%)	RR 1.11 (0.98 to 1.25)	79 more per 1000 (from 14 fewer to 179 more)	⊕⊕⊕O MODERATE
Any adverse events – donepezil (probability of experiencing ≥1; follow-up 12 weeks)										
2[1,2]	RCT	not serious	not serious	not serious	serious[4]	147/201 (73.1%)	55/80 (68.8%)	RR 1.05 (0.88 to 1.25)	34 more per 1000 (from 83 fewer to 172 more)	⊕⊕⊕O MODERATE
Any adverse events – rivastigmine (probability of experiencing ≥1; follow-up 20 weeks)										
1[3]	RCT	not serious	N/A	not serious	not serious	54/59 (91.5%)	46/61 (75.4%)	RR 1.21 (1.03 to 1.43)	158 more per 1000 (from 23 more to 324 more)	⊕⊕⊕⊕ HIGH
Serious adverse events – cholinesterase inhibitors (probability of experiencing ≥1; follow-up 12 to 20 weeks);); see Figure 19 for forest plot										
3[1–3]	RCT	not serious	not serious	not serious	serious[4]	23/260 (8.8%)	15/141 (10.9%)	RR 0.98 (0.53 to 1.82)	2 fewer per 1000 (from 51 fewer to 89 more)	⊕⊕⊕O MODERATE
Serious adverse events – donepezil (probability of experiencing ≥1; follow-up 12 weeks)										
2[1,2]	RCT	not serious	not serious	not serious	serious[4]	13/201 (6.5%)	7/80 (8.8%)	RR 0.73 (0.3 to 1.81)	24 fewer per 1000 (from 61 fewer to 71 more)	⊕⊕⊕O MODERATE
Serious adverse events – rivastigmine (probability of experiencing ≥1; follow-up 20 weeks)										
1[3]	RCT	not serious	N/A	not serious	serious[4]	10/59 (16.9%)	8/61 (13.1%)	RR 1.29 (0.55 to 3.05)	38 more per 1000 (from 59 fewer to 269 more)	⊕⊕⊕O MODERATE
Adverse events requiring treatment withdrawal – cholinesterase inhibitors (probability of experiencing; follow-up 12 to 20 weeks)); see Figure 20 for forest plot										
3[1–3]	RCT	not serious	not serious	not serious	serious[4]	25/260 (9.6%)	16/141 (11.3%)	RR 0.9 (0.49 to 1.63)	11 fewer per 1000 (from 58 fewer to 71 more)	⊕⊕⊕O MODERATE
Adverse events requiring treatment withdrawal – donepezil (probability of experiencing; follow-up 12 weeks)										
2[1,2]	RCT	not serious	not serious	not serious	serious[4]	18/201 (9%)	9/80 (11.3%)	RR 0.82 (0.39 to 1.74)	20 fewer per 1000 (from 69 fewer to 83 more)	⊕⊕⊕O MODERATE
Adverse events requiring treatment withdrawal – rivastigmine (probability of experiencing; follow-up 20 weeks)										
1[3]	RCT	not serious	N/A	not serious	serious[4]	7/59 (11.9%)	7/61 (11.5%)	RR 1.03 (0.39 to 2.77)	3 more per 1000 (from 70 fewer to 203 more)	⊕⊕⊕O MODERATE

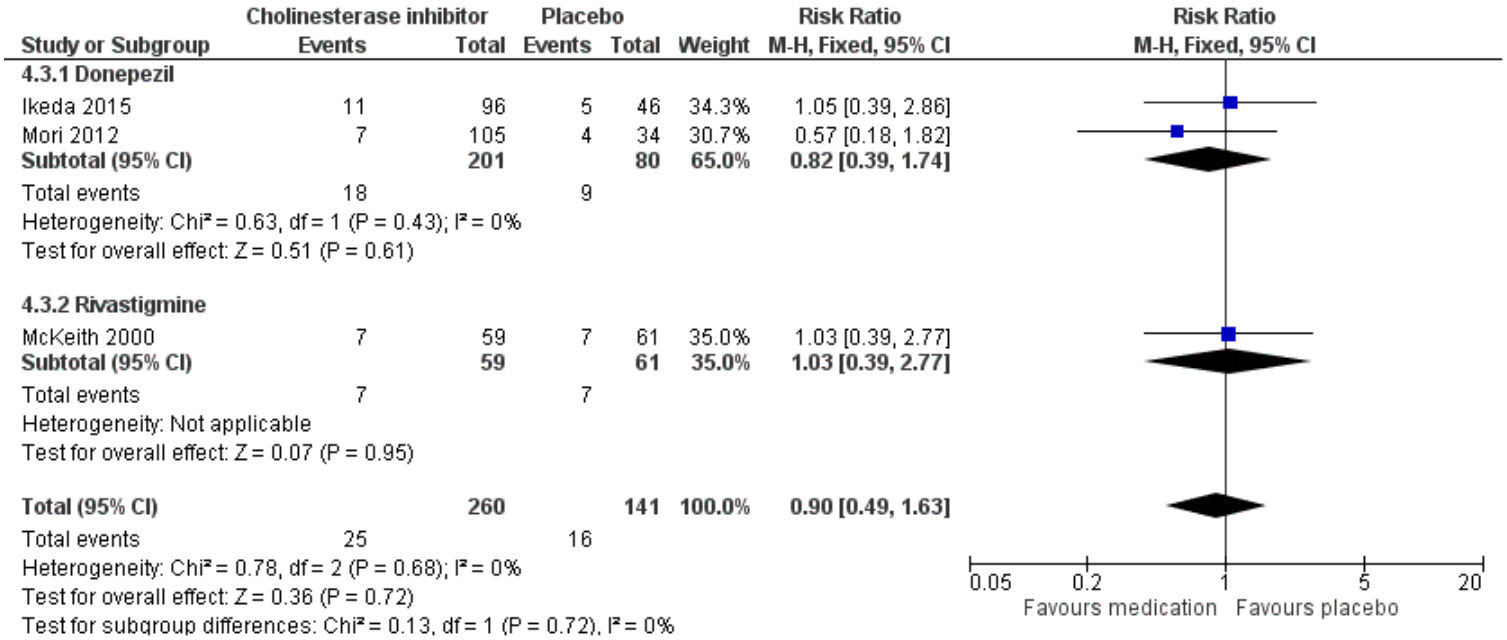
[1] Ikeda 2015; data for 2 active treatment groups were combined (donepezil 5mg and 10mg)
[2] Mori 2012; data for 3 active treatment groups were combined (donepezil 3mg, 5mg and 10mg)
[3] McKeith 2000
[4] At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit or no difference



DLB – cholinesterase inhibitor vs placebo: any adverse events (proportion of participants experiencing ≥1) – forest plot



DLB – cholinesterase inhibitor vs placebo: serious adverse events (proportion of participants experiencing ≥1) – forest plot

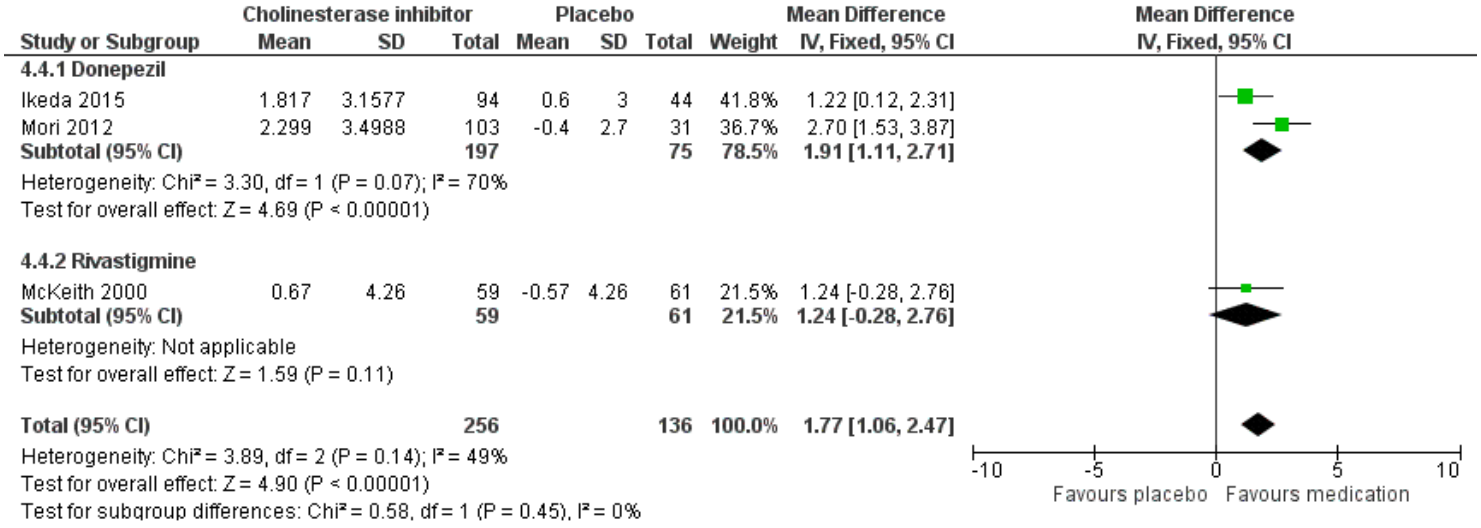


DLB – cholinesterase inhibitor vs placebo: adverse events requiring treatment withdrawal (proportion of participants experiencing) – forest plot

DLB – cholinesterase inhibitor vs. placebo: cognitive function

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Chl	Placebo	Mean difference (95% CI)	
MMSE – cholinesterase inhibitors (follow-up 12 to 20 weeks; range of scores: 0-30; higher is better); see Figure 21 for forest plot									
3[1–3]	RCT	not serious	serious[4]	not serious	not serious	256	136	1.77 higher (1.06 to 2.47 higher)	⊕⊕⊕O MODERATE
MMSE – donepezil (follow-up 12 weeks; range of scores: 0-30; higher is better)									
2[1,3]	RCT	not serious	serious[4]	not serious	not serious	197	75	1.91 higher (1.11 to 2.71 higher)	⊕⊕⊕O MODERATE
MMSE – rivastigmine (follow-up 20 weeks; range of scores: 0-30; higher is better)									
1[2]	RCT	not serious	N/A	not serious	serious[5]	59	61	1.24 higher (0.28 lower to 2.76 higher)	⊕⊕⊕O MODERATE

[1] Ikeda 2015; data for 2 active treatment groups were combined (donepezil 5mg and 10mg)
[2] McKeith 2000; data for this outcome taken from a Cochrane review; data not reported in published paper
[3] Mori 2012; data for 3 active treatment groups were combined (donepezil 3mg, 5mg and 10mg)
[4] i^2 >40% between studies
[5] At a 95% confidence level, data are consistent with appreciable benefit, appreciable harm or no difference



DLB – cholinesterase inhibitor vs placebo: MMSE – forest plot

DLB – cholinesterase inhibitor vs. placebo: global assessment

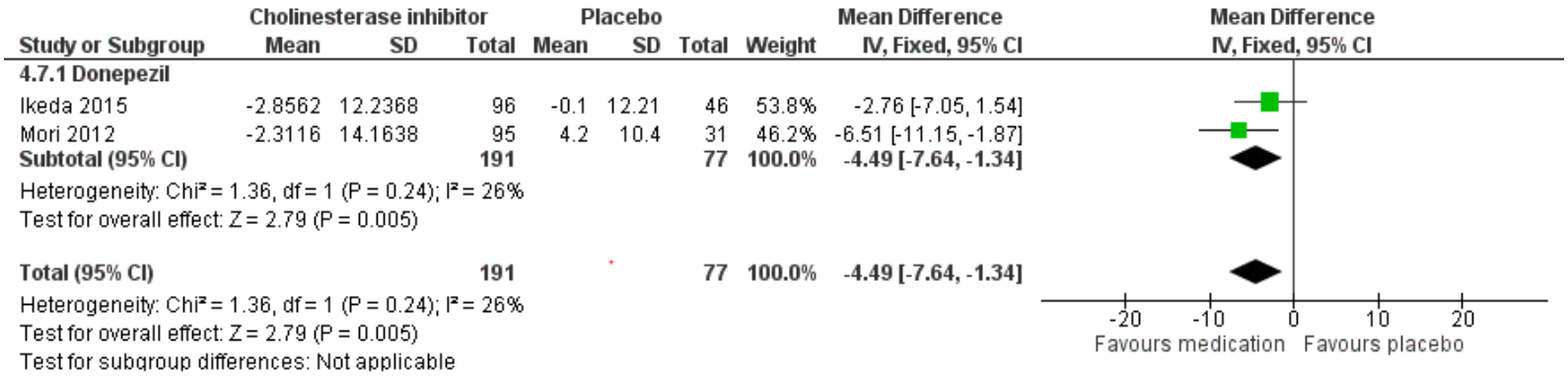
Quality assessment						No of patients		Effect (95% CI)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	ChI	Placebo		
CIBIC+ – donepezil (follow-up 12 weeks; range of scores: 1-7; lower is better)[1]									
1[2]	RCT	not serious	N/A	not serious	not serious	91	30	MD 1.17 lower (1.66 to 0.68 lower)	⊕⊕⊕⊕ HIGH
CIBIC+ – donepezil (at least minimal improvement; follow-up 12 weeks; higher is better)									
1[2]	RCT	not serious	N/A	not serious	not serious	62/91 (68.1%)	10/30 (33.3%)	RR 2.04 (1.21 to 3.46) 347 more per 1000 (from 70 more to 820 more)	⊕⊕⊕⊕ HIGH

[1] Mean and SD calculated from data presented in paper
[2] Mori 2012; data for 3 active treatment groups were combined (donepezil 3mg, 5mg and 10mg)

DLB – cholinesterase inhibitor vs. placebo: carer-reported outcomes

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	ChI	Placebo	Mean difference (95% CI)	
ZBI - donepezil (follow-up 12 weeks; lower is better); see Figure 22 for forest plot									
2[1,2]	RCT	not serious	not serious	not serious	not serious	191	77	4.49 lower (7.64 to 1.34 lower)	⊕⊕⊕⊕ HIGH

[1] Ikeda 2015; data for 2 active treatment groups were combined (donepezil 5mg and 10mg)
[2] Mori 2012; data for 3 active treatment groups were combined (donepezil 3mg, 5mg and 10mg)

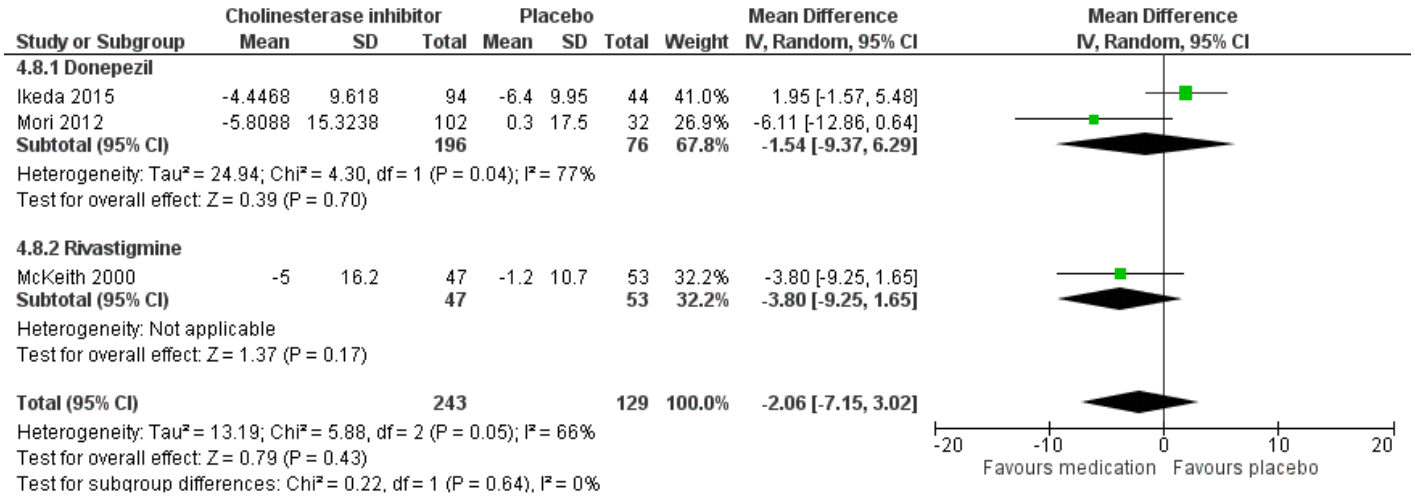


DLB – cholinesterase inhibitor (donepezil) vs placebo: ZBI – forest plot

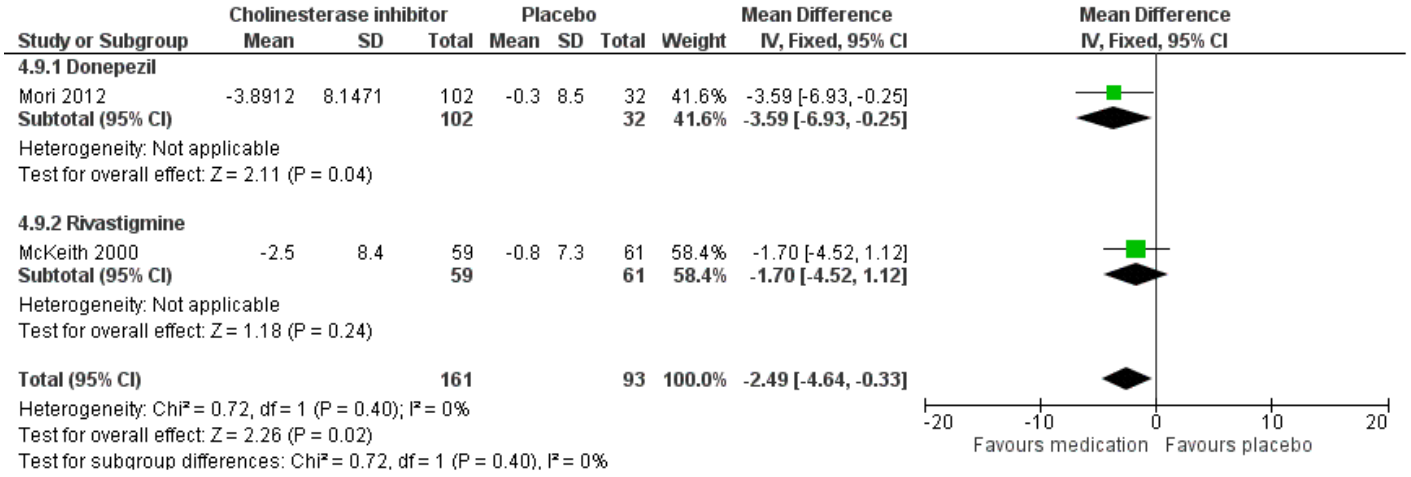
DLB – cholinesterase inhibitor vs. placebo: Other non-cognitive outcomes

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Chl	placebo	Mean difference (95% CI)	
NPI-10 item – cholinesterase inhibitors (follow-up 12 to 20 weeks; range of scores: 0-120; lower is better)[1]; see Figure 23 for forest plot									
3[2–4]	RCT	not serious	serious[5]	not serious	serious[6]	243	129	2.06 lower (7.15 lower to 3.02 higher)	⊕⊕OO LOW
NPI-10 item – donepezil (follow-up 12 weeks; range of scores: 0-120; lower is better)[1]									
2[2,4]	RCT	not serious	serious[5]	not serious	serious[6]	196	76	1.54 lower (9.37 lower to 6.29 higher)	⊕⊕OO LOW
NPI-10 item – rivastigmine (follow-up 20 weeks; range of scores: 0-120; lower is better)									
1[3]	RCT	not serious	N/A	not serious	serious[6]	47	53	3.8 lower (9.25 lower to 1.65 higher)	⊕⊕⊕O MODERATE
NPI-4 item – cholinesterase inhibitors (follow-up 12 to 20 weeks; range of scores: 0-48; lower is better)[7]; see Figure 24 for forest plot									
2[3,4]	RCT	not serious	not serious	not serious	not serious	161	93	2.49 lower (4.64 to 0.33 lower)	⊕⊕⊕⊕ HIGH
NPI-4 item – donepezil (follow-up 12 weeks; range of scores: 0-48; lower is better)[7]									
1[4]	RCT	not serious	N/A	not serious	not serious	102	32	3.59 lower (6.93 to 0.25 lower)	⊕⊕⊕⊕ HIGH
NPI-4 item – rivastigmine (follow-up 20 weeks; range of scores: 0-48; lower is better)[7]									
1[3]	RCT	not serious	N/A	not serious	serious[6]	59	61	1.7 lower (4.52 lower to 1.12 higher)	⊕⊕⊕O MODERATE
NPI-2 item – donepezil (follow-up 12 weeks; range of scores: 0-24; lower is better)[8]; see Figure 25 for forest plot									
2[2,4]	RCT	not serious	serious[5]	not serious	serious[6]	196	76	2.3 lower (6.32 lower to 1.72 higher)	⊕⊕OO LOW
UPDRS III – cholinesterase inhibitors (follow-up 12 weeks; lower is better)[1]; see Figure 26 for forest plot									
2[2,4]	RCT	serious[9]	not serious	not serious	not serious[10]	195	77	0.67 lower (2.08 lower to 0.73 higher)	⊕⊕⊕O MODERATE
UPDRS III – donepezil (follow-up 12 weeks; lower is better)[1]									
2[2,4]	RCT	not serious	not serious	not serious	not serious[10]	195	77	0.67 lower (2.08 lower to 0.73 higher)	⊕⊕⊕⊕ HIGH

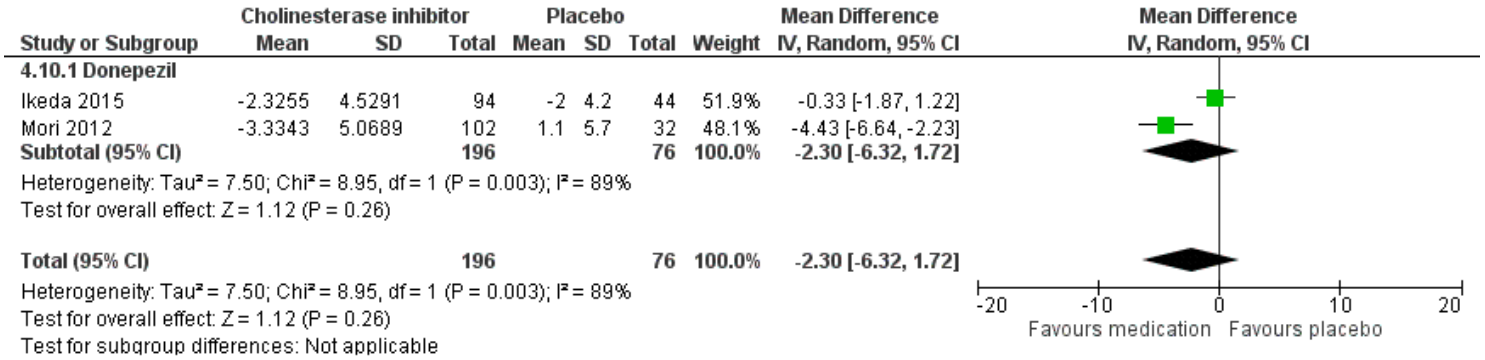
[1] SD not reported for this outcome in Ikeda 2015; calculated from SE reported in paper
[2] Ikeda 2015; data for 2 active treatment groups were combined (donepezil 5mg and 10mg)
[3] McKeith 2000
[4] Mori 2012; data for 3 active treatment groups were combined (donepezil 3mg, 5mg and 10mg)
[5] i^2 >40% between studies
[6] At a 95% confidence level, data are consistent with appreciable benefit, appreciable harm or no difference
[7] NPI 4-item consists of 4 NPI domains – hallucinations, delusions, dysphoria and apathy
[8] NPI 2-item consists of 2 NPI domains – hallucinations and cognitive fluctuation
[9] Data for outcome not presented in McKeith 2000. Study reported no significant difference between groups
[10] CI do not cross MID between 3.25 (Horvath et al., 2015) and 5 points (Schrag et al., 2006)



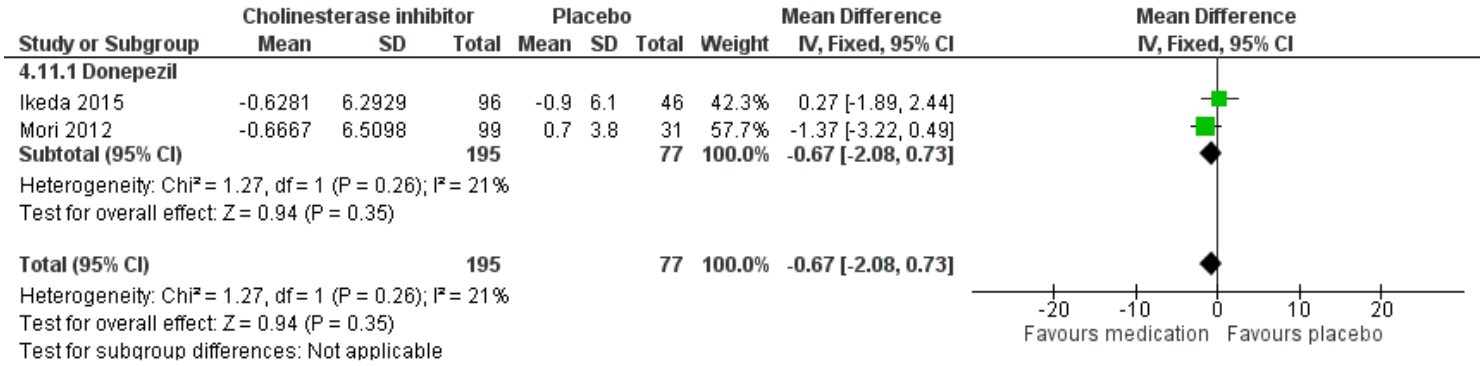
DLB – cholinesterase inhibitor vs placebo: NPI-10 item – forest plot



DLB – cholinesterase inhibitor vs placebo: NPI-4 item – forest plot



DLB – cholinesterase inhibitor (donepezil) vs placebo: NPI-2 item – forest plot



DLB – cholinesterase inhibitor (donepezil) vs placebo: UPDRS III – forest plot

Dementia with Lewy bodies – memantine

DLB – memantine vs. placebo: adverse events

Quality assessment						No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo	Relative (95% CI)	Absolute (95% CI)	
Any adverse events (probability of experiencing ≥1; follow-up 24 weeks)										
1[1]	RCT	not serious	N/A	not serious	serious[2]	18/34 (52.9%)	17/41 (41.5%)	RR 1.28 (0.79 to 2.07)	116 more per 1000 (from 87 fewer to 444 more)	⊕⊕⊕O MODERATE
Serious adverse events (probability of experiencing ≥1; follow-up 24 weeks)										
1[1]	RCT	not serious	N/A	not serious	very serious[2,3]	6/34 (17.6%)	3/41 (7.3%)	RR 2.41 (0.65 to 8.93)	103 more per 1000 (from 26 fewer to 580 more)	⊕⊕OO LOW
Adverse events requiring treatment withdrawal (probability of experiencing; follow-up 24 weeks)										
1[1]	RCT	not serious	N/A	not serious	very serious[2,3]	5/34 (14.7%)	7/41 (17.1%)	RR 0.86 (0.3 to 2.47)	24 fewer per 1000 (from 120 fewer to 251 more)	⊕⊕OO LOW

[1] *Emre 2010; data reported for DLB population only; study also included people with PDD*
[2] *At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit or no difference*
[3] *Very small numbers of events*

DLB – memantine vs. placebo: cognitive outcomes

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo	Mean difference (95% CI)	
Clock drawing test (follow-up 24 weeks; range of scores: 0-10; higher is better)									
1[1]	RCT	not serious	N/A	not serious	serious[2]	33	43	1.3 higher (0.51 lower to 3.11 higher)	⊕⊕⊕O MODERATE

[1] Emre 2010; data reported for DLB population only; study also included people with PDD
[2] At a 95% confidence level, data are consistent with appreciable benefit, appreciable harm or no difference

DLB – memantine vs. placebo: global assessment

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo	Mean difference (95% CI)	
ADCS-CGIC (follow-up 24 weeks; lower is better)									
1[1]	RCT	not serious	N/A	not serious	serious[2]	33	41	0.6 lower (1.22 lower to 0.02 higher)	⊕⊕⊕O MODERATE

[1] Emre 2010; data reported for DLB population only; study also included people with PDD
[2] At a 95% confidence level, data are consistent with appreciable benefit, appreciable harm or no difference

DLB – memantine vs. placebo: activities of daily living

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo	Mean difference (95% CI)	
ADCS-ADL (follow-up 24 weeks; range of scores: 0-78; higher is better)									
1[1]	RCT	not serious	N/A	not serious	serious[2]	33	41	1.6 higher (4.9 lower to 8.1 higher)	⊕⊕⊕O MODERATE

[1] Emre 2010; data reported for DLB population only; study also included people with PDD
[2] Wide 95% confidence intervals, data are consistent with appreciable benefit, appreciable harm or no difference

DLB – memantine vs. placebo: carer-reported outcomes

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo	Mean difference (95% CI)	
ZBI (follow-up 24 weeks; lower is better)									
1[1]	RCT	not serious	N/A	not serious	serious[2]	33	41	1.4 lower (6.66 lower to 3.86 higher)	⊕⊕⊕O MODERATE

[1] Emre 2010; data reported for DLB population only; study also included people with PDD
[2] Wide 95% confidence intervals, data are consistent with appreciable benefit, appreciable harm or no difference

DLB – memantine vs. placebo: other non-cognitive outcomes

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo	Mean difference (95% CI)	
NPI-12 item (follow-up 24 weeks; range of scores: 0-144; lower is better)									
1[1]	RCT	not serious	N/A	not serious	serious[2]	33	41	6 lower (12.23 lower to 0.23 higher)	⊕⊕⊕O MODERATE
UPDRS III (follow-up 24 weeks; lower is better)									
1[1]	RCT	not serious	N/A	not serious	serious[2,3]	33	41	1.4 lower (5.52 lower to 2.72 higher)	⊕⊕⊕O MODERATE

[1] Emre 2010; data reported for DLB population only; study also included people with PDD
[2] Wide 95% confidence intervals, data are consistent with appreciable benefit, appreciable harm or no difference
[3]CI cross the MID between 3.25 (Horvath et al., 2015) and 5 points (Schrag et al., 2006)

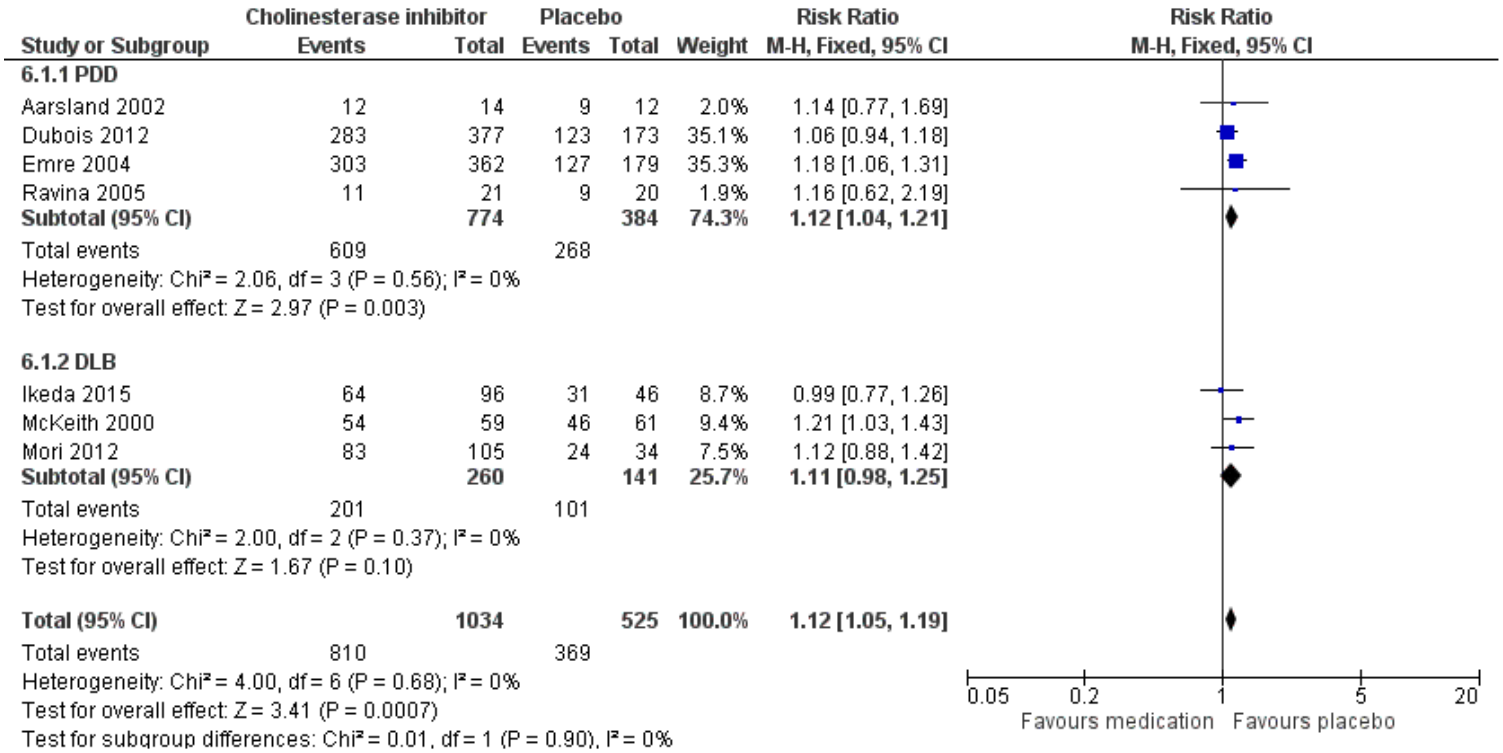
Mixed population (PDD or DLB) – cholinesterase inhibitors

PDD/DLB – cholinesterase inhibitor vs. placebo: adverse events

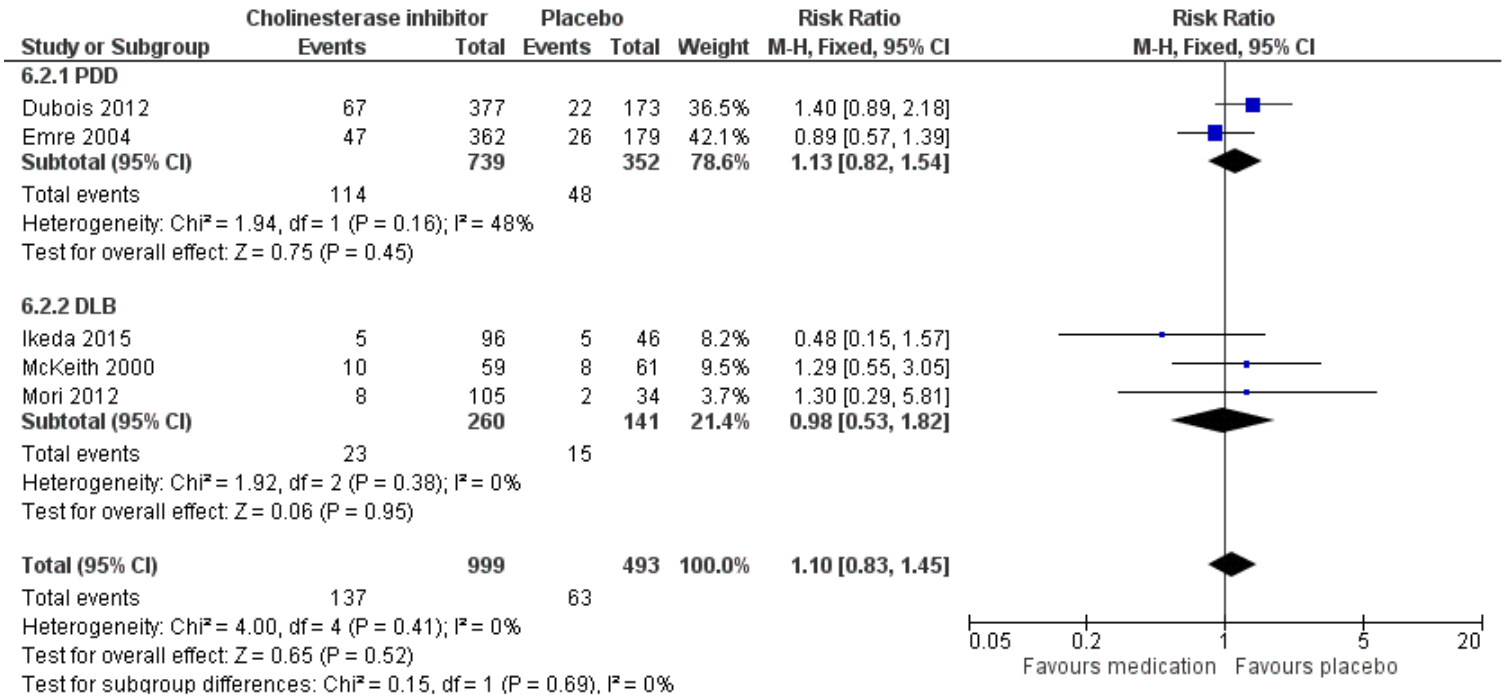
Quality assessment						No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	ChI	Placebo	Relative (95% CI)	Absolute (95% CI)	
Any adverse events – cholinesterase inhibitors (probability of experiencing ≥1; follow-up 10 to 24 weeks; lower is better); see Figure 27 for forest plot										
7[1–7]	RCT	not serious	not serious	not serious	not serious	810/1034 (78.3%)	369/525 (70.3%)	RR 1.12 (1.05 to 1.19)	84 more per 1000 (from 35 more to 134 more)	⊕⊕⊕⊕ HIGH
Any adverse events – donepezil (probability of experiencing ≥1; follow-up 10 to 24 weeks; lower is better)										
5[1,2,4,6,7]	RCT	not serious	not serious	not serious	serious[8]	453/613 (73.9%)	196/285 (68.8%)	RR 1.06 (0.97 to 1.16)	41 more per 1000 (from 21 fewer to 110 more)	⊕⊕⊕O MODERATE
Any adverse events – rivastigmine (probability of experiencing ≥1; follow-up 20 to 24 weeks; lower is better)										
2[3,5]	RCT	not serious	not serious	not serious	not serious	357/421 (84.8%)	173/240 (72.1%)	RR 1.19 (1.09 to 1.3)	137 more per 1000 (from 65 more to 216 more)	⊕⊕⊕⊕ HIGH
Serious adverse events – cholinesterase inhibitors (probability of experiencing ≥1; follow-up 12 to 24 weeks; lower is better); see Figure 28 for forest plot										
5[2–6]	RCT	not serious	not serious	not serious	serious[8]	137/999 (13.7%)	63/493 (12.8%)	RR 1.10 (0.83 to 1.45)	13 more per 1000 (from 22 fewer to 58 more)	⊕⊕⊕O MODERATE
Serious adverse events – donepezil (probability of experiencing ≥1; follow-up 12 to 24 weeks; lower is better)										
3[2,4,6]	RCT	not serious	not serious	not serious	serious[8]	80/578 (13.8%)	29/253 (11.5%)	RR 1.23 (0.83 to 1.84)	26 more per 1000 (from 19 fewer to 96 more)	⊕⊕⊕O MODERATE
Serious adverse events – rivastigmine (probability of experiencing ≥1; follow-up 20 to 24 weeks; lower is better)										
2[3,5]	RCT	not serious	not serious	not serious	serious[8]	57/421 (13.5%)	34/240 (14.2%)	RR 0.97 (0.65 to 1.43)	4 fewer per 1000 (from 50 fewer to 61 more)	⊕⊕⊕O MODERATE
Adverse events requiring treatment withdrawal – cholinesterase inhibitors (probability of experiencing; follow-up 10 to 24 weeks; lower is better); see Figure 29 for forest plot										
6[1–6]	RCT	not serious	not serious	not serious	not serious	147/1013 (14.5%)	49/505 (9.7%)	RR 1.50 (1.10 to 2.04)	49 more per 1000 (from 10 more to 101 more)	⊕⊕⊕⊕ HIGH

Adverse events requiring treatment withdrawal – donepezil (probability of experiencing; follow-up 10 to 24 weeks; lower is better)									
4[1,2,4,6]	RCT	not serious	not serious	not serious	serious[8]	78/592 (13.2%)	28/265 (10.6%)	RR 1.25 (0.84 to 1.87)	26 more per 1000 (from 17 fewer to 92 more)	⊕⊕⊕O MODERATE
Adverse events requiring treatment withdrawal – rivastigmine (probability of experiencing; follow-up 20 to 24 weeks; lower is better)									
2[3,5]	RCT	not serious	not serious	not serious	not serious	69/421 (16.4%)	21/240 (8.8%)	RR 1.88 (1.17 to 3.03)	77 more per 1000 (from 15 more to 178 more)	⊕⊕⊕⊕ HIGH

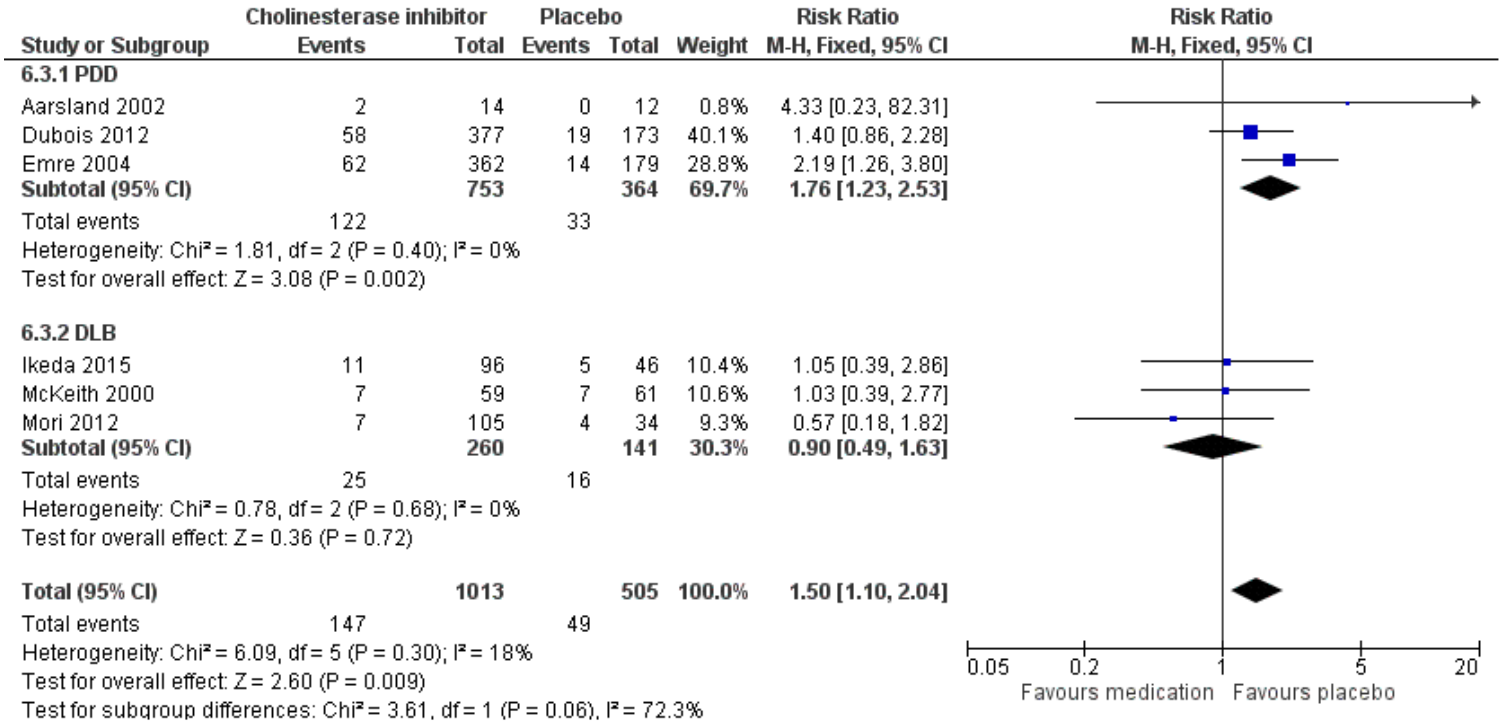
[1] *Aarsland 2002*
[2] *Dubois 2012; data for 2 active treatment groups were combined (donepezil 5mg and 10mg). Mean and standard deviation calculated from data reported in paper*
[3] *Emre 2004*
[4] *Ikeda 2015; data for 2 active treatment groups were combined (donepezil 5mg and 10mg)*
[5] *McKeith 2000*
[6] *Mori 2012; data for 3 active treatment groups were combined (donepezil 3mg, 5mg and 10mg)*
[7] *Ravina 2005*
[8] *At a 95% confidence level, data are consistent with appreciable benefit, appreciable harm or no difference*



PDD/DLB – cholinesterase inhibitor vs placebo: any adverse events (proportion of participants experiencing ≥1) – forest plot



PDD/DLB – cholinesterase inhibitor vs placebo: serious adverse events (proportion of participants experiencing ≥1) – forest plot

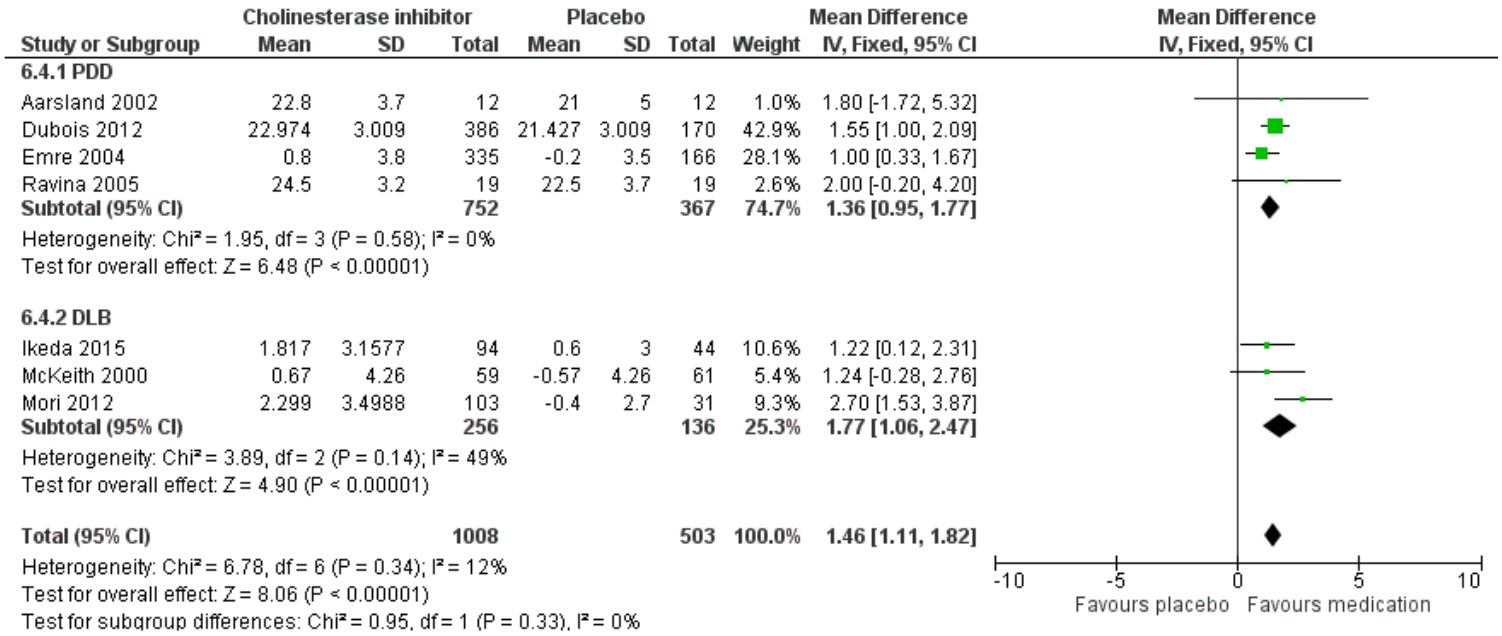


PDD/DLB – cholinesterase inhibitor vs placebo: adverse events requiring treatment withdrawal (proportion of participants experiencing) – forest plot

PDD/DLB – cholinesterase inhibitor vs. placebo: cognitive outcomes

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Chl	Placebo	Mean difference (95% CI)	
MMSE – cholinesterase inhibitors (follow-up 10 to 24 weeks; range of scores: 0-30; higher is better); see Figure 30 for forest plot									
7[1–7]	RCT	not serious	not serious	not serious	not serious	1008	503	1.46 higher (1.11 to 1.82 higher)	⊕⊕⊕⊕ HIGH
MMSE – donepezil (follow-up 10 to 24 weeks; range of scores: 0-30; higher is better)									
5[1,2,4,6,7]	RCT	not serious	not serious	not serious	not serious	614	276	1.68 higher (1.24 to 2.11 higher)	⊕⊕⊕⊕ HIGH
MMSE – rivastigmine (follow-up 20 to 24 weeks; range of scores: 0-30; higher is better)									
2[3,5]	RCT	not serious	not serious	not serious	not serious	394	227	1.04 higher (0.43 to 1.65 higher)	⊕⊕⊕⊕ HIGH

[1] *Aarsland 2002*
[2] *Dubois 2012; data for 2 active treatment groups were combined (donepezil 5mg and 10mg). Mean and standard deviation calculated from data reported in paper*
[3] *Emre 2004*
[4] *Ikeda 2015; data for 2 active treatment groups were combined (donepezil 5mg and 10mg)*
[5] *McKeith 2000*
[6] *Mori 2012; data for 3 active treatment groups were combined (donepezil 3mg, 5mg and 10mg)*
[7] *Ravina 2005*

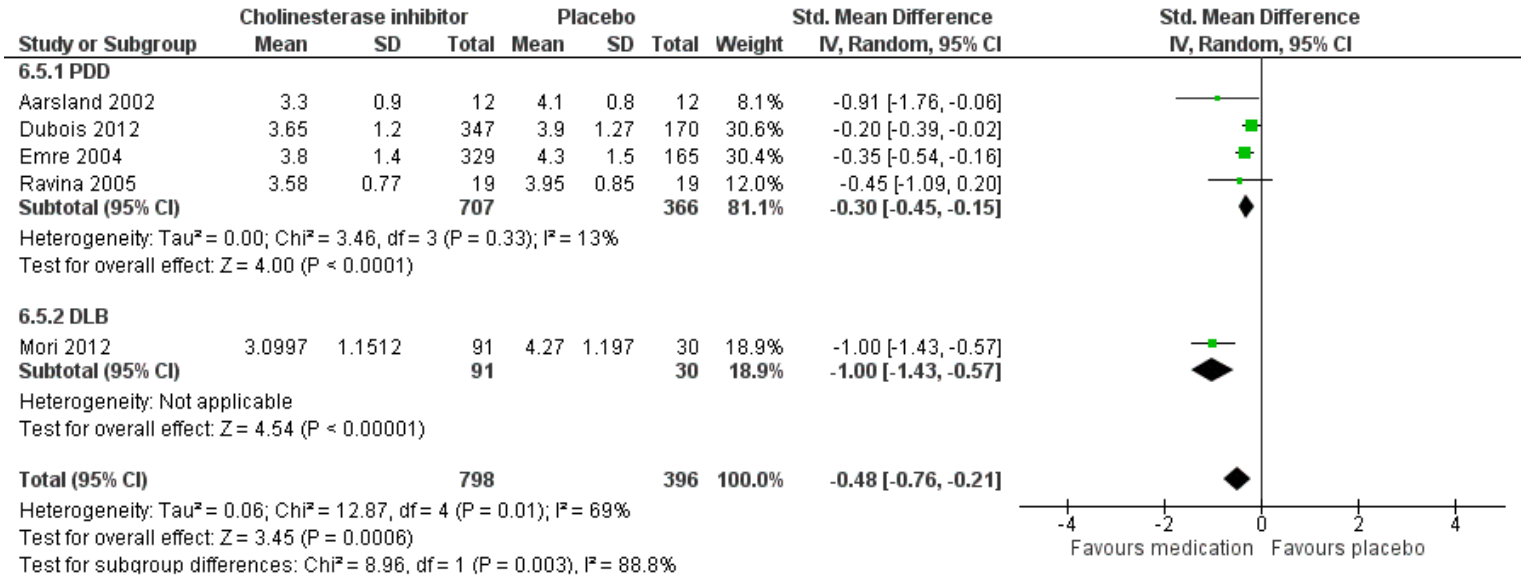


PDD/DLB – cholinesterase inhibitor vs placebo: MMSE – forest plot

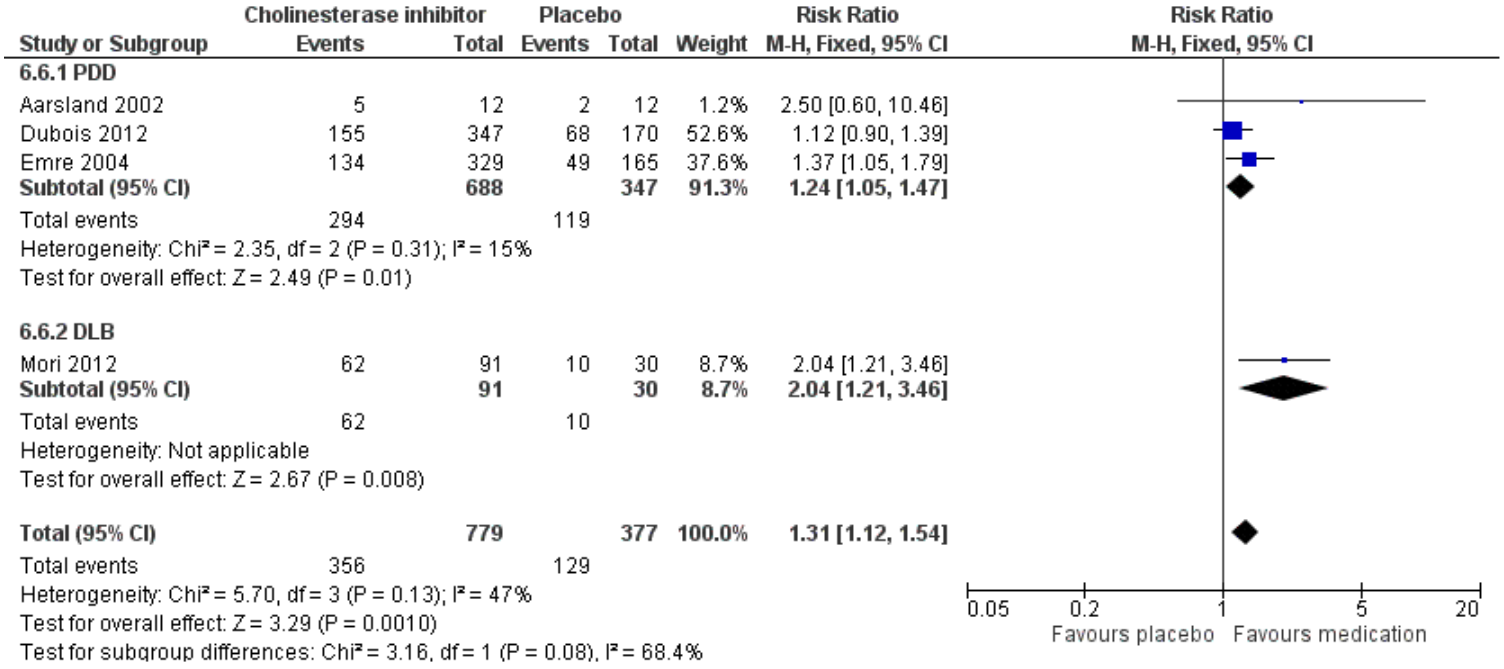
PDD/DLB – cholinesterase inhibitor vs. placebo: global assessment

Quality assessment						No of patients		Effect (95% CI)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	ChI	Placebo		
Global function – cholinesterase inhibitors (follow-up 10 to 24 weeks; measured with: CIBIC+, ADCS-CGIC or CGIC; range of scores: 1-7; lower is better); see Figure 31 for forest plot									
5[1–5]	RCT	not serious	serious[6]	not serious	not serious	798	396	SMD 0.48 lower (0.76 to 0.21 lower)	⊕⊕⊕O MODERATE
Global function – donepezil (follow-up 10 to 24 weeks; measured with: CIBIC+, ADCS-CGIC or CGIC; range of scores: 1-7; lower is better)									
4[1,2,3,5]	RCT	not serious	serious[6]	not serious	not serious	469	231	SMD 0.6 lower (1.08 to 0.11 lower)	⊕⊕⊕O MODERATE
Global response – cholinesterase inhibitors (at least minimal improvement; follow-up 10 to 24 weeks; measured with: CIBIC+ or ADCS-CGIC; higher is better); see Figure 32 for forest plot									
4[1–4]	RCT	not serious	not serious	not serious	not serious	356/779 (45.7%)	129/377 (34.2%)	RR 1.31 (1.12 to 1.54) 106 more per 1000 (from 41 more to 185 more)	⊕⊕⊕⊕ HIGH
Global response – donepezil (at least minimal improvement; follow-up 10 to 24 weeks; measured with: CIBIC+ or ADCS-CGIC; higher is better)									
3[1,2,4]	RCT	not serious	serious[6]	not serious	not serious	222/450 (49.3%)	80/212 (37.7%)	RR 1.27 (1.04 to 1.55) 102 more per 1000 (from 15 more to 208 more)	⊕⊕⊕O MODERATE

[1] Aarsland 2002
[2] Dubois 2012; data for 2 active treatment groups were combined (donepezil 5mg and 10mg). Mean and standard deviation calculated from data reported in paper
[3] Emre 2004
[4] Mori 2012; data for 3 active treatment groups were combined (donepezil 3mg, 5mg and 10mg)
[5] Ravina 2005
[6] Heterogeneity >40% between studies



PDD/DLB – cholinesterase inhibitor vs placebo: global function (different measures) – forest plot

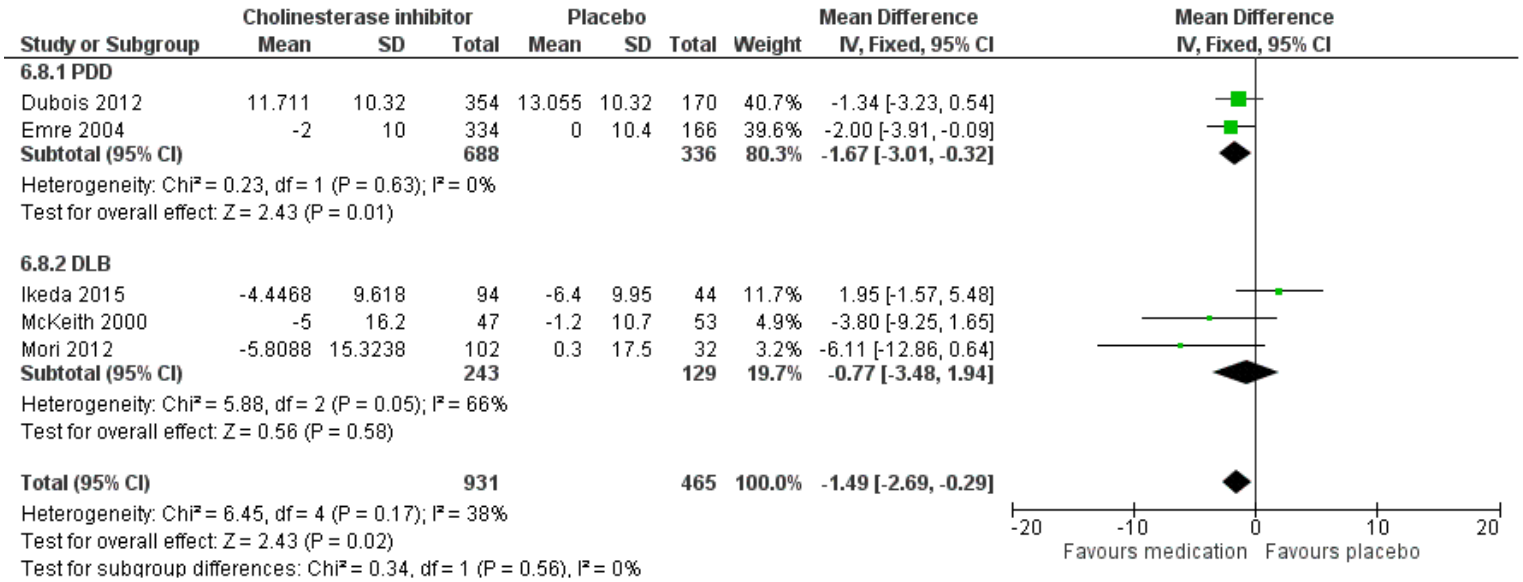


PDD/DLB – cholinesterase inhibitor vs placebo: global response (at least minimal improvement) – forest plot

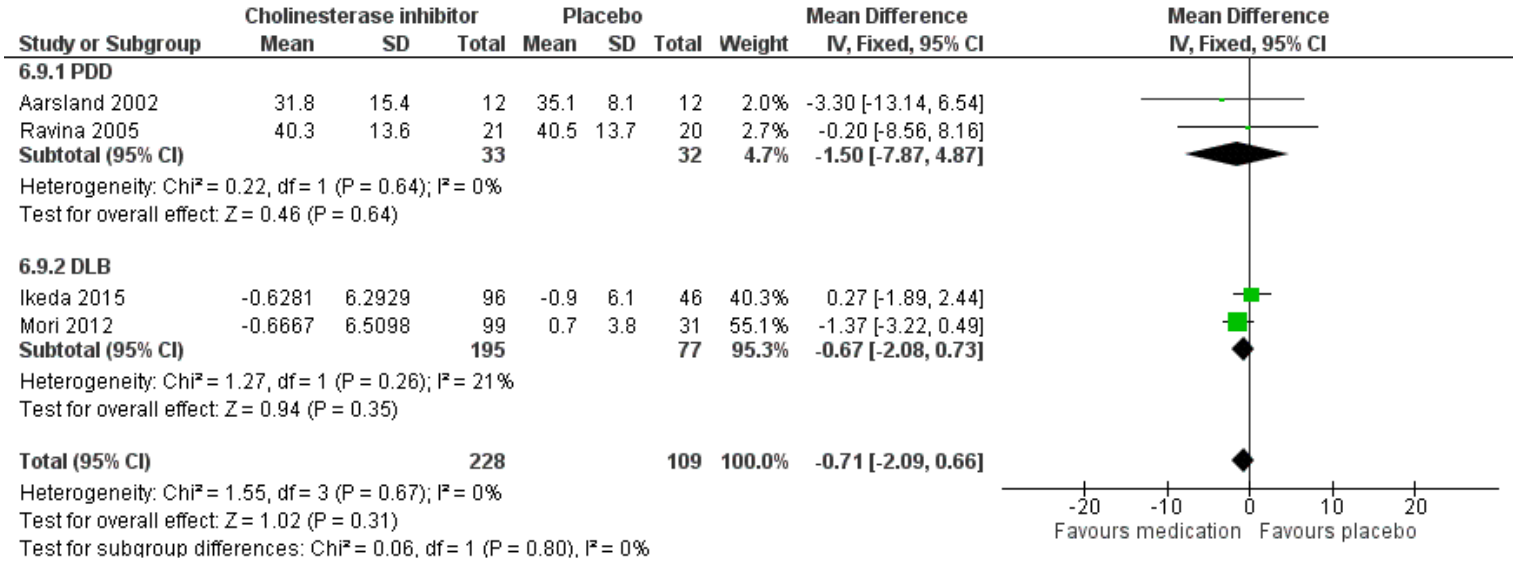
PDD/DLB – cholinesterase inhibitor vs. placebo: other non-cognitive outcomes

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Chl	Placebo	Mean difference (95% CI)	
NPI-10 item – cholinesterase inhibitors (follow-up 12 to 24 weeks; range of scores: 0-120; lower is better)[1]; see Figure 33 for forest plot									
5[2–6]	RCT	not serious[7]	not serious	not serious	not serious	931	465	1.49 lower (2.69 to 0.29 lower)	⊕⊕⊕⊕ HIGH
NPI-10 item – donepezil (follow-up 12 to 24 weeks; range of scores: 0-120; lower is better)[1]									
3[2,4,6]	RCT	not serious[7]	serious[8]	not serious	serious[9]	550	246	0.92 lower (2.54 lower to 0.69 higher)	⊕⊕OO LOW
NPI-10 item – rivastigmine (follow-up 20 to 24 weeks; range of scores: 0-120; lower is better)									
2[3,5]	RCT	not serious	not serious	not serious	not serious	381	219	2.2 lower (4 to 0.39 lower)	⊕⊕⊕⊕ HIGH
UPDRS III – donepezil (follow-up 24 weeks; lower is better); see Figure 34 for forest plot									
4[4,6,10,11]	RCT	serious[12]	not serious	not serious	not serious[13]	228	109	0.71 lower (2.09 lower to 0.66 higher)	⊕⊕⊕O MODERATE

[1] SD not reported for this outcome in Ikeda 2015; calculated from SE reported in paper
[2] Dubois 2012; data for 2 active treatment groups were combined (donepezil 5mg and 10mg). Mean and standard deviation calculated from data reported in paper
[3] Emre 2004
[4] Ikeda 2015; data for 2 active treatment groups were combined (donepezil 5mg and 10mg)
[5] McKeith 2000
[6] Mori 2012; data for 3 active treatment groups were combined (donepezil 3mg, 5mg and 10mg)
[7] Data for this outcome not reported in Aarsland 2002. This represents a very small proportion of the total participants in the analysis, therefore quality assessment not downgraded
[8] Heterogeneity > 40% between studies
[9] At a 95% confidence level, data are consistent with appreciable benefit, appreciable harm or no difference
[10] Aarsland 2002
[11] Ravina 2005
[12]Data for outcome not reported in 3 large RCTs (Dubois 2012, Emre 2004 and McKeith 2000). Papers stated no significant difference between groups
[13]CI do not cross the MID between 3.25 (Horvath et al., 2015) and 5 points (Schrag et al., 2006)



PDD/DLB – cholinesterase inhibitor vs placebo: NPI-10 item – forest plot



PDD/DLB – cholinesterase inhibitor vs placebo: UPDRS III – forest plot

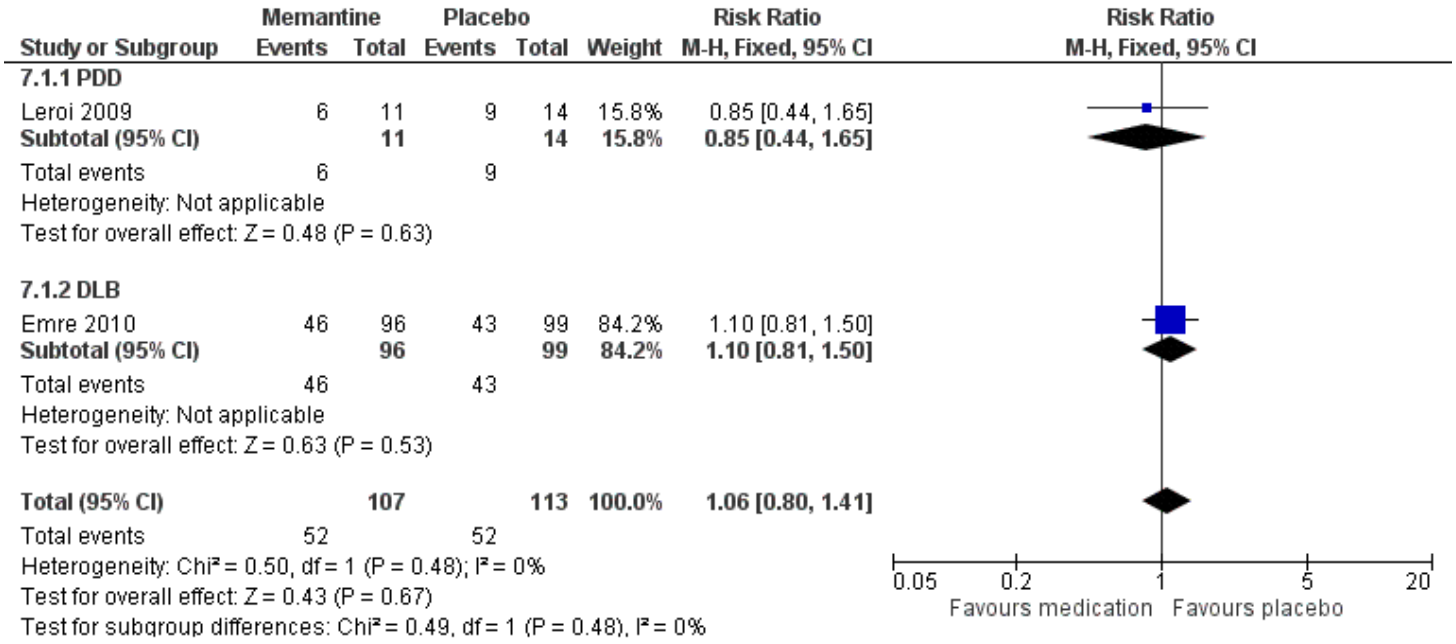
Mixed population (PDD or DLB) – memantine

PDD/DLB – memantine vs. placebo: adverse events

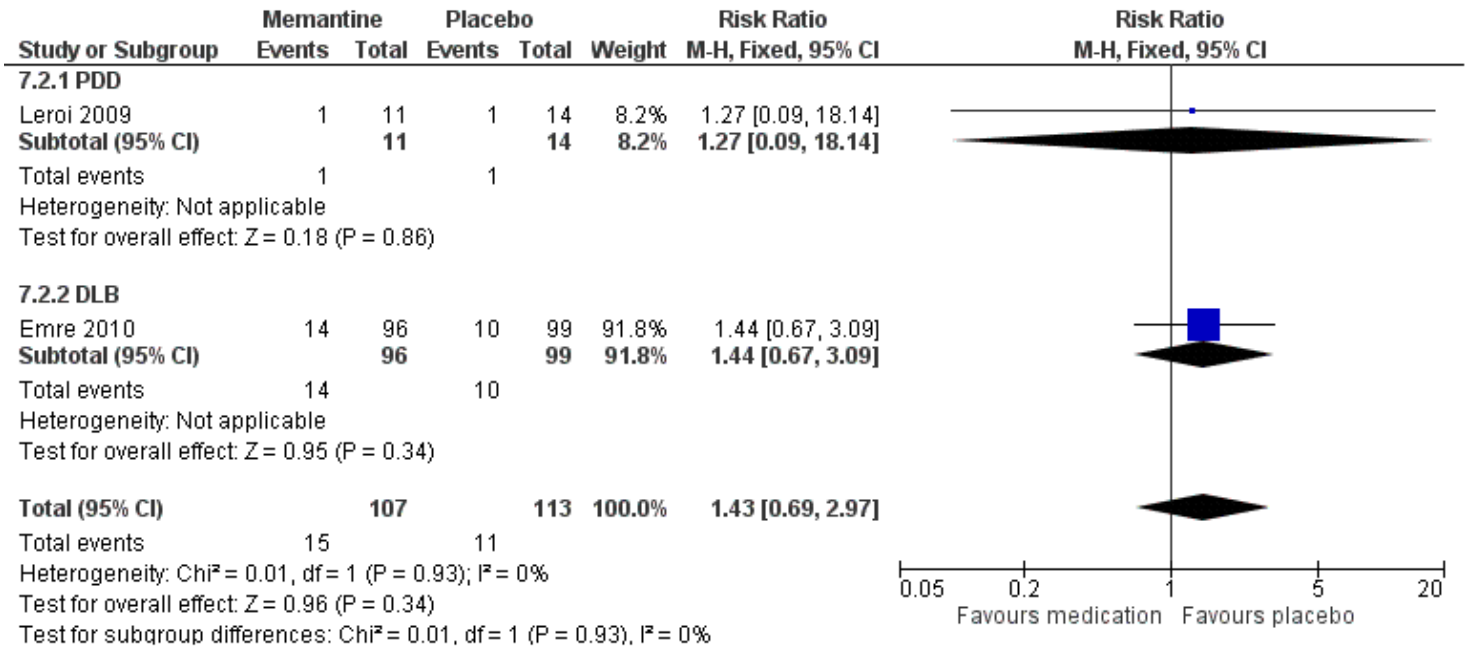
Quality assessment						No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo	Relative (95% CI)	Absolute (95% CI)	
Any adverse events (probability of experiencing ≥1; follow-up 16 to 24 weeks; lower is better); see Figure 35 for forest plot										
2[1,2]	RCT	not serious	not serious	not serious	serious[3]	52/107 (48.6%)	52/113 (46%)	RR 1.06 (0.8 to 1.41)	28 more per 1000 (from 92 fewer to 189 more)	⊕⊕⊕O MODERATE
Serious adverse events (probability of experiencing ≥1; follow-up 16 to 24 weeks; lower is better); see Figure 36 for forest plot										
2[1,2]	RCT	not serious	not serious	not serious	serious[3]	15/107 (14%)	11/113 (9.7%)	RR 1.43 (0.69 to 2.97)	42 more per 1000 (from 30 fewer to 192 more)	⊕⊕⊕O MODERATE
Adverse events requiring treatment withdrawal (probability of experiencing; follow-up 16 to 24 weeks; lower is better); see Figure 37 for forest plot										
2[2,4]	RCT	not serious	not serious	serious[5]	serious[3]	18/130 (13.8%)	21/137 (15.3%)	RR 0.91 (0.51 to 1.63)	14 fewer per 1000 (from 75 fewer to 97 more)	⊕⊕OO LOW

[1] *Emre 2010; data reported for total population (PDD and DLB)*

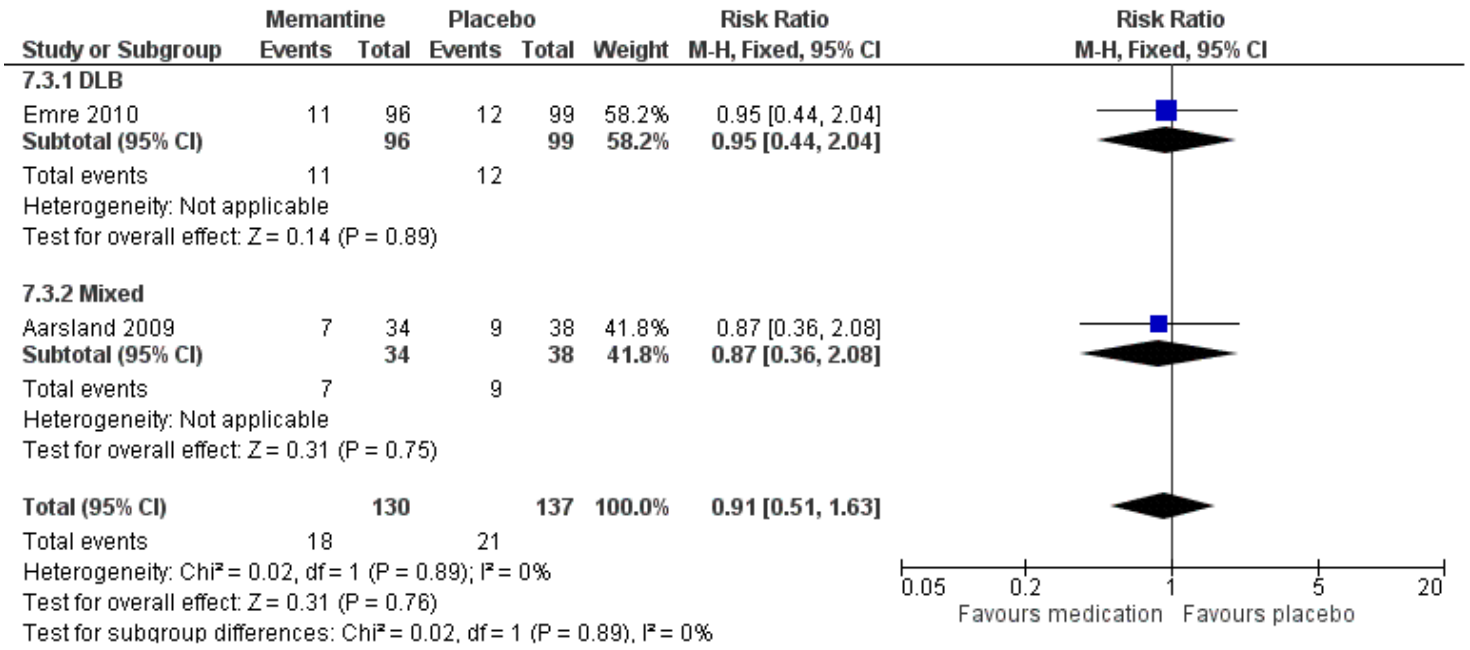
[2] *Leroi 2009; not clear if adverse event data reported at end of active treatment (16 weeks) or end of drug withdrawal phase (22 weeks)*
[3] *At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit or no difference*
[4] *Aarsland 2009*
[5] *Both studies included people who were also taking a cholinesterase inhibitor*



PDD/DLB – memantine vs placebo: any adverse events (proportion of participants experiencing ≥1) – forest plot



PDD/DLB – memantine vs placebo: serious adverse events (proportion of participants experiencing ≥1) – forest plot

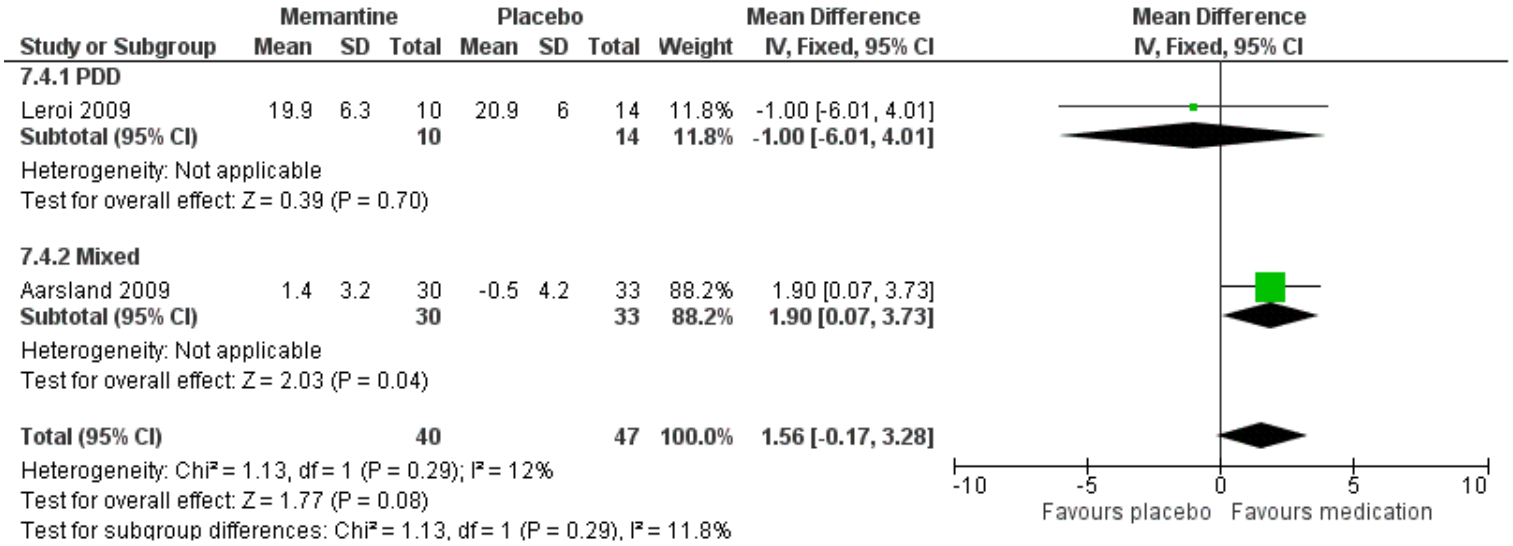


PDD/DLB – memantine vs placebo: adverse events requiring treatment withdrawal (proportion of participants experiencing) – forest plot

PDD/DLB – memantine vs. placebo: cognitive outcomes

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo	Mean difference (95% CI)	
MMSE (follow-up 16 to 24 weeks; range of scores: 0-30; higher is better); see Figure 38 for forest plot									
2[1,2]	RCT	not serious	not serious	serious[3]	serious[3]	40	47	1.56 higher (0.17 lower to 3.28 higher)	⊕⊕OO LOW

[1] Aarsland 2009
[2] Leroi 2009; data reported for end of drug treatment phase (16 weeks)
[3] Both studies included people who were also taking a cholinesterase inhibitor
[4] At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit or no difference

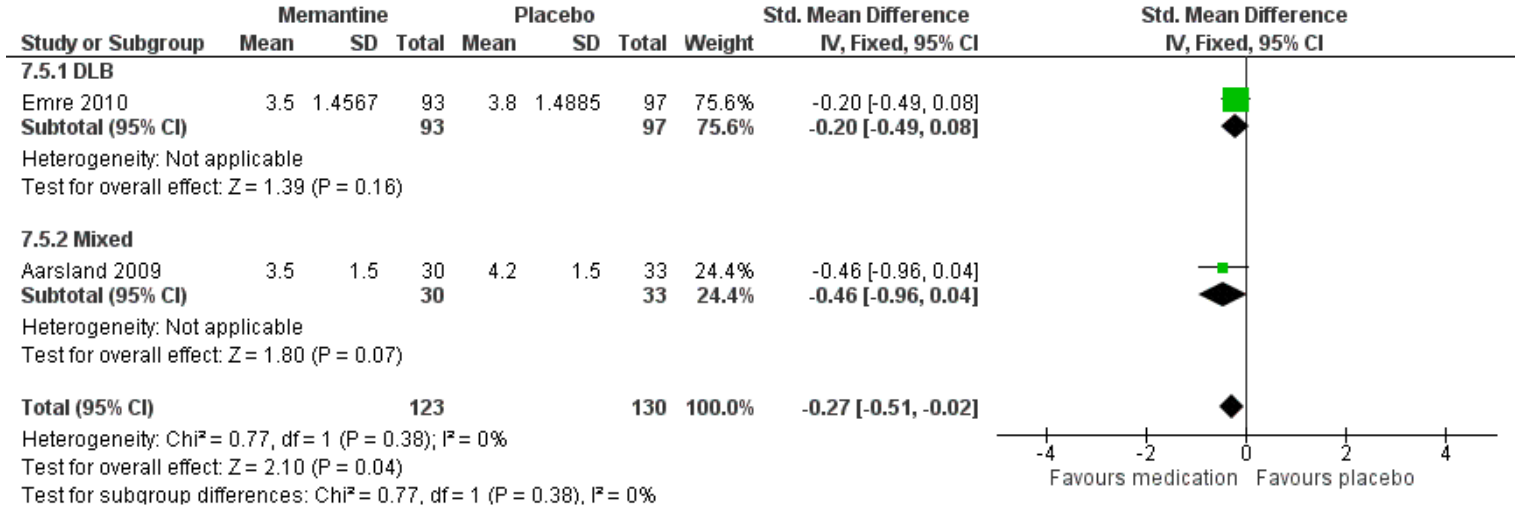


PDD/DLB – memantine vs placebo: MMSE – forest plot

PDD/DLB – memantine vs. placebo: global assessment

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo	Standardised mean difference (95% CI)	
Global function (follow-up 24 weeks; measured with: ADCS-CGIC or CGIC; range of scores: 1-7; lower is better); see Figure 39 for forest plot									
2[1,2]	RCT	not serious	not serious	not serious	not serious	123	130	0.27 lower (0.51 to 0.02 lower)	⊕⊕⊕⊕ HIGH

[1] *Aarsland 2009*
[2] *Emre 2010; data reported for total population (PDD and DLB)*

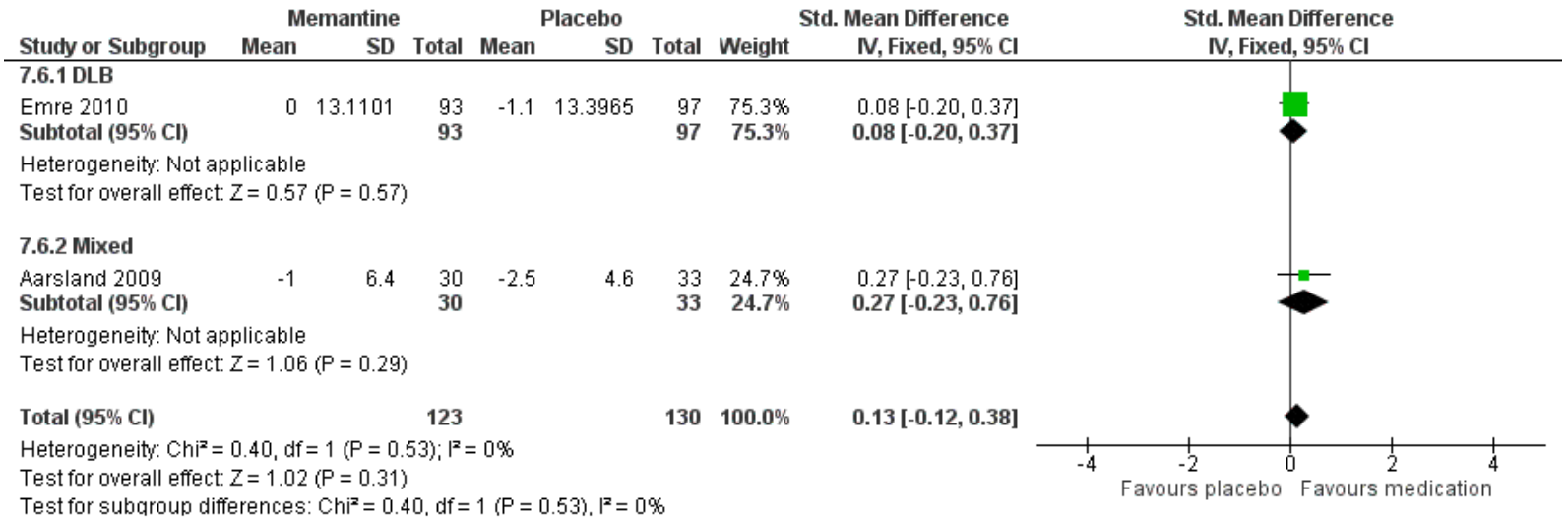


PDD/DLB – memantine vs placebo: global function (different measures) – forest plot

PDD/DLB – memantine vs. placebo: activities of daily living

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo	Standardised mean difference (95% CI)	
ADL (follow-up 24 weeks; measured with: ADCS-ADL or DAD; higher is better); see Figure 40 for forest plot									
2[1,2]	RCT	not serious	not serious	not serious	serious[3]	123	130	0.13 higher (0.12 lower to 0.38 higher)	⊕⊕⊕O MODERATE

[1] Aarsland 2009
[2] Emre 2010; data reported for total population (PDD and DLB)
[3] At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit or no difference

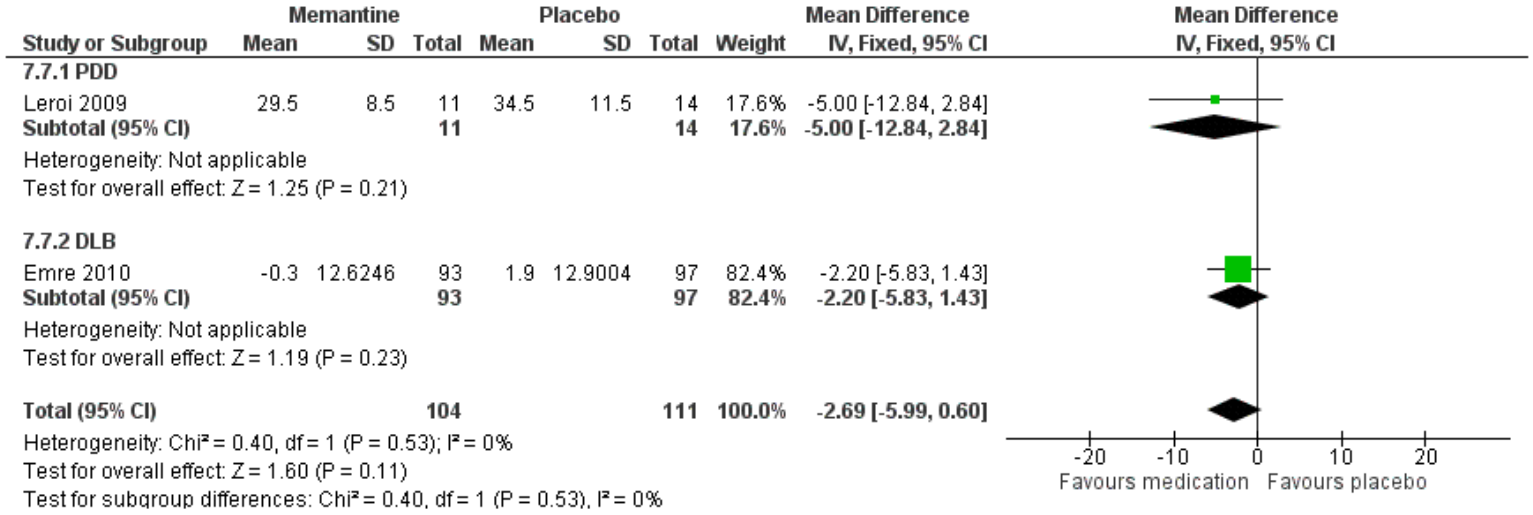


PDD/DLB – memantine vs placebo: activities of daily living (different measures) – forest plot

PDD/DLB – memantine vs. placebo: carer-reported outcomes

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo	Mean difference (95% CI)	
ZBI (follow-up 16 to 24 weeks; lower is better); see Figure 41 for forest plot									
2[1,2]	RCT	not serious	not serious	not serious	serious[3]	104	111	2.69 lower (5.99 lower to 0.6 higher)	⊕⊕⊕O MODERATE

[1] Emre 2010; data reported for total population (PDD and DLB)
[2] Leroi 2009; data reported for end of drug treatment phase (16 weeks)
[3] At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit or no difference

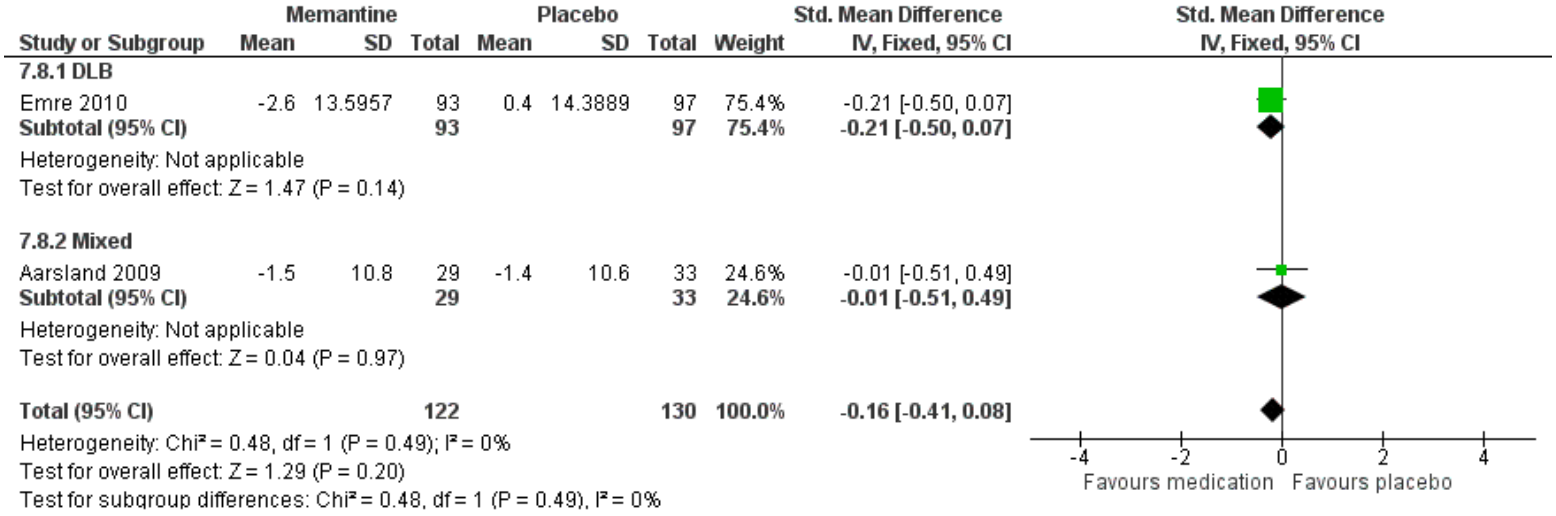


PDD/DLB – memantine vs placebo: ZBI – forest plot

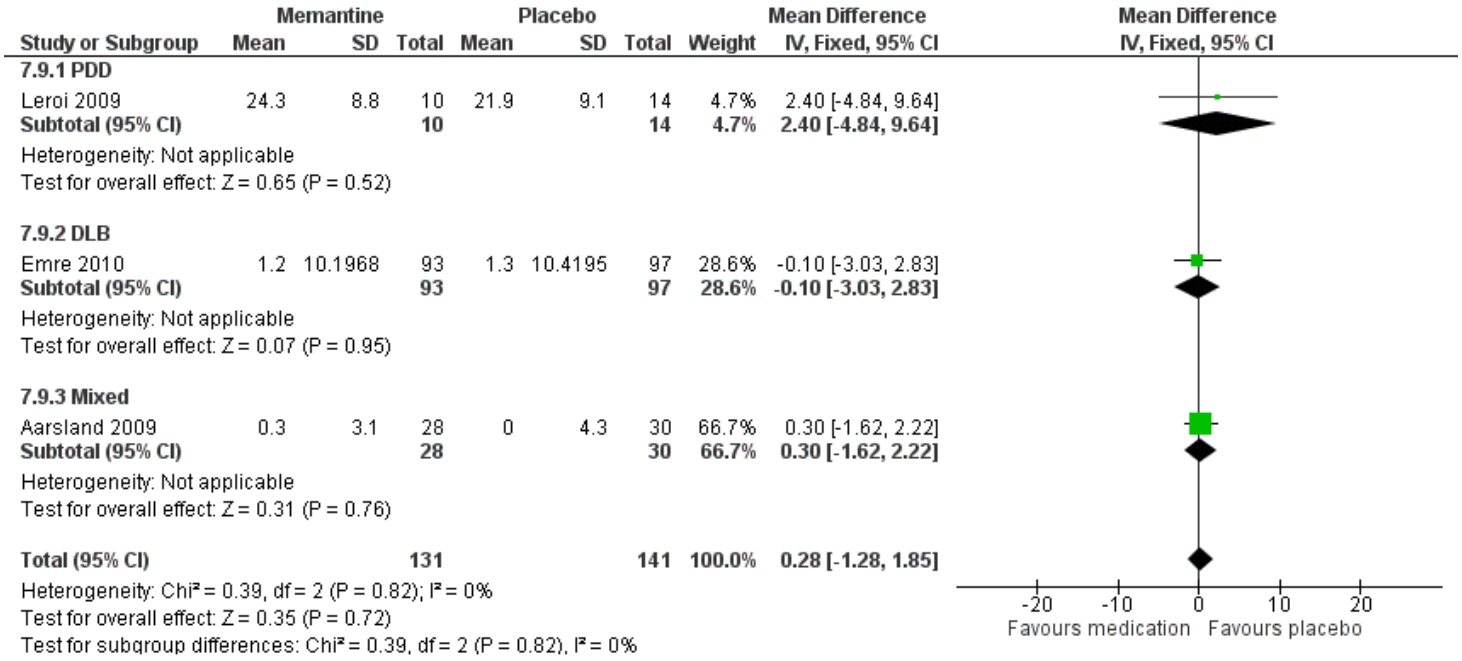
PDD/DLB – memantine vs. placebo: other non-cognitive outcomes

Quality assessment						No of patients		Effect (95% CI)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Memantine	Placebo		
NPI (follow-up 16 to 24 weeks; measured with: NPI-10 item or NPI 12-item; lower is better)[1]; see Figure 42 for forest plot									
2[2,3]	RCT	not serious	not serious	not serious	serious[4]	122	130	SMD 0.16 lower (0.41 lower to 0.08 higher)	⊕⊕⊕O MODERATE
UPDRS III (follow-up 16 to 24 weeks; lower is better); see Figure 43 for forest plot									
2[2,3]	RCT	not serious	not serious	not serious	not serious[5]	131	141	MD 0.28 higher (1.28 lower to 1.85 higher)	⊕⊕⊕⊕ HIGH

[1] Data from Leroi 2009 could not be included in this analysis due to inconsistent outcome reporting
[2] Aarsland 2009
[3] Emre 2010; data reported for total population (PDD and DLB)
[4] At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit or no difference
[5]CI do not cross the MID between 3 (Horvath et al., 2015) and 5 points (Schrag et al., 2006)



PDD/DLB – memantine vs placebo: NPI (different measures) – forest plot



PDD/DLB – memantine vs placebo: UPDRS III – forest plot

Network meta-analyses

Any adverse events

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Adverse events					
9 Aarsland 2002, Dubois 2012, Ikeda 2015, Mori 2012, Ravina 2005, Emre 2004, McKeith 2000, Emre 2010, Leroi 2009	Not serious	Not serious	Not serious[1]	Not serious	High
[1]Considered not serious as population, interventions, comparator and outcomes are as defined in protocol					

Serious adverse events

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Serious adverse events					
7 Dubois 2012, Ikeda 2015, Mori 2012, Emre 2004, McKeith 2000, Emre 2010, Leroi 2009	Not serious	Not serious	Not serious[1]	Not serious	High
[1]Considered not serious as population, interventions, comparator and outcomes are as defined in protocol					

Adverse events requiring treatment withdrawal

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Adverse events requiring treatment withdrawal					
8 Aarsland 2002, Dubois 2012, Ikeda 2015, Mori 2012, Emre 2004, McKeith 2000, Aarsland 2009, Emre 2010	Not serious	Not serious	Not serious[1]	Not serious	High
[1]Considered not serious as population, interventions, comparator and outcomes are as defined in protocol					

MMSE

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in MMSE scores					
9 Aarsland 2002, Dubois 2012, Ikeda 2015, Mori 2012, Ravina 2005, Emre 2004, McKeith 2000, Aarsland 2009, Emre 2010	Not serious	Not serious	Not serious[1]	Not serious	High
[1]Considered not serious as population, interventions, comparator and outcomes are as defined in protocol					

Clincial global function

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in clinical global function (various measures)					
7	Not serious	Serious[1]	Not serious[2]	Not serious	Moderate

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Aarsland 2002, Dubois 2012, Mori 2012, Ravina 2005, Emre 2004, Aarsland 2009, Emre 2010					

[1]Considerable between study heterogeneity (i^2>40%)
[2]Considered not serious as population, interventions, comparator and outcomes are as defined in protocol

NPI

Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in NPI scores					
8 Dubois 2012, Ikeda 2015, Mori 2012, Emre 2004, McKeith 2000, Aarsland 2009, Emre 2010, Leroi 2009	Not serious	Not serious	Not serious[1]	Not serious	High

[1]Considered not serious as population, interventions, comparator and outcomes are as defined in protocol

UPDRS III (motor subscale)

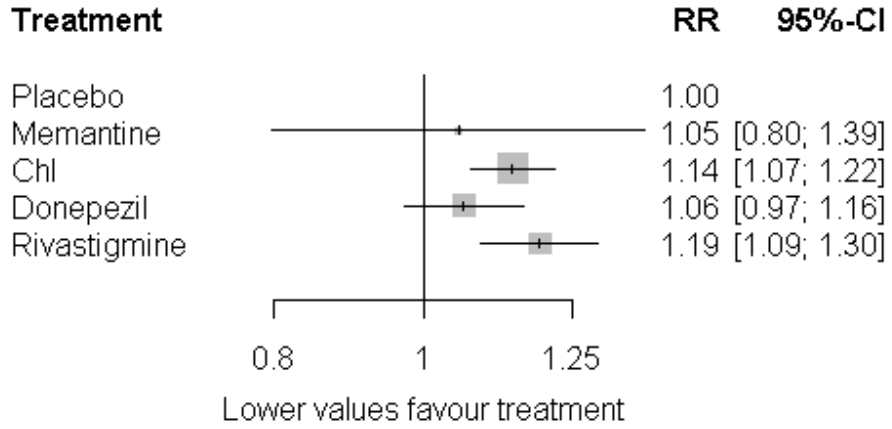
Quality assessment					
Number of RCTs	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Quality**
Change in UPDRS III (motor) scores					
7 Aarsland 2002, Ikeda 2015, Mori 2012, Ravina 2005, Aarsland 2009, Emre 2010, Leroi 2009	Serious[1]	Not serious	Not serious[2]	Serious[3]	Low

[1]Some studies do not report measure of variation
[2]Considered not serious as population, interventions, comparator and outcomes are as defined in protocol
[3]Analysis could not differentiate between any clinically distinct options

Network meta-analyses

Mixed population (PDD or DLB)

PDD/DLB – any adverse events – FE model



Differences between treatments – relative risk and 95% confidence interval

	Placebo	Memantine	Chl	Donepezil	Rivastigmine
Placebo	N/A				
Memantine	1.05 (0.80, 1.39)	N/A			
Chl	1.14 (1.07, 1.22)	1.08 (0.81, 1.44)	N/A		
Donepezil	1.06 (0.97, 1.16)	1.01 (0.75, 1.35)	N/A	N/A	
Rivastigmine	1.19 (1.09, 1.30)	1.13 (0.84, 1.51)	N/A	1.12 (0.99, 1.27)	N/A

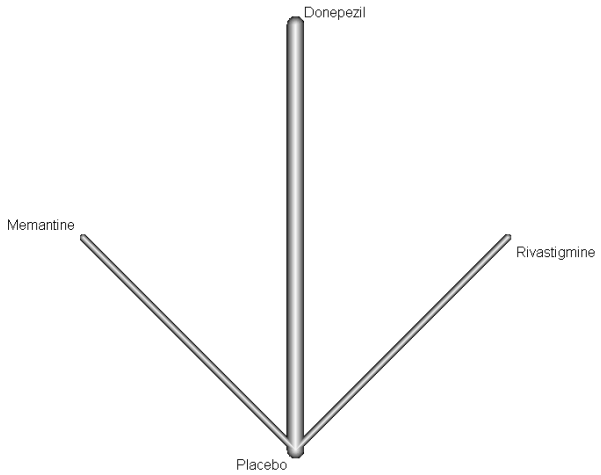
Quantifying heterogeneity/inconsistency:

tau^2 < 0.0001; I^2 = 0%

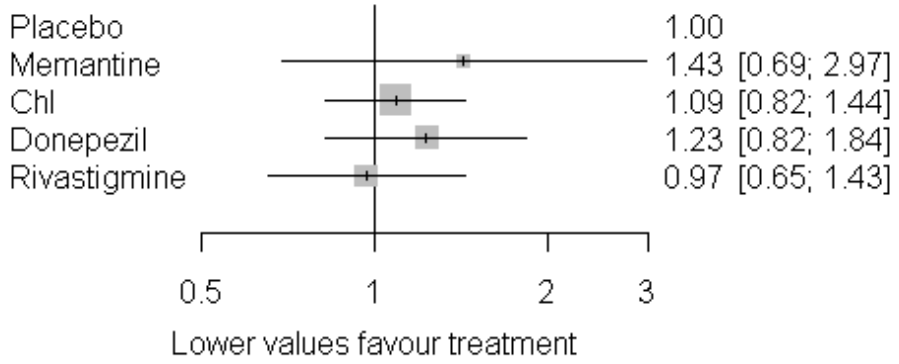
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
1.31	6	0.971

Network graph:



PDD/DLB – serious adverse events – FE model



Differences between treatments – relative risk and 95% confidence interval

	Placebo	Memantine	Chl	Donepezil	Rivastigmine
Placebo	N/A				
Memantine	1.43 (0.69, 2.97)	N/A			
Chl	1.09 (0.82, 1.44)	0.76 (0.35, 1.67)	N/A		
Donepezil	1.23 (0.82, 1.84)	0.86 (0.37, 1.98)	N/A	N/A	
Rivastigmine	0.97 (0.65, 1.43)	0.68 (0.29, 1.55)	N/A	0.79 (0.45, 1.38)	N/A

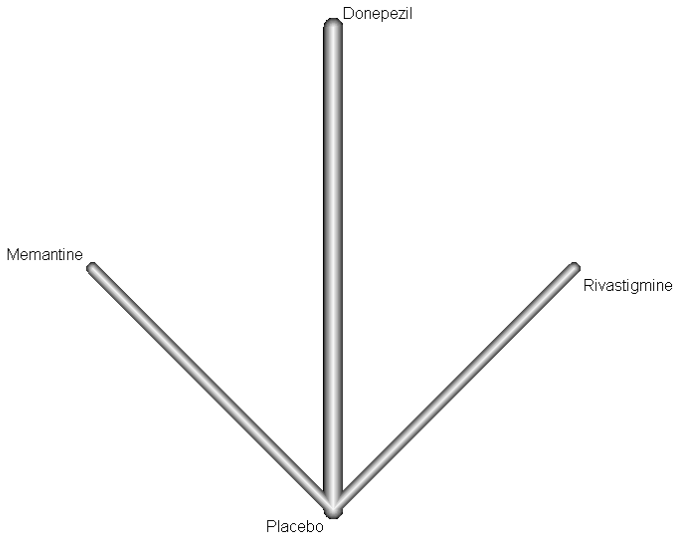
Quantifying heterogeneity/inconsistency:

tau^2 < 0.0001; I^2 = 0%

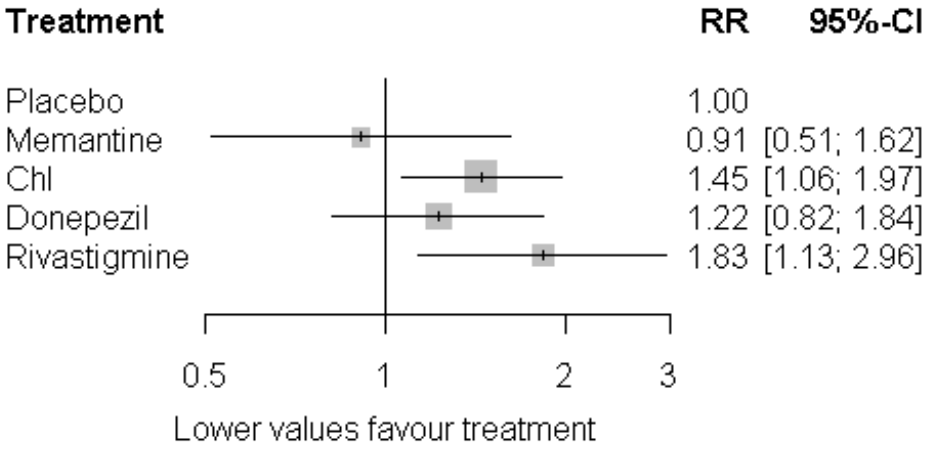
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
3.3	4	0.5087

Network graph:



PDD/DLB – adverse events requiring treatment withdrawal – FE model



Differences between treatments – relative risk and 95% confidence interval

	Placebo	Memantine	Chl	Donepezil	Rivastigmine
Placebo	N/A				
Memantine	0.91 (0.51, 1.62)	N/A			
Chl	1.45 (1.06, 1.97)	1.59 (0.82, 3.05)	N/A		
Donepezil	1.22 (0.82, 1.84)	1.34 (0.66, 2.72)	N/A	N/A	
Rivastigmine	1.83 (1.13, 2.96)	2.01 (0.95, 4.26)	N/A	1.50 (0.80, 2.80)	N/A

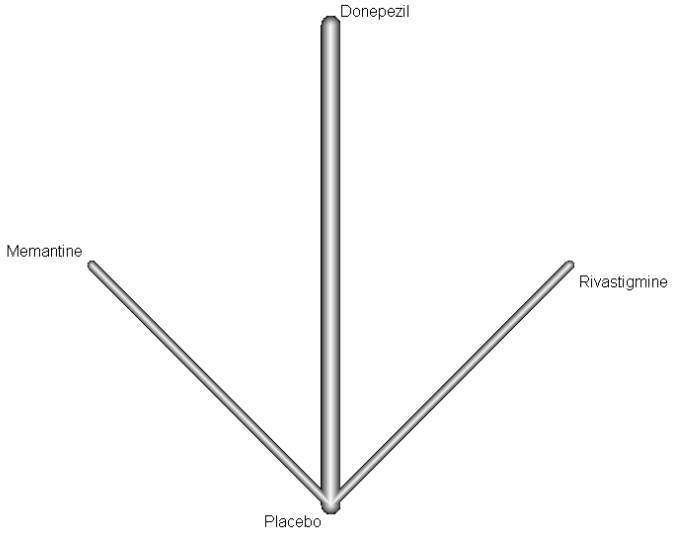
Quantifying heterogeneity/inconsistency:

tau^2 < 0.0001; I^2 = 0%

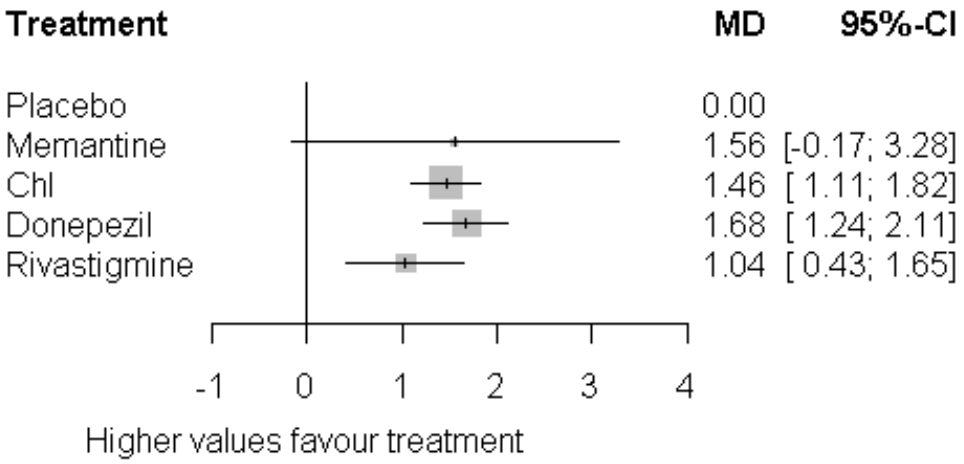
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
4.49	5	0.4819

Network graph:



PDD/DLB – MMSE – FE model



Differences between treatments – mean difference and 95% confidence interval

	Placebo	Memantine	Chl	Donepezil	Rivastigmine
Placebo	N/A				
Memantine	1.56 (-0.17, 3.28)	N/A			
Chl	1.46 (1.11, 1.82)	-0.09 (-1.85, 1.66)	N/A		
Donepezil	1.68 (1.24, 2.11)	0.12 (-1.66, 1.90)	N/A	N/A	
Rivastigmine	1.04 (0.43, 1.65)	-0.52 (-2.35, 1.31)	N/A	-0.64 (-1.39, 0.11)	N/A

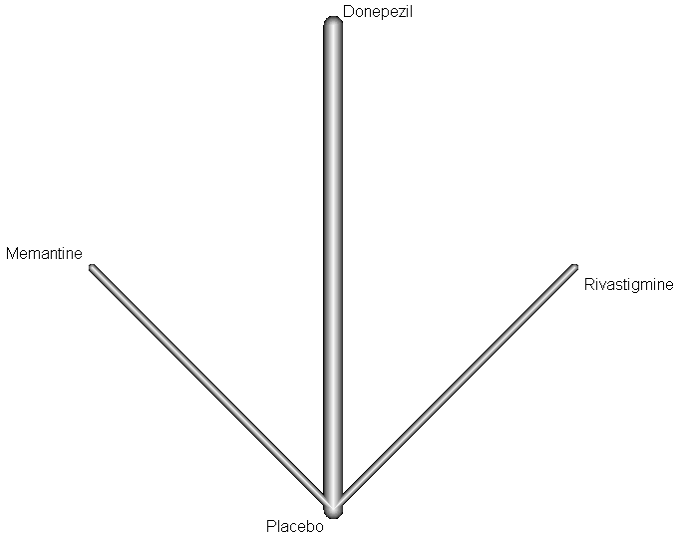
Quantifying heterogeneity/inconsistency:

tau^2 < 0.0001; I^2 = 0%

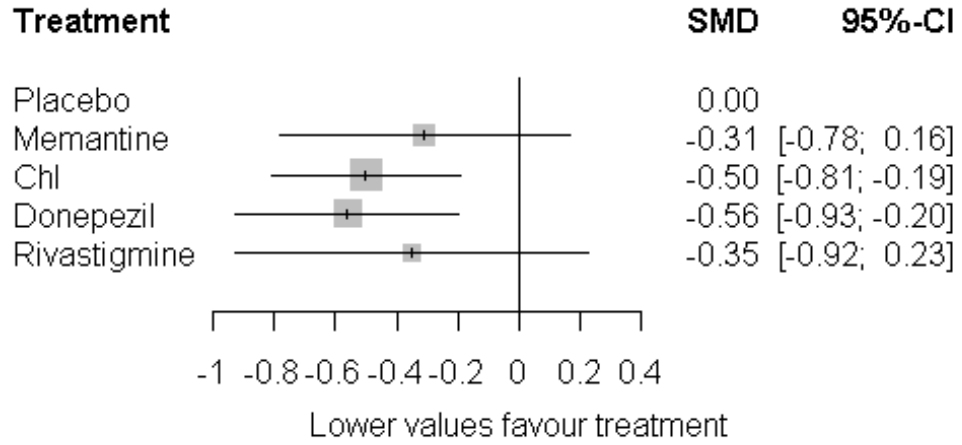
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
5.15	6	0.5243

Network graph:



PDD/DLB – global function – RE model



Differences between treatments – standardised mean difference and 95% confidence interval

	Placebo	Memantine	Chl	Donepezil	Rivastigmine
Placebo	N/A				
Memantine	-0.31 (-0.78, 0.16)	N/A			
Chl	-0.50 (-0.81, -0.19)	-0.19 (-0.76, 0.37)	N/A		
Donepezil	-0.56 (-0.93, -0.20)	-0.25 (-0.85, 0.34)	N/A	N/A	
Rivastigmine	-0.35 (-0.92, 0.23)	-0.04 (-0.78, 0.70)	N/A	0.21 (-0.47, 0.90)	N/A

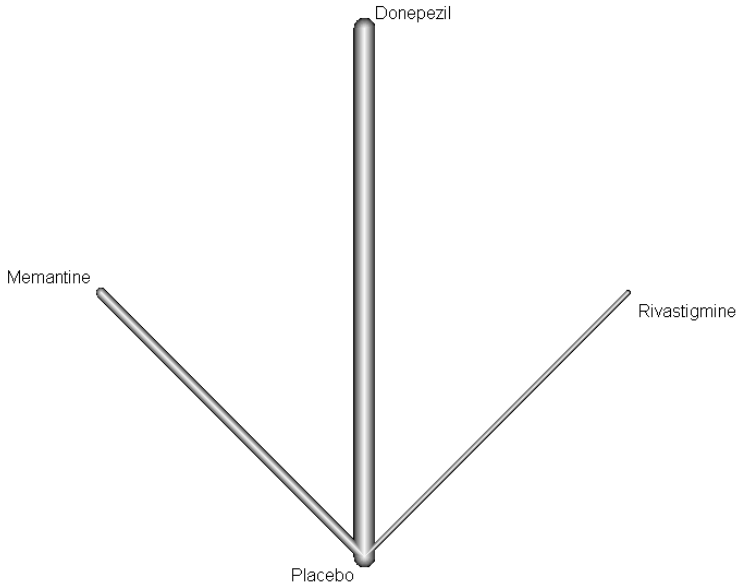
Quantifying heterogeneity/inconsistency:

tau^2 = 0.1182; I^2 = 70.7%

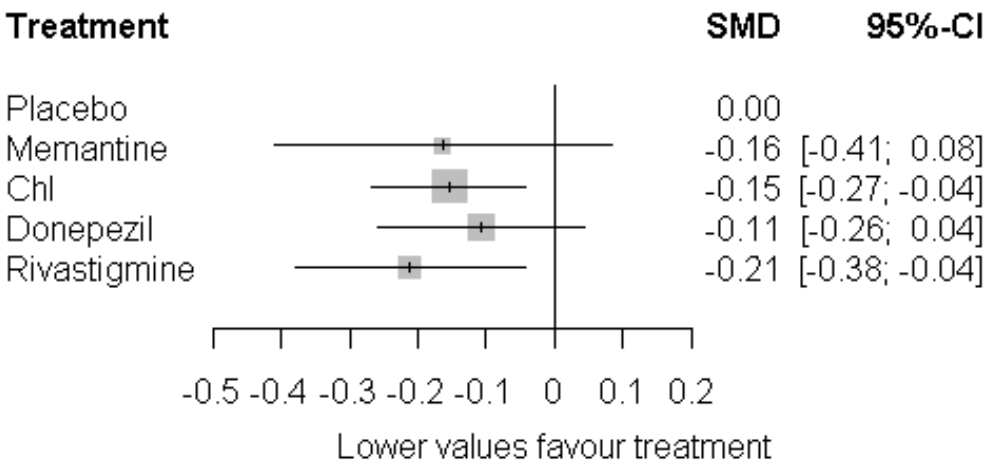
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
13.63	4	0.0086

Network graph:



PDD/DLB – NPI – FE model



Differences between treatments – standardised mean difference and 95% confidence interval

	Placebo	*Memantine*	*Chl*	*Donepezil*	*Rivastigmine*
Placebo	*N/A*				
Memantine	*-0.16 (-0.41, 0.08)*	*N/A*			
Chl	*-0.15 (-0.27, -0.04)*	*0.01 (-0.26, 0.28)*	*N/A*		
Donepezil	*-0.11 (-0.26, 0.04)*	*0.06 (-0.23, 0.35)*	*N/A*	*N/A*	
Rivastigmine	*-0.21 (-0.38, -0.04)*	*-0.05 (-0.35, 0.25)*	*N/A*	*-0.10 (-0.33, 0.12)*	*N/A*

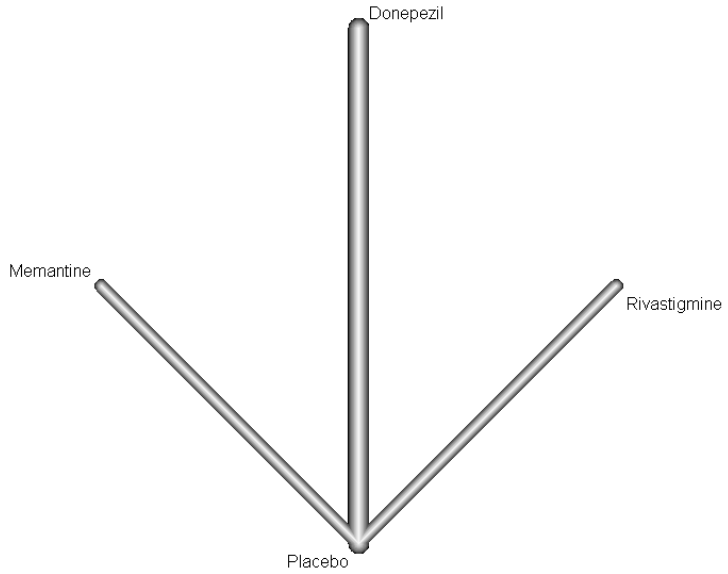
Quantifying heterogeneity/inconsistency:

tau^2 = 0.0090; I^2 = 24.7%

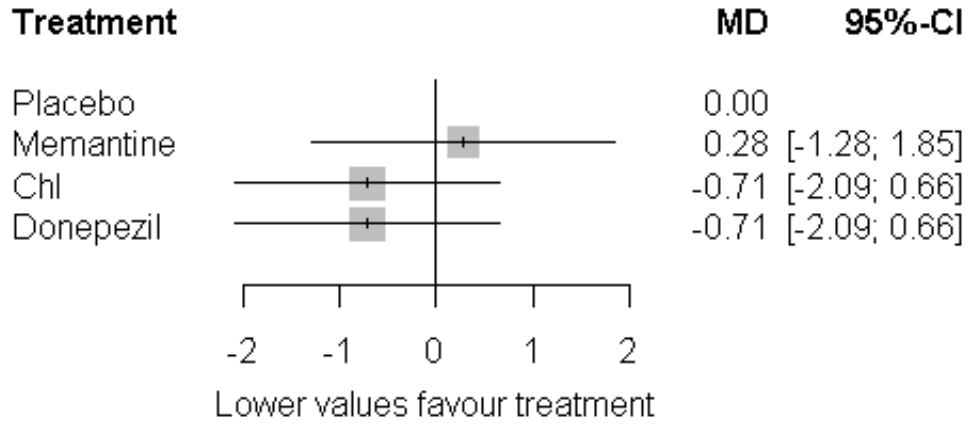
Test of heterogeneity/inconsistency:

Q	d.f.	p.value
5.31	4	0.2565

Network graph:



PDD/DLB – UPDRS III – FE model



Differences between treatments – mean difference and 95% confidence interval

	Placebo	*Memantine*	*Chl*	*Donepezil*
Placebo	*N/A*			
Memantine	*0.28 (-1.28, 1.85)*	*N/A*		
Chl	*-0.71 (-2.09, 0.66)*	*-1.00 (-3.08, 1.09)*	*N/A*	
Donepezil	*-0.71 (-2.09, 0.66)*	*-1.00 (-3.08, 1.09)*	*N/A*	*N/A*

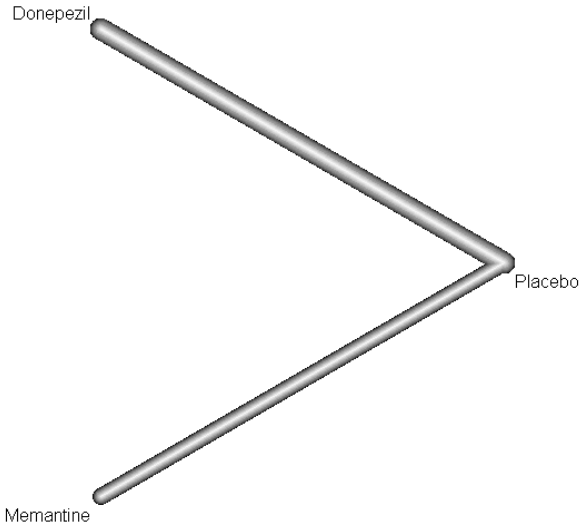
Quantifying heterogeneity/inconsistency:

tau^2 < 0.0001; I^2 = 0%

Test of heterogeneity/inconsistency:

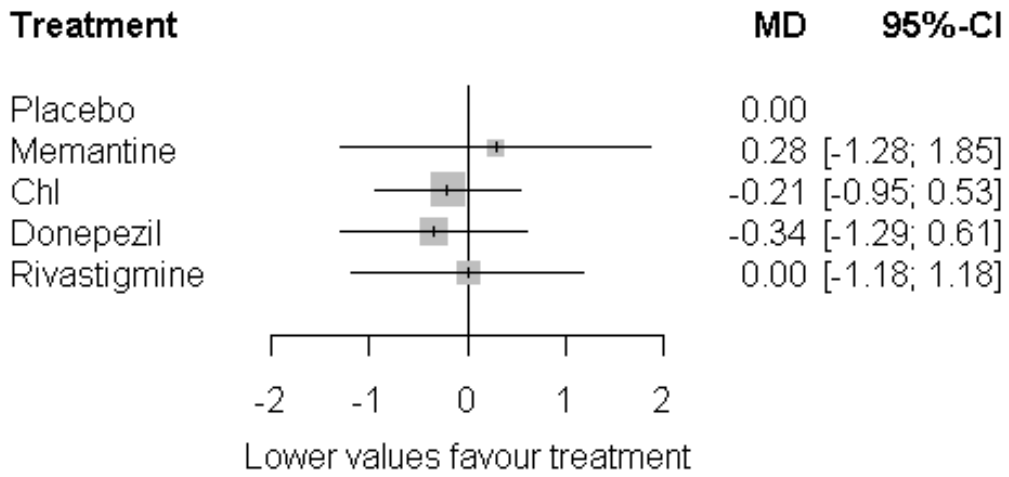
Q	d.f.	p.value
1.95	5	0.8566

Network graph:



PDD/DLB – UPDRS III sensitivity analysis – FE model

For this sensitivity analysis, in the 3 studies where the UPDRS III was measured but reported only as “non-significant”, an effect size of 0 was assumed, and a SD imputed based on the pooled SD from the other trials of cholinesterase inhibitors versus placebo.



Differences between treatments – mean difference and 95% confidence interval

	Placebo	*Memantine*	*Chl*	*Donepezil*	*Rivastigmine*
Placebo	*N/A*				
Memantine	*0.28 (-1.28, 1.85)*	*N/A*			
Chl	*-0.21 (-0.95, 0.53)*	*-0.49 (-2.22, 1.24)*	*N/A*		
Donepezil	*-0.34 (-1.29, 0.61)*	*-0.63 (-2.46, 1.21)*	*N/A*	*N/A*	
Rivastigmine	*0.00 (-1.18, 1.18)*	*-0.28 (-2.24, 1.68)*	*N/A*	*0.34 (-1.17, 1.86)*	*N/A*

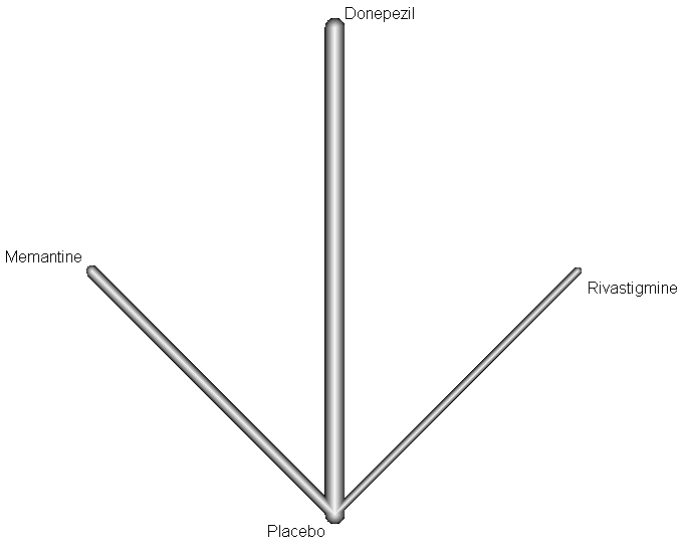
Quantifying heterogeneity/inconsistency:

tau^2 < 0.0001; I^2 = 0%

Test of heterogeneity/inconsistency:

Q	d.f.	p.value
2.48	7	0.9284

Network graph:



E.5 Non-pharmacological management of motor and non-motor symptoms

E.5.1 Physiotherapy and physical activity

Gait Outcomes

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
2 or 6 Minute Walk Test	10	Serious[1]	Not serious	Not serious	Not serious	MD 0.33 [0.11, 0.55]	Moderate
10 or 20m Walk Test	6	Serious[1]	Serious[2]	Not serious	Serious[4]	MD 0.02 [-0.63, 0.67]	Very Low
Speed	24	Serious[1]	Not serious	Not serious	Not serious	MD 0.06 [0.04, 0.08]	Moderate
Cadence (steps/min)	9	Serious[1]	Not serious	Not serious	Serious[4]	MD 0.06 [-1.67, 1.78]	Low
Stride Length (m)	10	Serious[1]	Not serious	Not serious	Not serious	MD 0.06 [0.02, 0.10]	Moderate
Step Length (m)	7	Serious[1]	Not serious	Not serious	Serious[4]	MD 0.02 [-0.00, 0.04]	Low
Freezing of Gait Questionnaire	4	Serious[1]	Not serious	Serious[3]	Not serious	MD -1.41 [-2.63, -0.19]	Low

[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2>40%); [3]Serious indirectness: The GDG did not feel that the freezing of gate questionnaire was an adequate measure to quantify the severity and frequency of freezing in people with PD; [4]Non-significant result

Functional Mobility and Balance Outcomes

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Timed Up and Go	17	Serious[1]	Serious[2]	Not serious	Serious[4]	MD -1.09 [-1.57, -0.60]	Very Low
Functional Reach (cm)	6	Serious[1]	Serious[2]	Not serious	Not serious	MD 2.82 [1.08, 4.55]	Low

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Berg Balance Scale	11	Serious[1]	Serious[2]	Not serious	Serious[5]	MD 3.28 [1.96, 4.59]	Very Low
Activity Specific Balance Confidence	3	Serious[1]	Not serious	Not serious	Serious[6]	MD 2.40 [-2.78, 7.57]	Low
Falls Efficacy Scale	8	Serious[1]	Serious[2]	Serious[7]	Serious[6]	MD -3.59 [-7.55, 0.38]	Very Low
Number of people falling	2	Serious[1]	Serious[2]	Not serious	Serious[6]	OR 0.53 [0.20, 1.43]	Very Low

[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2>40%); [3]Serious indirectness: The GDG did not feel that the freezing of gate questionnaire was an adequate measure to quantify the severity and frequency of freezing in people with PD; [4]Serious imprecision: MCIC = 11s was deemed clinically meaningful by the GDG; [5]Serious imprecision: MCIC = 5 points was deemed clinically meaningful by the GDG; [6]Non-significant results; [7]Serious indirection: The GDG did not feel that the falls efficacy scale was an adequate measure to quantify the severity and frequency of falls in people with PD

Clinical-Rated Disability

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
UPDRS Total	7	Serious[1]	Serious[2]	Not serious	Serious[3]	MD -5.32 [-8.34, -2.30]	Very low
UPDRS Mental	4	Serious[1]	Not serious	Not serious	Not serious	MD -0.43 [-0.82, -0.05]	Moderate
UPDRS II (ADL)	7	Serious[1]	Not serious	Not serious	Not serious[4]	MD -1.63 [-2.42, -0.84]	Moderate
UPDRS III (motor)	23	Serious[1]	Serious[2]	Not serious	Serious[5]	MD -4.24 [-5.90, -2.58]	Very low

[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2>40%); [3]CI cross the MID of 7.3 points (Schrag et al., 2006); [4]CI do not cross the MID of 3 points (Schrag et al., 2006); [5]CI cross the MID of 3.25 (Horvath et al., 2015) and 5 points (Schrag et al, 2006)

Clinical-rated QoL

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
PDQ-39 Summary Index	14	Serious[1]	Serious[2]	Not serious	Serious[4]	MD -4.74 [-8.08, -1.39]	Very low
PDQ-39 Mobility	4	Serious[1]	Not serious	Not serious	Serious[3]	MD -2.31 [-6.55, 1.92]	Low
[1]Individual study(ies) at risk of bias; [2]Considerable between study heterogeneity (i^2>40%); [3]Non-significant result; [4]CI cross the MID of 1.6 points (Peto et al., 2001)							

PD REHAB (Clarke et al., 2016)

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
NEADL Summary Index (at 3 months)	1	Not serious	N/A	Serious[1]	Serious[2]	MD 0.5 [-0.7, 1.7]	Low
NEADL Summary Index (at 15 months)	1	Not serious	N/A	Serious[1]	Serious[2]	MD 0.07 [-0.64, 0.77]	Low
PDQ-39 Summary Index (at 3 months)	1	Not serious	N/A	Serious[1]	Not serious[3]	MD 0.007 [-1.5, 1.5]	Moderate
PDQ-39 Summary Index (at 15 months)	1	Not serious	N/A	Serious[1]	Serious[4]	MD -1.55 [-2.62, -0.47]	Low
EQ-5D quotient (at 3 months)	1	Not serious	N/A	Serious[1]	Serious[2]	MD -0.03 [-0.07, -0.002]	Low
EQ-5D quotient (at 15 months)	1	Not serious	N/A	Serious[1]	Not serious	MD 0.02 [0.00007, 0.03]	Moderate
SF-12 physical (carers – at 3 months)	1	Not serious	N/A	Serious[1]	Serious[2]	MD -0.6 [-2.3, 1.2]	Low
SF-12 mental (carers – at 3 months)	1	Not serious	N/A	Serious[1]	Not serious	MD -2.1 [-3.9, -0.3]	Moderate

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
[1]Considered serious as intervention is not as defined in protocol [2]Non-significant result [3]CI does not cross the MID of 1.6 points (Peto et al., 2001) [4]CI cross the MID of 1.6 points (Peto et al., 2001)							

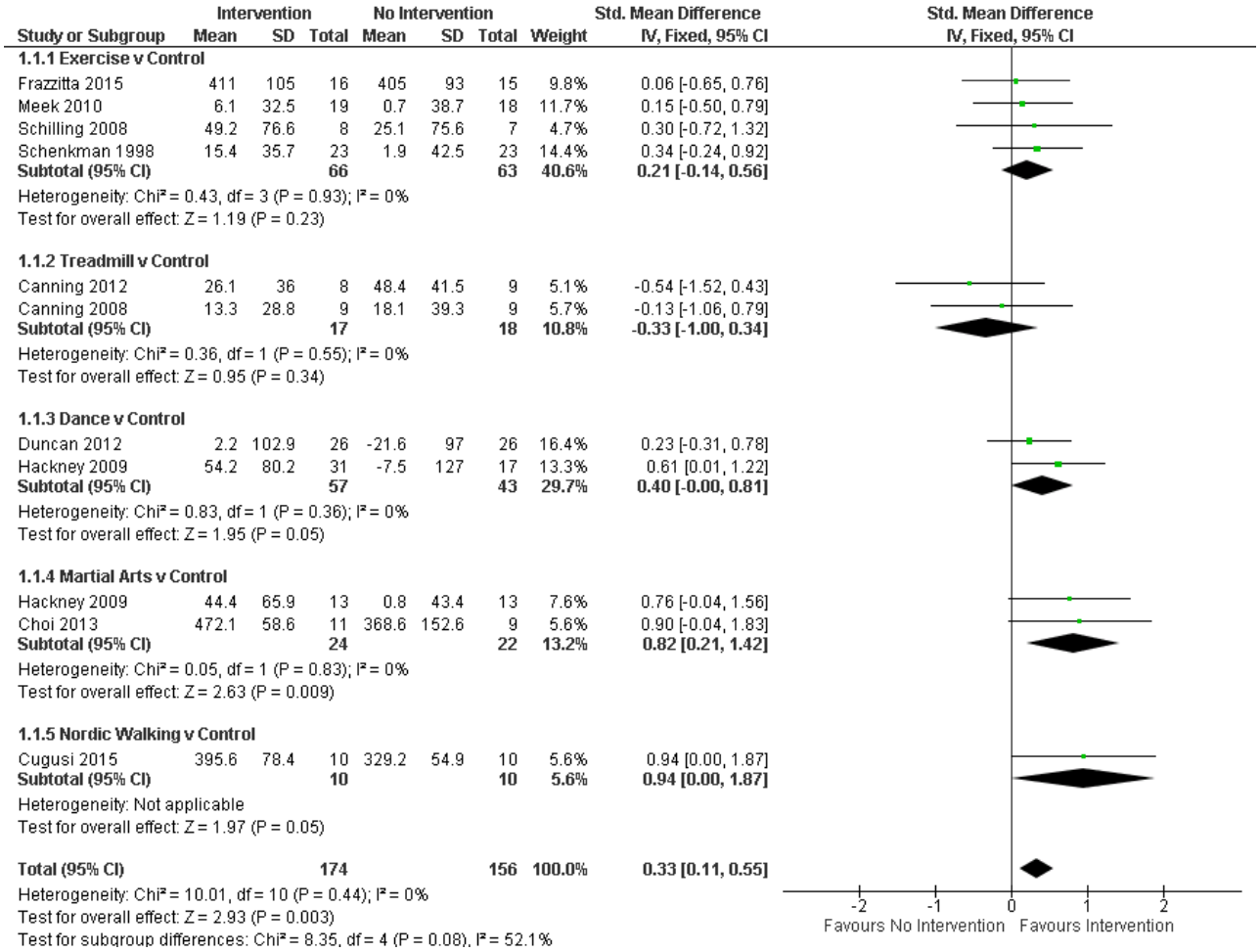
Alexander Technique

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Self-assessment PD disability scale (SPDDS) at best	1	Not serious	N/A	Not serious	Not serious	MD -3.5 (-7.7 to -0.0)	High
Self-assessment PD disability scale (SPDDS) at worst	1	Not serious	N/A	Not serious	Not serious	MD -6.3 (-11.8 to -0.9)	High
Beck Depression Inventory	1	Not serious	N/A	Not serious	Serious[1]	MD -0.9 (-2.6 to 0.9)	Moderate
[1]Non-significant results							

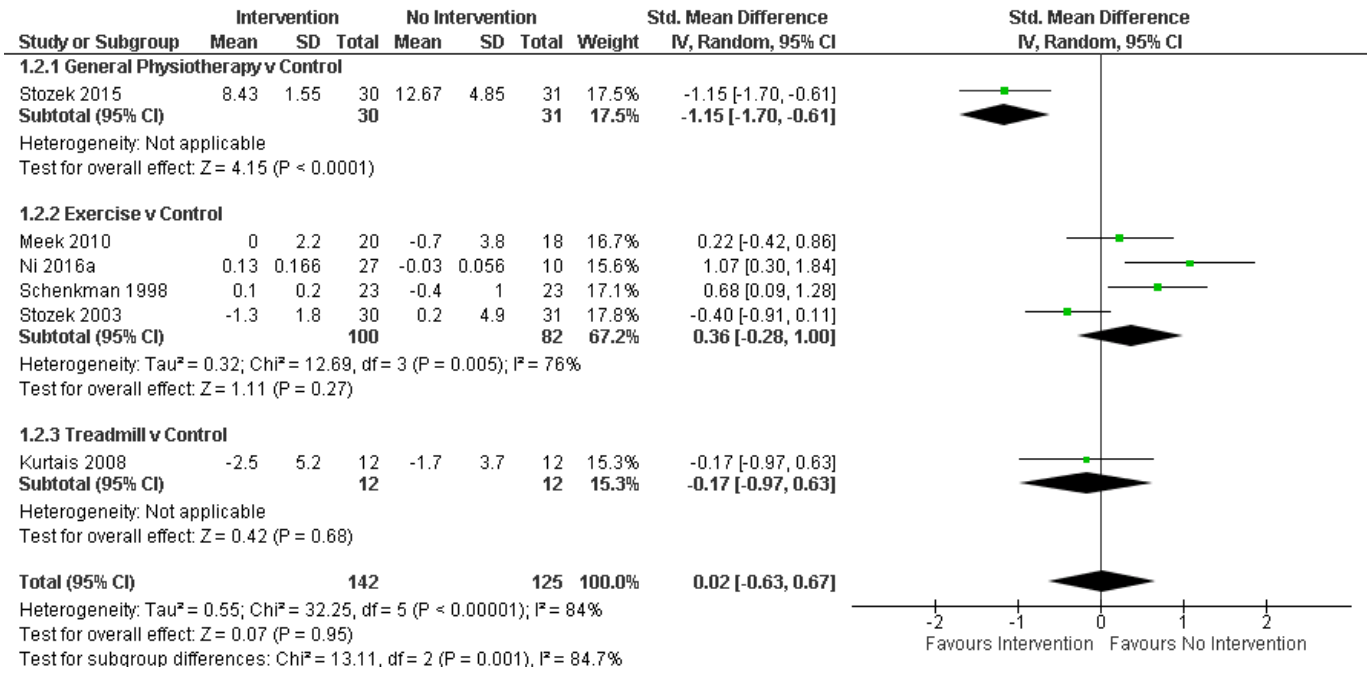
Forest plots

Gait Outcomes

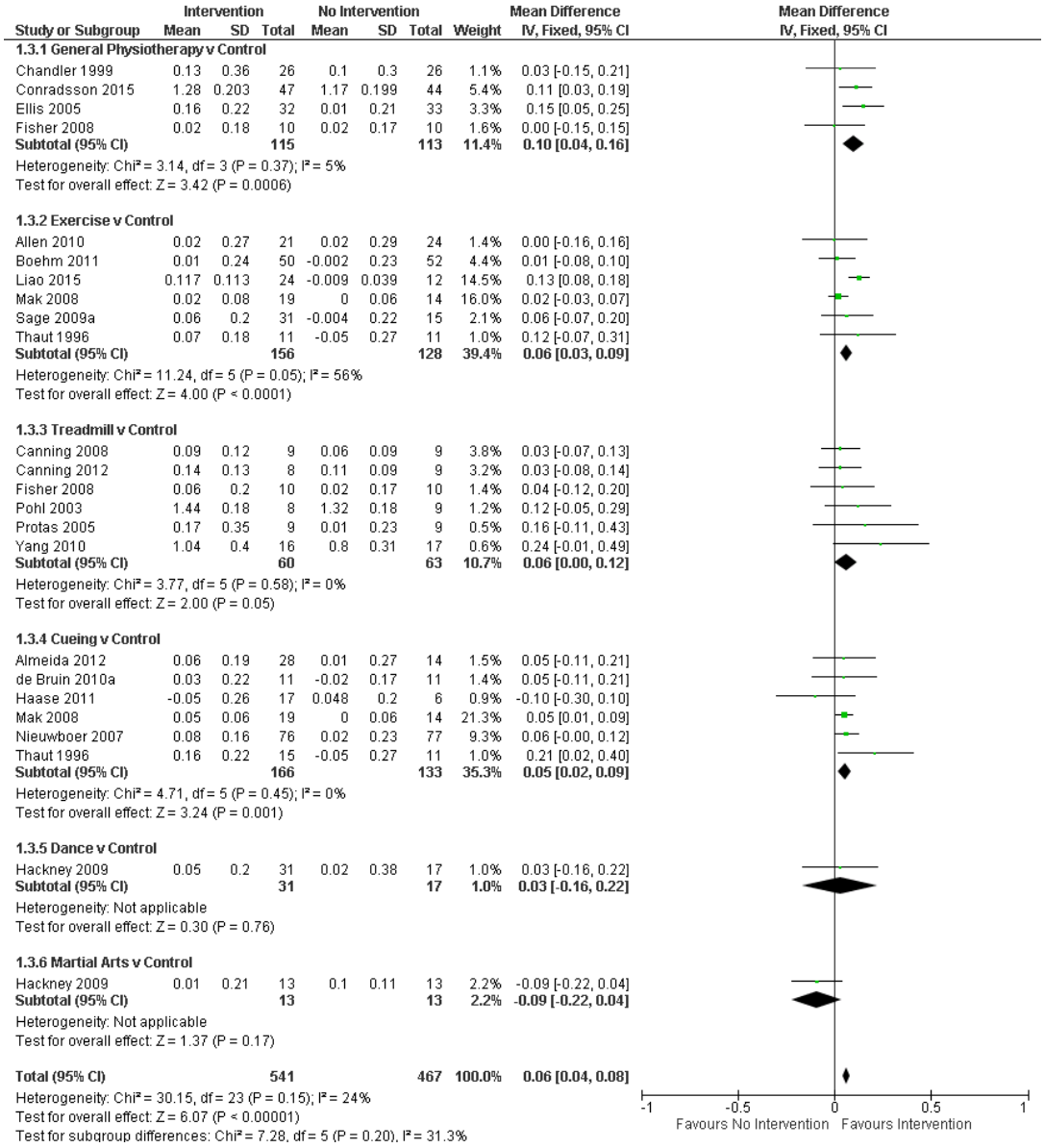
2 or 6 Minute Walk Test



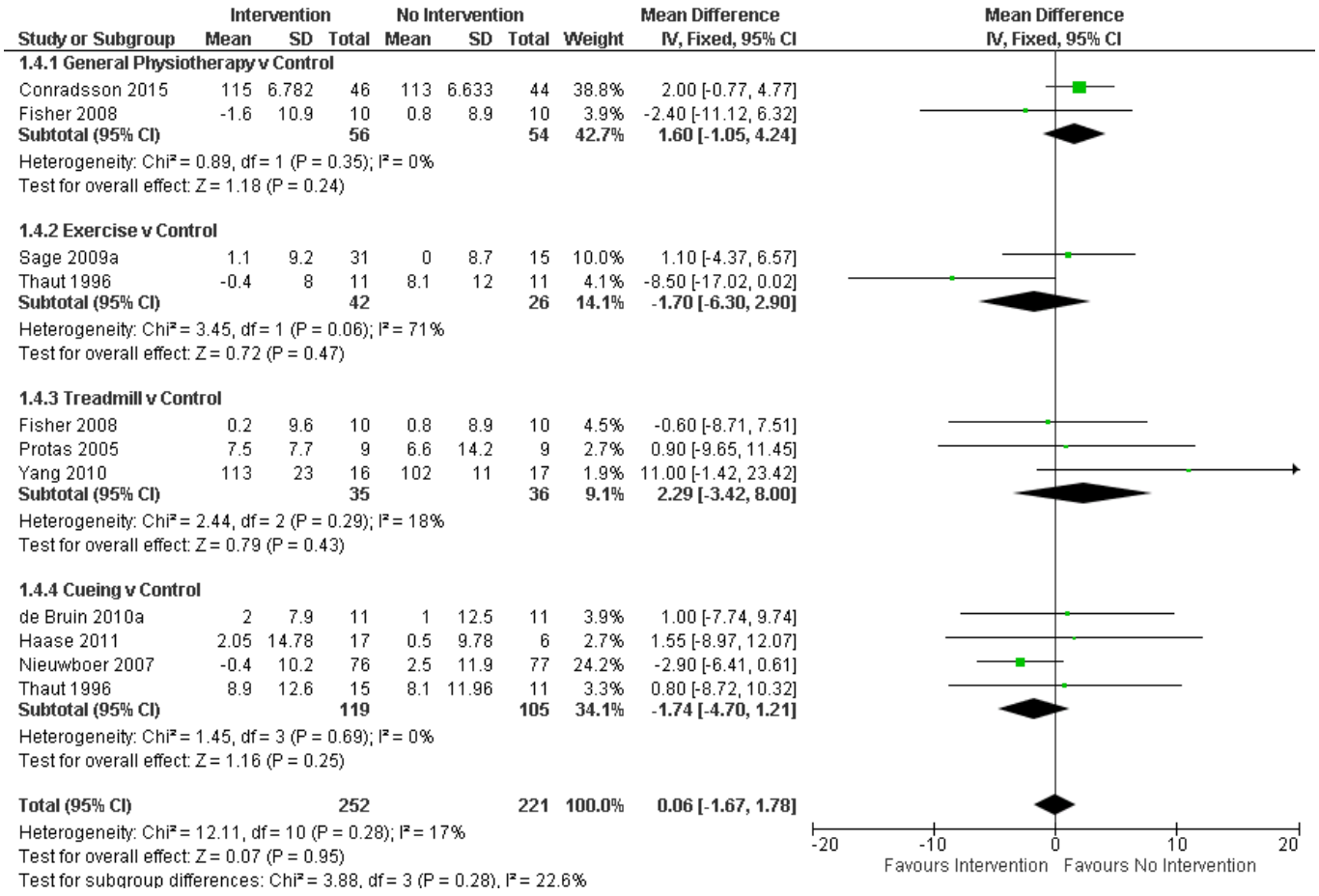
10 or 20m Walk test



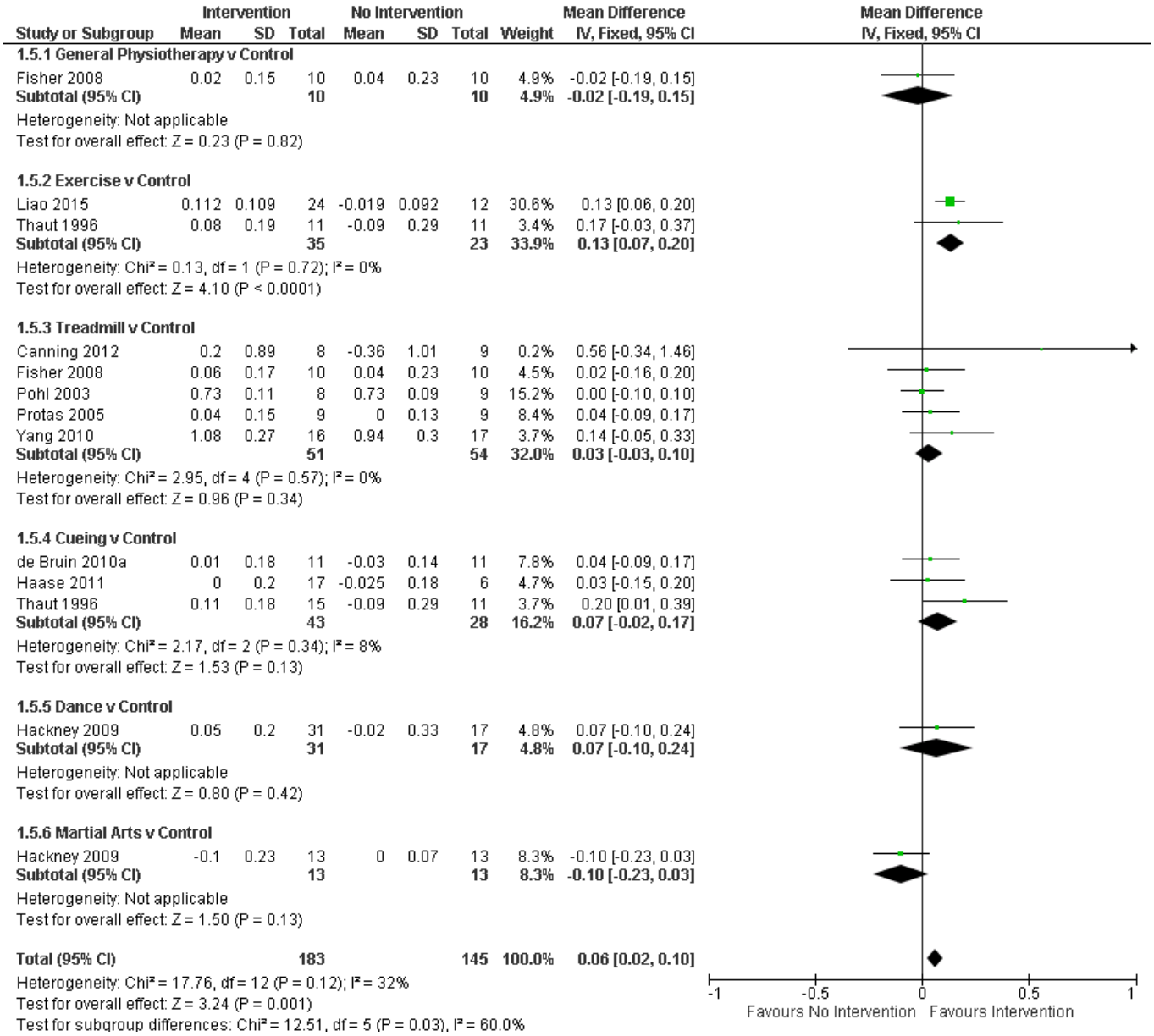
Speed


Intervention
No Intervention
Mean Difference
Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Fixed, 95% CI
IV, Fixed, 95% CI
1.3.1 General Physiotherapy v Control
Chandler 1999 0.13 0.36 26 0.1 0.3 26 1.1% 0.03 [-0.15, 0.21]
Conradsson 2015 1.28 0.203 47 1.17 0.199 44 5.4% 0.11 [0.03, 0.19]
Ellis 2005 0.16 0.22 32 0.01 0.21 33 3.3% 0.15 [0.05, 0.25]
Fisher 2008 0.02 0.18 10 0.02 0.17 10 1.6% 0.00 [-0.15, 0.15]
Subtotal (95% CI) 115 113 11.4% 0.10 [0.04, 0.16]
Heterogeneity: Chi² = 3.14, df = 3 (P = 0.37); I² = 5%
Test for overall effect: Z = 3.42 (P = 0.0006)
1.3.2 Exercise v Control
Allen 2010 0.02 0.27 21 0.02 0.29 24 1.4% 0.00 [-0.16, 0.16]
Boehm 2011 0.01 0.24 50 -0.002 0.23 52 4.4% 0.01 [-0.08, 0.10]
Liao 2015 0.117 0.113 24 -0.009 0.039 12 14.5% 0.13 [0.08, 0.18]
Mak 2008 0.02 0.08 19 0 0.06 14 16.0% 0.02 [-0.03, 0.07]
Sage 2009a 0.06 0.2 31 -0.004 0.22 15 2.1% 0.06 [-0.07, 0.20]
Thaut 1996 0.07 0.18 11 -0.05 0.27 11 1.0% 0.12 [-0.07, 0.31]
Subtotal (95% CI) 156 128 39.4% 0.06 [0.03, 0.09]
Heterogeneity: Chi² = 11.24, df = 5 (P = 0.05); I² = 56%
Test for overall effect: Z = 4.00 (P < 0.0001)
1.3.3 Treadmill v Control
Canning 2008 0.09 0.12 9 0.06 0.09 9 3.8% 0.03 [-0.07, 0.13]
Canning 2012 0.14 0.13 8 0.11 0.09 9 3.2% 0.03 [-0.08, 0.14]
Fisher 2008 0.06 0.2 10 0.02 0.17 10 1.4% 0.04 [-0.12, 0.20]
Pohl 2003 1.44 0.18 8 1.32 0.18 9 1.2% 0.12 [-0.05, 0.29]
Protas 2005 0.17 0.35 9 0.01 0.23 9 0.5% 0.16 [-0.11, 0.43]
Yang 2010 1.04 0.4 16 0.8 0.31 17 0.6% 0.24 [-0.01, 0.49]
Subtotal (95% CI) 60 63 10.7% 0.06 [0.00, 0.12]
Heterogeneity: Chi² = 3.77, df = 5 (P = 0.58); I² = 0%
Test for overall effect: Z = 2.00 (P = 0.05)
1.3.4 Cueing v Control
Almeida 2012 0.06 0.19 28 0.01 0.27 14 1.5% 0.05 [-0.11, 0.21]
de Bruin 2010a 0.03 0.22 11 -0.02 0.17 11 1.4% 0.05 [-0.11, 0.21]
Haase 2011 -0.05 0.26 17 0.048 0.2 6 0.9% -0.10 [-0.30, 0.10]
Mak 2008 0.05 0.06 19 0 0.06 14 21.3% 0.05 [0.01, 0.09]
Nieuwboer 2007 0.08 0.16 76 0.02 0.23 77 9.3% 0.06 [-0.00, 0.12]
Thaut 1996 0.16 0.22 15 -0.05 0.27 11 1.0% 0.21 [0.02, 0.40]
Subtotal (95% CI) 166 133 35.3% 0.05 [0.02, 0.09]
Heterogeneity: Chi² = 4.71, df = 5 (P = 0.45); I² = 0%
Test for overall effect: Z = 3.24 (P = 0.001)
1.3.5 Dance v Control
Hackney 2009 0.05 0.2 31 0.02 0.38 17 1.0% 0.03 [-0.16, 0.22]
Subtotal (95% CI) 31 17 1.0% 0.03 [-0.16, 0.22]
Heterogeneity: Not applicable
Test for overall effect: Z = 0.30 (P = 0.76)
1.3.6 Martial Arts v Control
Hackney 2009 0.01 0.21 13 0.1 0.11 13 2.2% -0.09 [-0.22, 0.04]
Subtotal (95% CI) 13 13 2.2% -0.09 [-0.22, 0.04]
Heterogeneity: Not applicable
Test for overall effect: Z = 1.37 (P = 0.17)
Total (95% CI) 541 467 100.0% 0.06 [0.04, 0.08]
Heterogeneity: Chi² = 30.15, df = 23 (P = 0.15); I² = 24%
Test for overall effect: Z = 6.07 (P < 0.00001)
Test for subgroup differences: Chi² = 7.28, df = 5 (P = 0.20), I² = 31.3%
-1
-0.5
0
0.5
1
Favours No Intervention
Favours Intervention

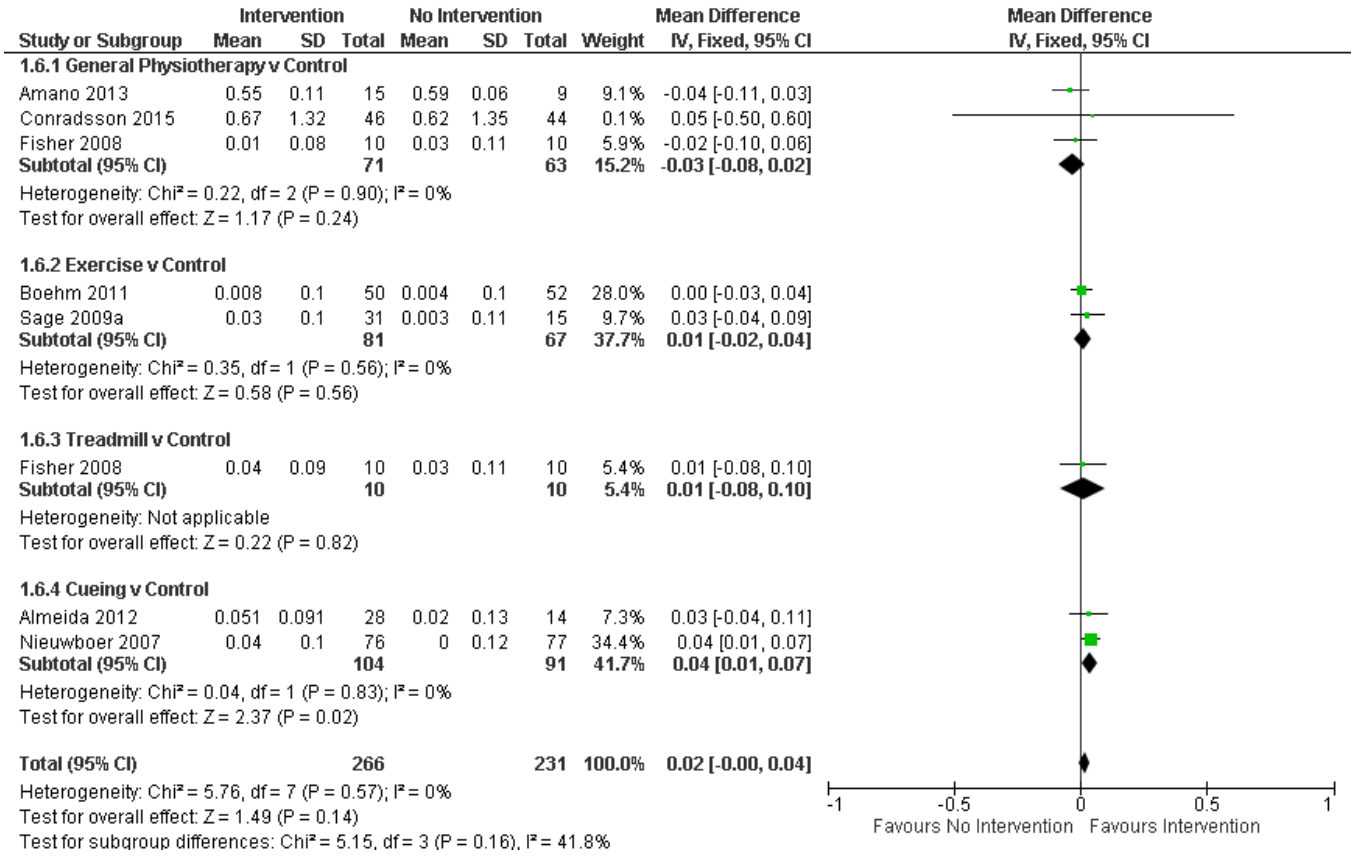
Cadence (steps/min)



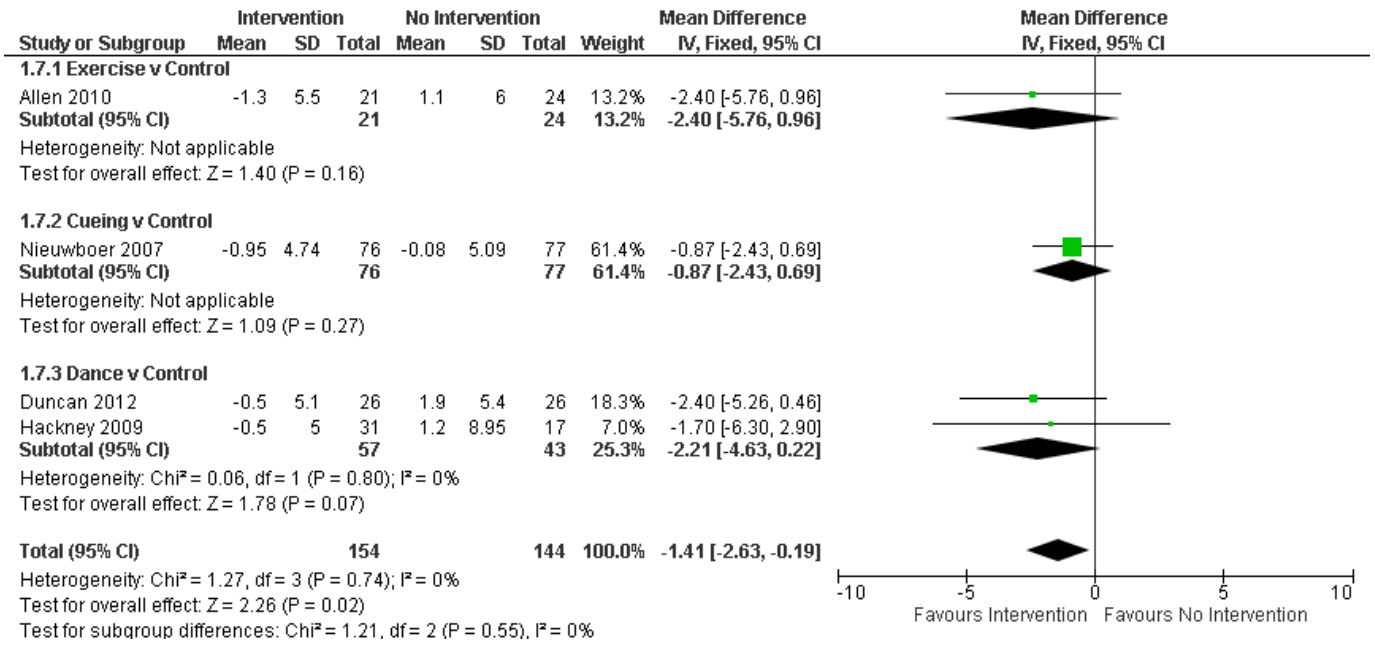
Stride Length (m)


Intervention
No Intervention
Mean Difference
Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Fixed, 95% CI
IV, Fixed, 95% CI
1.5.1 General Physiotherapy v Control
Fisher 2008 0.02 0.15 10 0.04 0.23 10 4.9% -0.02 [-0.19, 0.15]
Subtotal (95% CI) 10 10 4.9% -0.02 [-0.19, 0.15]
Heterogeneity: Not applicable
Test for overall effect: Z = 0.23 (P = 0.82)
1.5.2 Exercise v Control
Liao 2015 0.112 0.109 24 -0.019 0.092 12 30.6% 0.13 [0.06, 0.20]
Thaut 1996 0.08 0.19 11 -0.09 0.29 11 3.4% 0.17 [-0.03, 0.37]
Subtotal (95% CI) 35 23 33.9% 0.13 [0.07, 0.20]
Heterogeneity: Chi² = 0.13, df = 1 (P = 0.72); I² = 0%
Test for overall effect: Z = 4.10 (P < 0.0001)
1.5.3 Treadmill v Control
Canning 2012 0.2 0.89 8 -0.36 1.01 9 0.2% 0.56 [-0.34, 1.46]
Fisher 2008 0.06 0.17 10 0.04 0.23 10 4.5% 0.02 [-0.16, 0.20]
Pohl 2003 0.73 0.11 8 0.73 0.09 9 15.2% 0.00 [-0.10, 0.10]
Protas 2005 0.04 0.15 9 0 0.13 9 8.4% 0.04 [-0.09, 0.17]
Yang 2010 1.08 0.27 16 0.94 0.3 17 3.7% 0.14 [-0.05, 0.33]
Subtotal (95% CI) 51 54 32.0% 0.03 [-0.03, 0.10]
Heterogeneity: Chi² = 2.95, df = 4 (P = 0.57); I² = 0%
Test for overall effect: Z = 0.96 (P = 0.34)
1.5.4 Cueing v Control
de Bruin 2010a 0.01 0.18 11 -0.03 0.14 11 7.8% 0.04 [-0.09, 0.17]
Haase 2011 0 0.2 17 -0.025 0.18 6 4.7% 0.03 [-0.15, 0.20]
Thaut 1996 0.11 0.18 15 -0.09 0.29 11 3.7% 0.20 [0.01, 0.39]
Subtotal (95% CI) 43 28 16.2% 0.07 [-0.02, 0.17]
Heterogeneity: Chi² = 2.17, df = 2 (P = 0.34); I² = 8%
Test for overall effect: Z = 1.53 (P = 0.13)
1.5.5 Dance v Control
Hackney 2009 0.05 0.2 31 -0.02 0.33 17 4.8% 0.07 [-0.10, 0.24]
Subtotal (95% CI) 31 17 4.8% 0.07 [-0.10, 0.24]
Heterogeneity: Not applicable
Test for overall effect: Z = 0.80 (P = 0.42)
1.5.6 Martial Arts v Control
Hackney 2009 -0.1 0.23 13 0 0.07 13 8.3% -0.10 [-0.23, 0.03]
Subtotal (95% CI) 13 13 8.3% -0.10 [-0.23, 0.03]
Heterogeneity: Not applicable
Test for overall effect: Z = 1.50 (P = 0.13)
Total (95% CI) 183 145 100.0% 0.06 [0.02, 0.10]
Heterogeneity: Chi² = 17.76, df = 12 (P = 0.12); I² = 32%
Test for overall effect: Z = 3.24 (P = 0.001)
Test for subgroup differences: Chi² = 12.51, df = 5 (P = 0.03), I² = 60.0%
-1
-0.5
0
0.5
1
Favours No Intervention
Favours Intervention

Step Length (m)

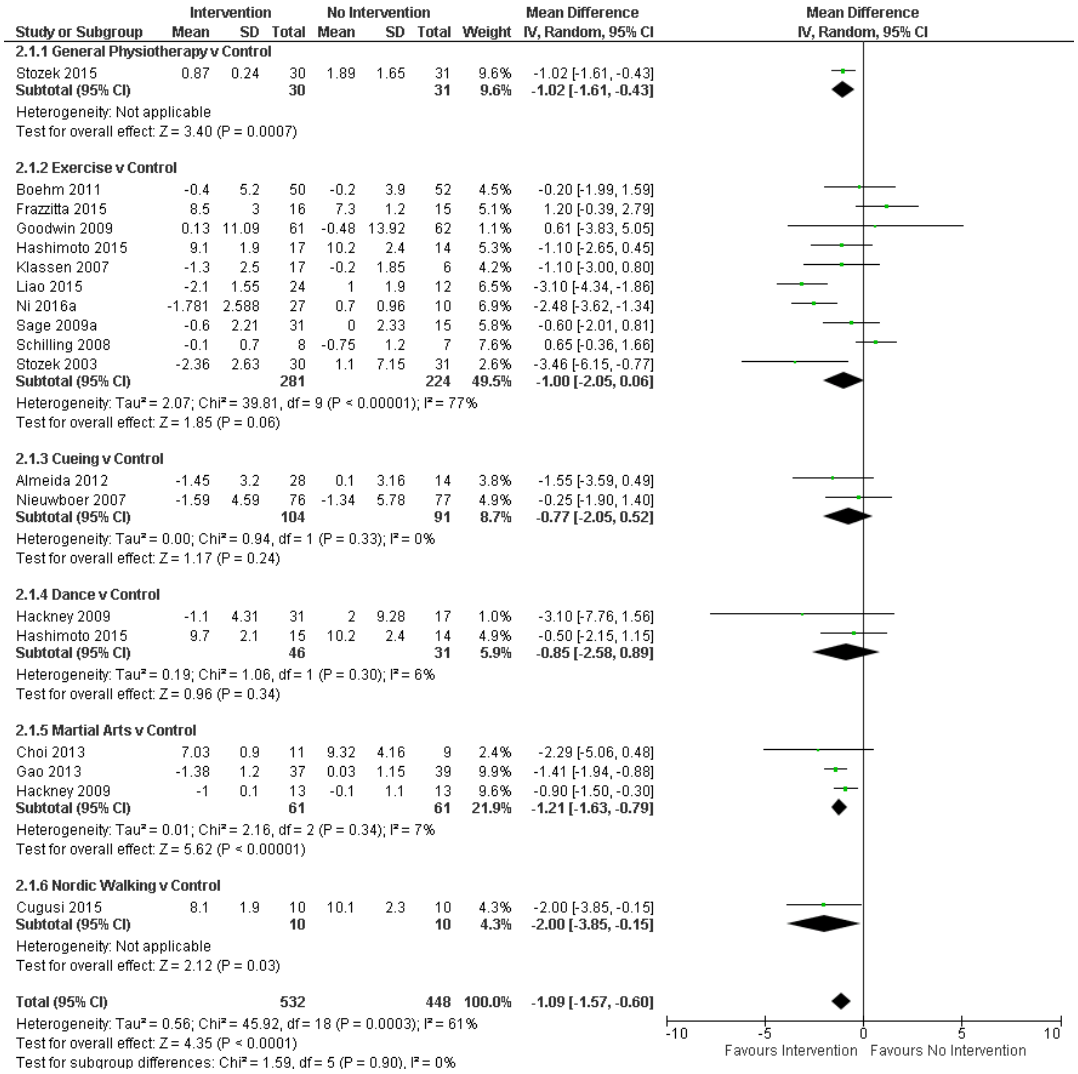


Freezing of Gait Questionnaire

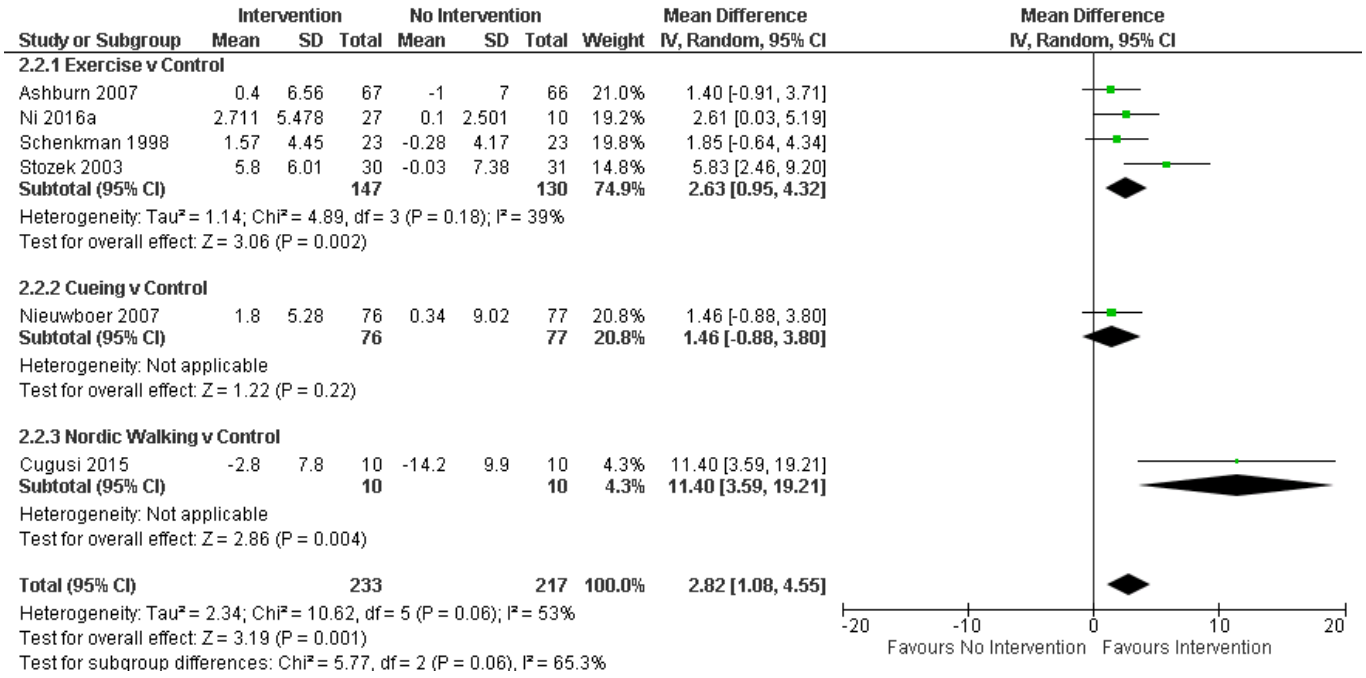


Functional Mobility and Balance Outcomes

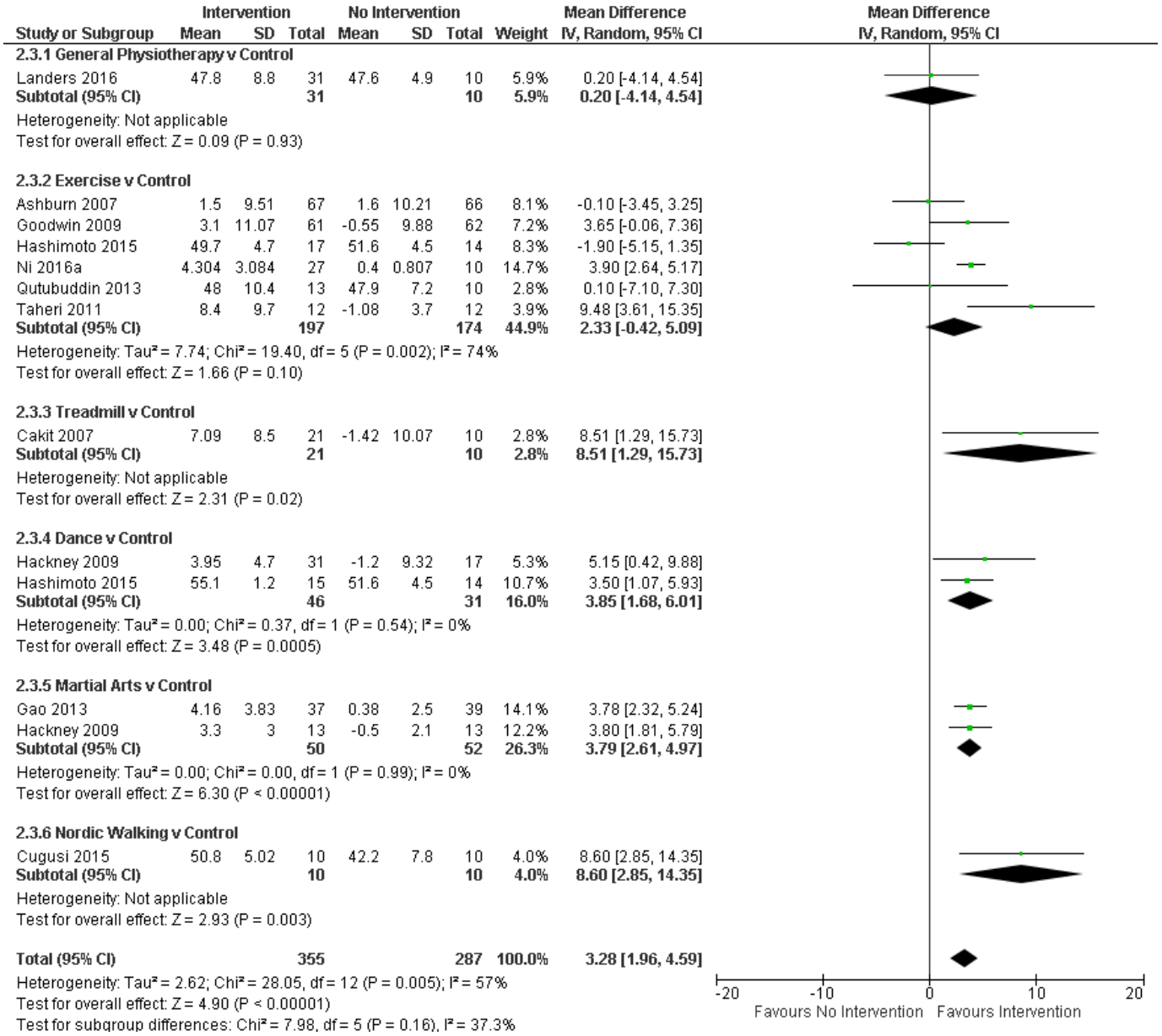
Timed Up and Go



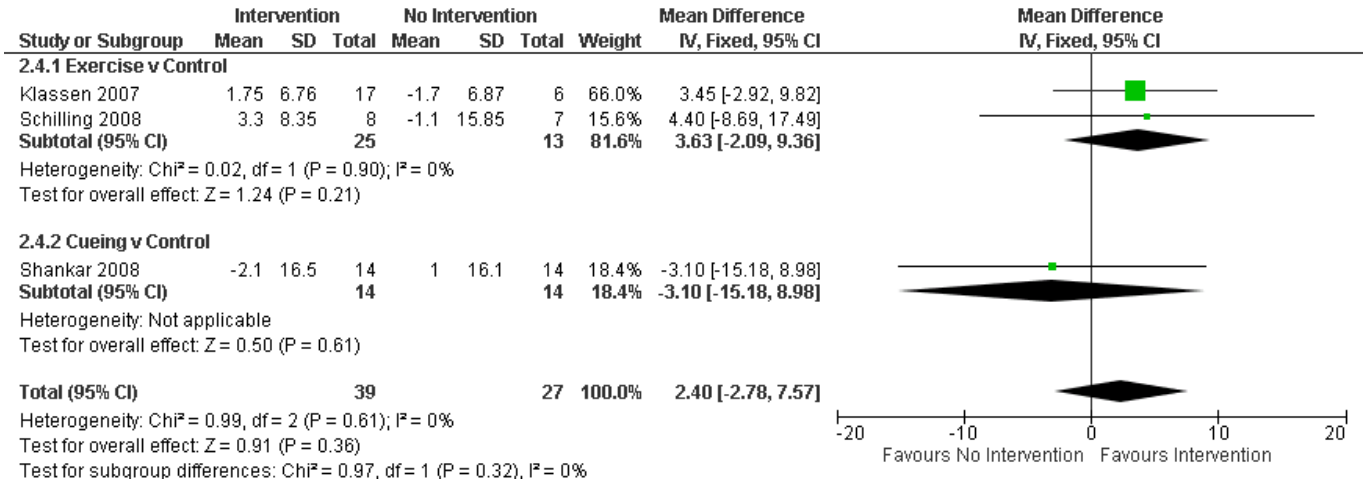
Functional Reach (cm)



Berg Balance Scale

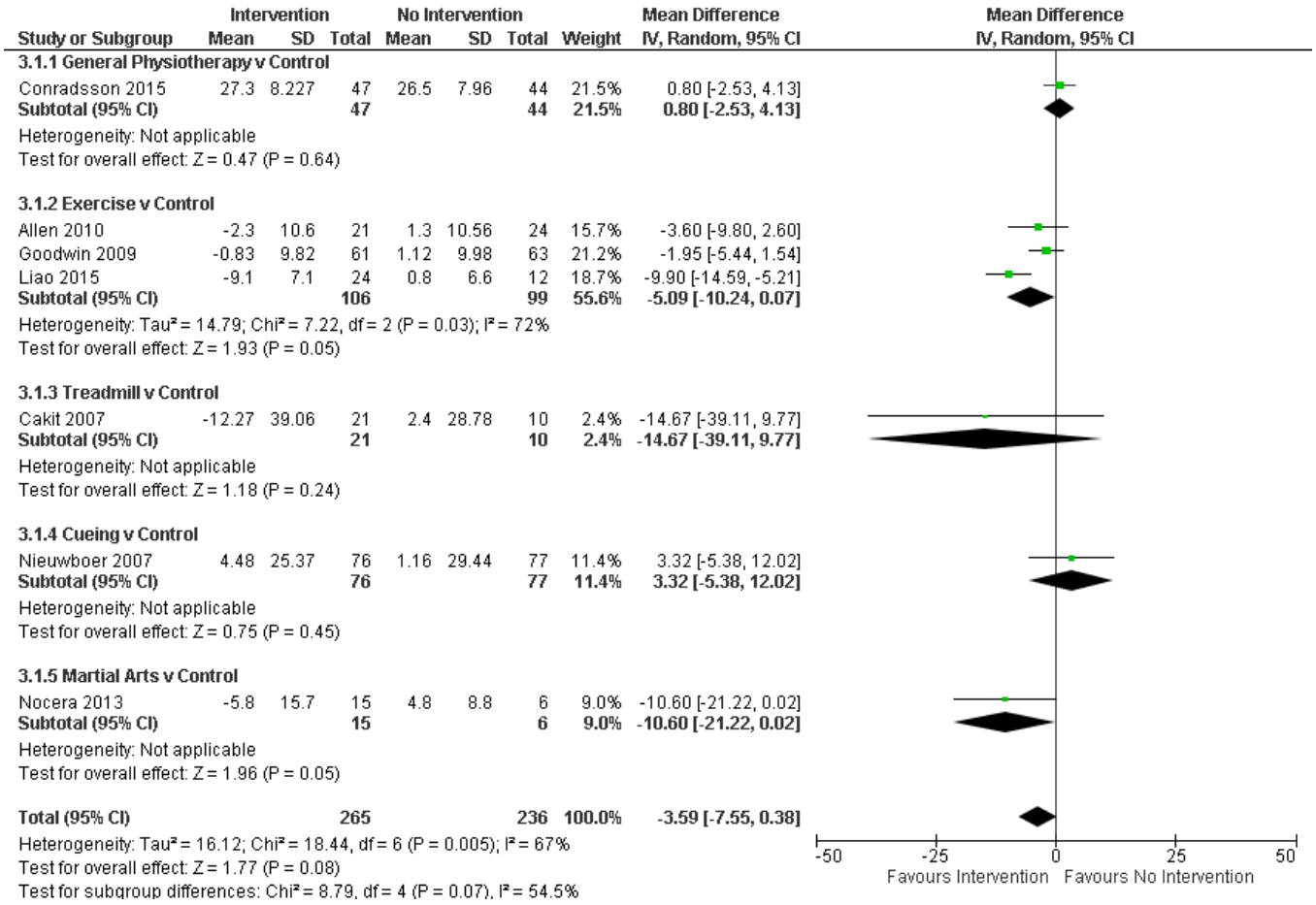

Intervention
No Intervention
Mean Difference
Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Random, 95% CI
IV, Random, 95% CI
2.3.1 General Physiotherapy v Control
Landers 2016 47.8 8.8 31 47.6 4.9 10 5.9% 0.20 [-4.14, 4.54]
Subtotal (95% CI) 31 10 5.9% 0.20 [-4.14, 4.54]
Heterogeneity: Not applicable
Test for overall effect: Z = 0.09 (P = 0.93)
2.3.2 Exercise v Control
Ashburn 2007 1.5 9.51 67 1.6 10.21 66 8.1% -0.10 [-3.45, 3.25]
Goodwin 2009 3.1 11.07 61 -0.55 9.88 62 7.2% 3.65 [-0.06, 7.36]
Hashimoto 2015 49.7 4.7 17 51.6 4.5 14 8.3% -1.90 [-5.15, 1.35]
Ni 2016a 4.304 3.084 27 0.4 0.807 10 14.7% 3.90 [2.64, 5.17]
Qutubuddin 2013 48 10.4 13 47.9 7.2 10 2.8% 0.10 [-7.10, 7.30]
Taheri 2011 8.4 9.7 12 -1.08 3.7 12 3.9% 9.48 [3.61, 15.35]
Subtotal (95% CI) 197 174 44.9% 2.33 [-0.42, 5.09]
Heterogeneity: Tau² = 7.74; Chi² = 19.40, df = 5 (P = 0.002); I² = 74%
Test for overall effect: Z = 1.66 (P = 0.10)
2.3.3 Treadmill v Control
Cakit 2007 7.09 8.5 21 -1.42 10.07 10 2.8% 8.51 [1.29, 15.73]
Subtotal (95% CI) 21 10 2.8% 8.51 [1.29, 15.73]
Heterogeneity: Not applicable
Test for overall effect: Z = 2.31 (P = 0.02)
2.3.4 Dance v Control
Hackney 2009 3.95 4.7 31 -1.2 9.32 17 5.3% 5.15 [0.42, 9.88]
Hashimoto 2015 55.1 1.2 15 51.6 4.5 14 10.7% 3.50 [1.07, 5.93]
Subtotal (95% CI) 46 31 16.0% 3.85 [1.68, 6.01]
Heterogeneity: Tau² = 0.00; Chi² = 0.37, df = 1 (P = 0.54); I² = 0%
Test for overall effect: Z = 3.48 (P = 0.0005)
2.3.5 Martial Arts v Control
Gao 2013 4.16 3.83 37 0.38 2.5 39 14.1% 3.78 [2.32, 5.24]
Hackney 2009 3.3 3 13 -0.5 2.1 13 12.2% 3.80 [1.81, 5.79]
Subtotal (95% CI) 50 52 26.3% 3.79 [2.61, 4.97]
Heterogeneity: Tau² = 0.00; Chi² = 0.00, df = 1 (P = 0.99); I² = 0%
Test for overall effect: Z = 6.30 (P < 0.00001)
2.3.6 Nordic Walking v Control
Cugusi 2015 50.8 5.02 10 42.2 7.8 10 4.0% 8.60 [2.85, 14.35]
Subtotal (95% CI) 10 10 4.0% 8.60 [2.85, 14.35]
Heterogeneity: Not applicable
Test for overall effect: Z = 2.93 (P = 0.003)
Total (95% CI) 355 287 100.0% 3.28 [1.96, 4.59]
Heterogeneity: Tau² = 2.62; Chi² = 28.05, df = 12 (P = 0.005); I² = 57%
Test for overall effect: Z = 4.90 (P < 0.00001)
Test for subgroup differences: Chi² = 7.98, df = 5 (P = 0.16), I² = 37.3%
-20
-10
0
10
20
Favours No Intervention
Favours Intervention

Activity Specific Balance Confidence

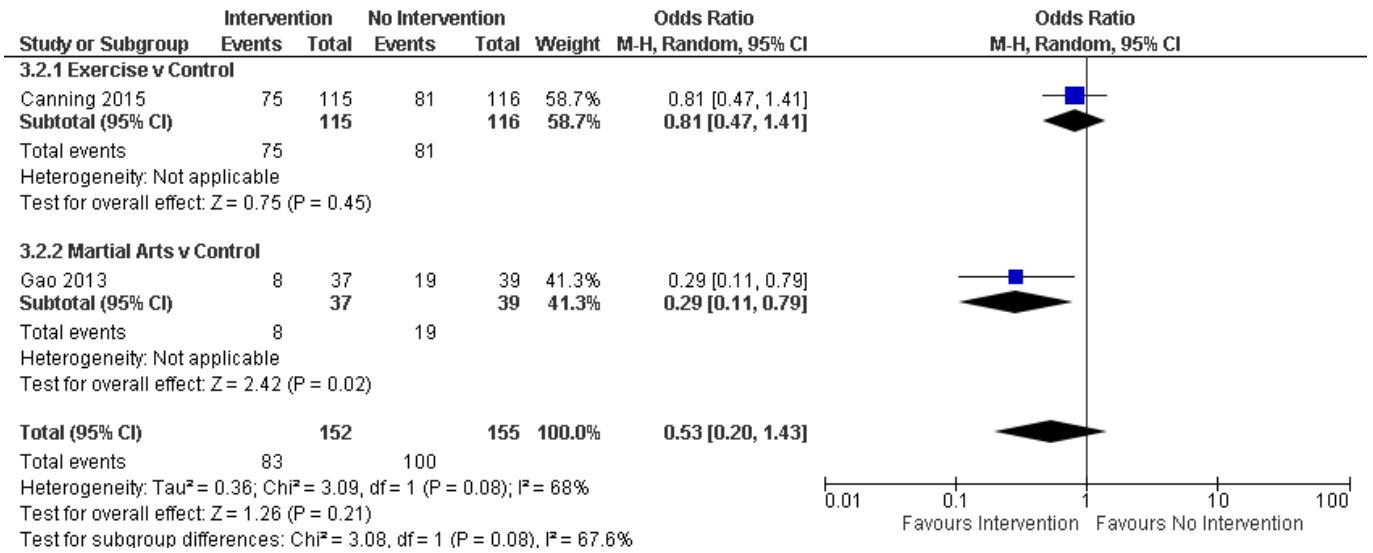


Falls

Falls Efficacy Scale

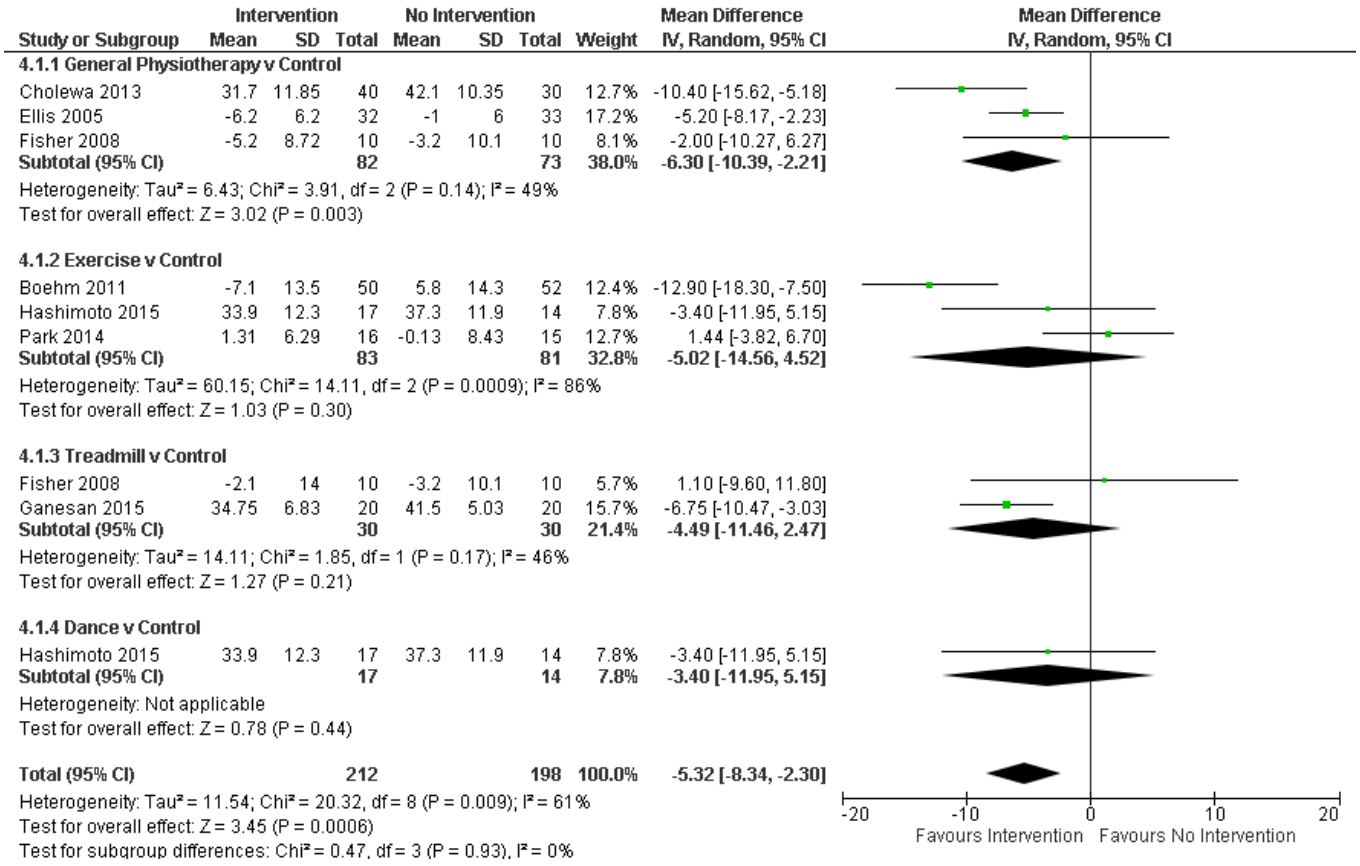


Number of people falling

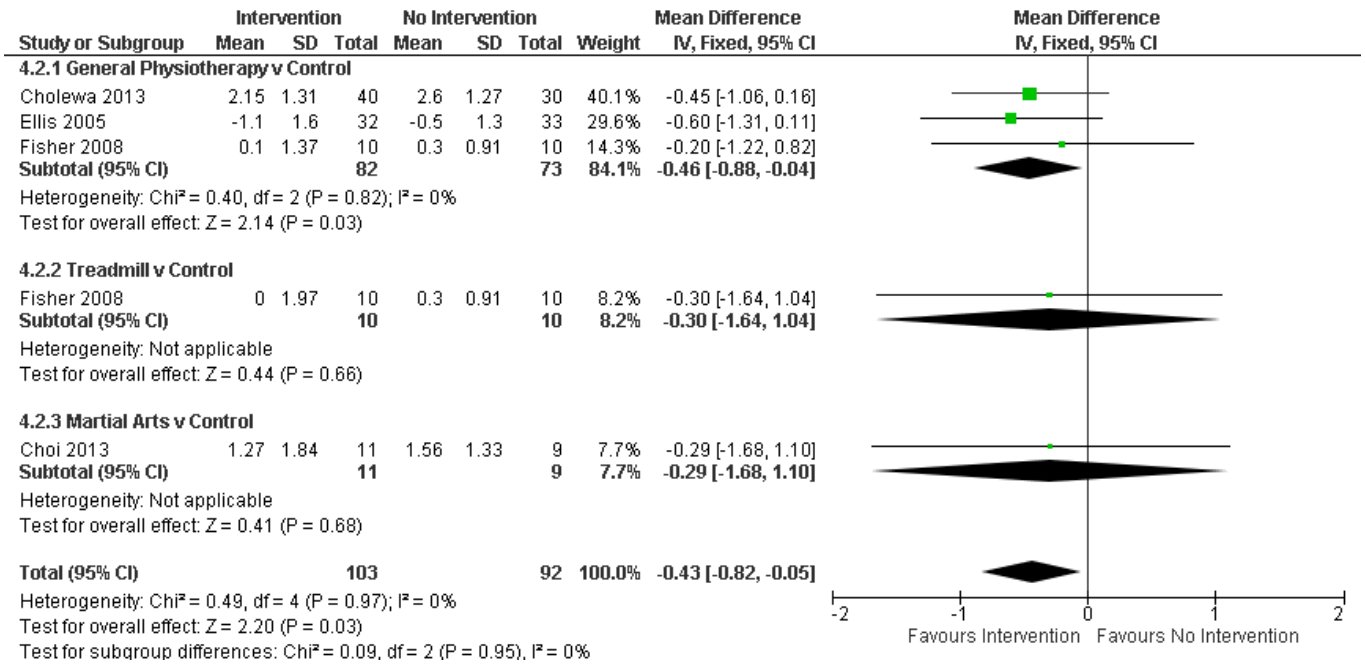


Clinical-Rated Disability

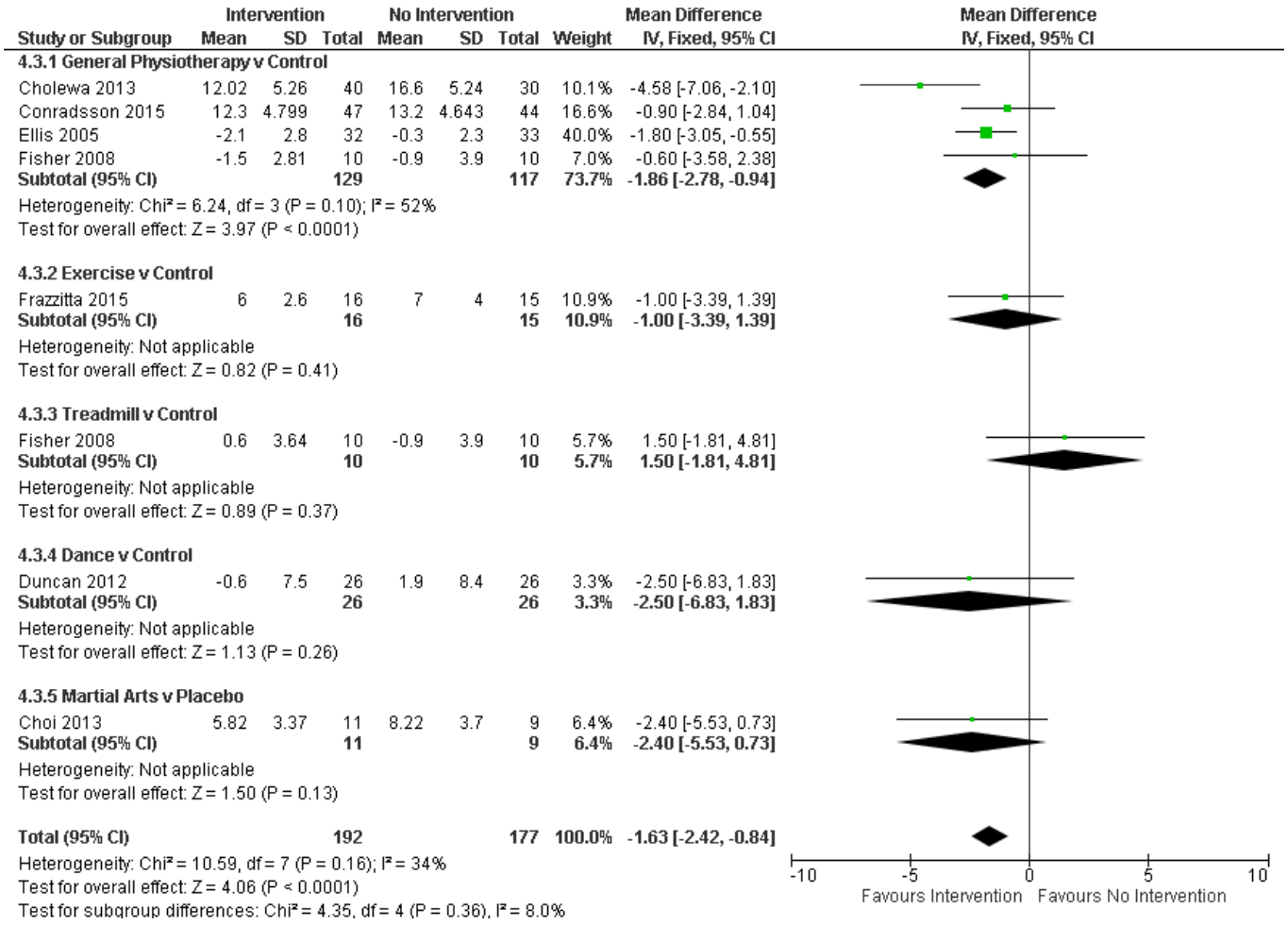
UPDRS Total



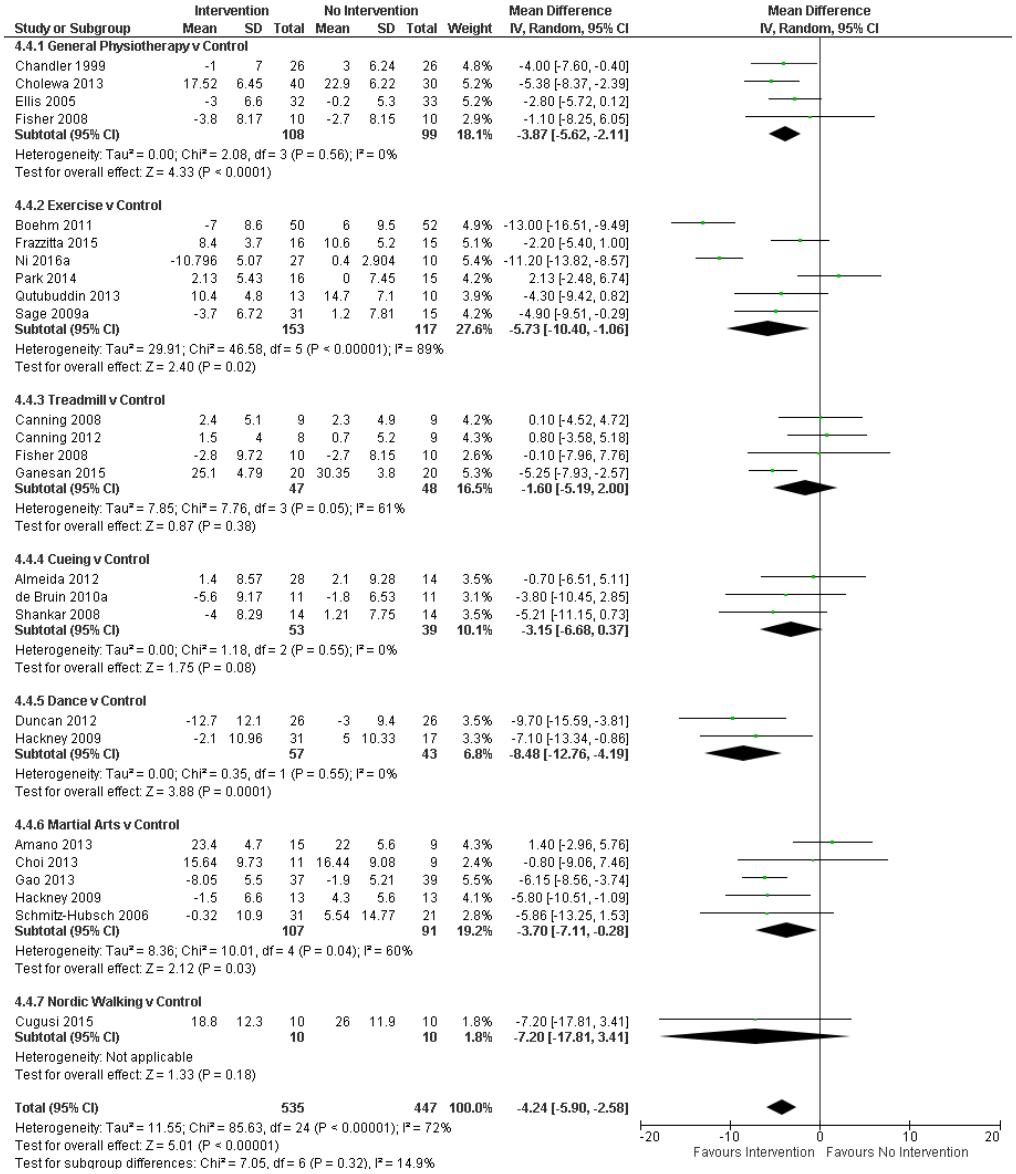
UPDRS Mental



UPDRS ADL

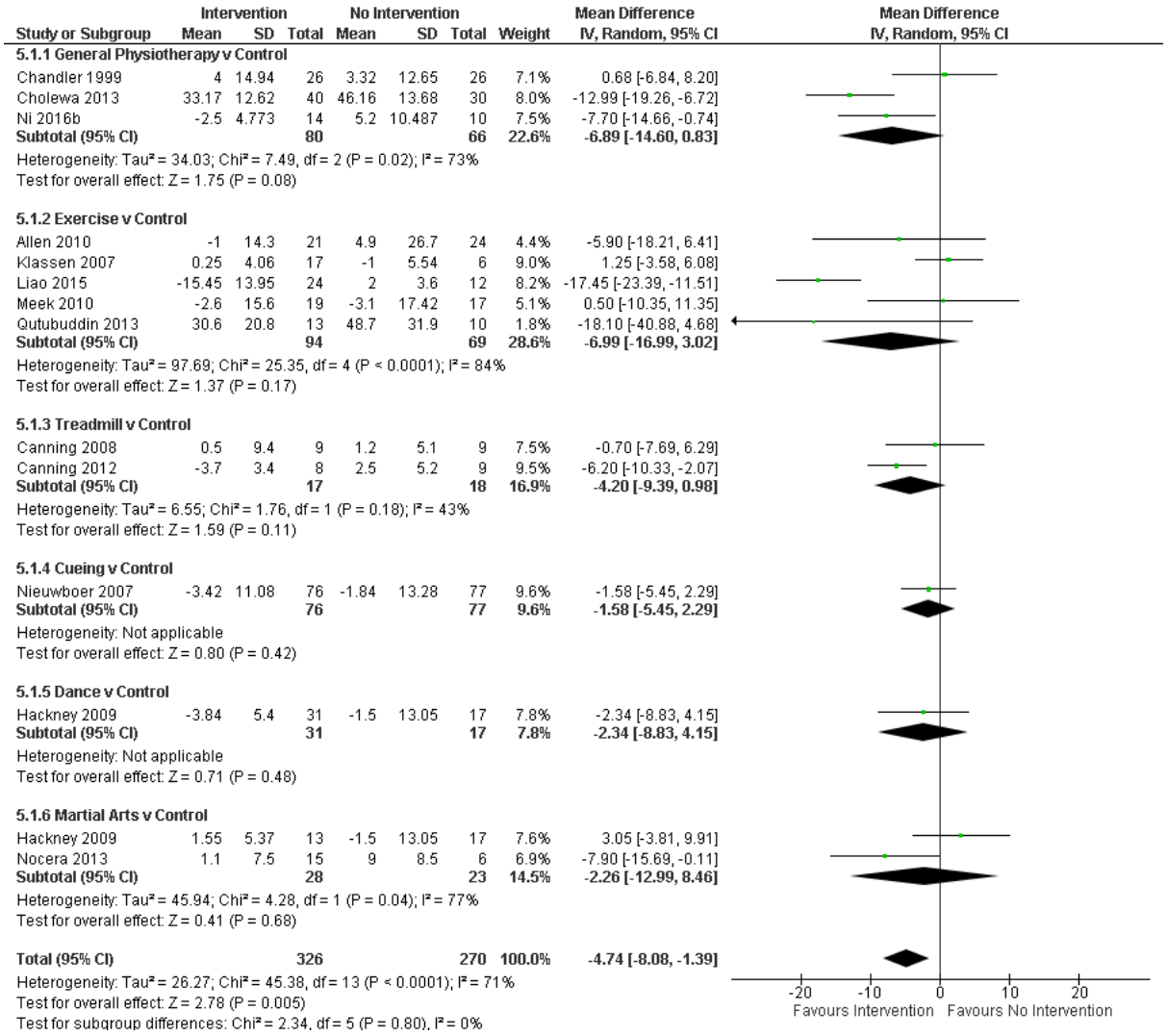

Intervention
No Intervention
Mean Difference
Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Fixed, 95% CI
IV, Fixed, 95% CI
4.3.1 General Physiotherapy v Control
Cholewa 2013 12.02 5.26 40 16.6 5.24 30 10.1% -4.58 [-7.06, -2.10]
Conradsson 2015 12.3 4.799 47 13.2 4.643 44 16.6% -0.90 [-2.84, 1.04]
Ellis 2005 -2.1 2.8 32 -0.3 2.3 33 40.0% -1.80 [-3.05, -0.55]
Fisher 2008 -1.5 2.81 10 -0.9 3.9 10 7.0% -0.60 [-3.58, 2.38]
Subtotal (95% CI) 129 117 73.7% -1.86 [-2.78, -0.94]
Heterogeneity: Chi² = 6.24, df = 3 (P = 0.10); I² = 52%
Test for overall effect: Z = 3.97 (P < 0.0001)
4.3.2 Exercise v Control
Frazzitta 2015 6 2.6 16 7 4 15 10.9% -1.00 [-3.39, 1.39]
Subtotal (95% CI) 16 15 10.9% -1.00 [-3.39, 1.39]
Heterogeneity: Not applicable
Test for overall effect: Z = 0.82 (P = 0.41)
4.3.3 Treadmill v Control
Fisher 2008 0.6 3.64 10 -0.9 3.9 10 5.7% 1.50 [-1.81, 4.81]
Subtotal (95% CI) 10 10 5.7% 1.50 [-1.81, 4.81]
Heterogeneity: Not applicable
Test for overall effect: Z = 0.89 (P = 0.37)
4.3.4 Dance v Control
Duncan 2012 -0.6 7.5 26 1.9 8.4 26 3.3% -2.50 [-6.83, 1.83]
Subtotal (95% CI) 26 26 3.3% -2.50 [-6.83, 1.83]
Heterogeneity: Not applicable
Test for overall effect: Z = 1.13 (P = 0.26)
4.3.5 Martial Arts v Placebo
Choi 2013 5.82 3.37 11 8.22 3.7 9 6.4% -2.40 [-5.53, 0.73]
Subtotal (95% CI) 11 9 6.4% -2.40 [-5.53, 0.73]
Heterogeneity: Not applicable
Test for overall effect: Z = 1.50 (P = 0.13)
Total (95% CI) 192 177 100.0% -1.63 [-2.42, -0.84]
Heterogeneity: Chi² = 10.59, df = 7 (P = 0.16); I² = 34%
Test for overall effect: Z = 4.06 (P < 0.0001)
Test for subgroup differences: Chi² = 4.35, df = 4 (P = 0.36), I² = 8.0%
-10
-5
0
5
10
Favours Intervention
Favours No Intervention

UPDRS Motor

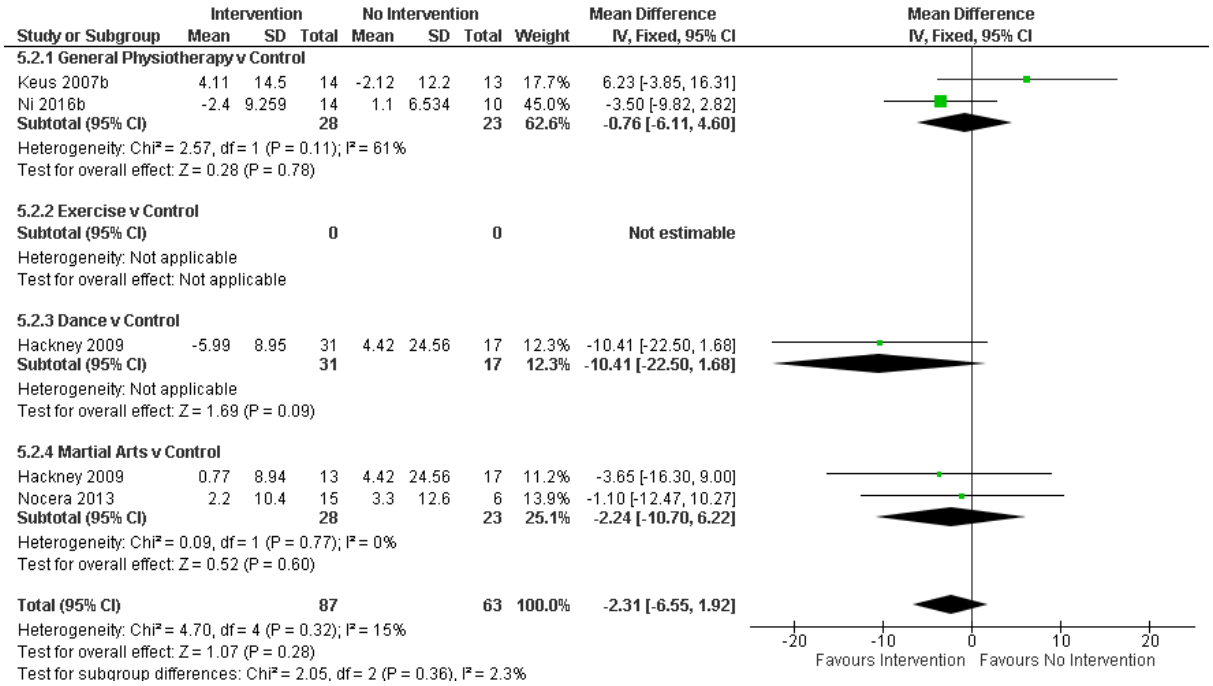

Intervention
No Intervention
Mean Difference
Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Random, 95% CI
IV, Random, 95% CI
4.4.1 General Physiotherapy v Control
Chandler 1999 -1 7 26 3 6.24 26 4.8% -4.00 [-7.60, -0.40]
Cholewa 2013 17.52 6.45 40 22.9 6.22 30 5.2% -5.38 [-8.37, -2.39]
Ellis 2005 -3 6.6 32 -0.2 5.3 33 5.2% -2.80 [-5.72, 0.12]
Fisher 2008 -3.8 8.17 10 -2.7 8.15 10 2.9% -1.10 [-8.25, 6.05]
Subtotal (95% CI) 108 99 18.1% -3.87 [-5.62, -2.11]
Heterogeneity: Tau² = 0.00; Chi² = 2.08, df = 3 (P = 0.56); I² = 0%
Test for overall effect: Z = 4.33 (P < 0.0001)
4.4.2 Exercise v Control
Boehm 2011 -7 8.6 50 6 9.5 52 4.9% -13.00 [-16.51, -9.49]
Frazzitta 2015 8.4 3.7 16 10.6 5.2 15 5.1% -2.20 [-5.40, 1.00]
Ni 2016a -10.796 5.07 27 0.4 2.904 10 5.4% -11.20 [-13.82, -8.57]
Park 2014 2.13 5.43 16 0 7.45 15 4.2% 2.13 [-2.48, 6.74]
Qutubuddin 2013 10.4 4.8 13 14.7 7.1 10 3.9% -4.30 [-9.42, 0.82]
Sage 2009a -3.7 6.72 31 1.2 7.81 15 4.2% -4.90 [-9.51, -0.29]
Subtotal (95% CI) 153 117 27.6% -5.73 [-10.40, -1.06]
Heterogeneity: Tau² = 29.91; Chi² = 46.58, df = 5 (P < 0.00001); I² = 89%
Test for overall effect: Z = 2.40 (P = 0.02)
4.4.3 Treadmill v Control
Canning 2008 2.4 5.1 9 2.3 4.9 9 4.2% 0.10 [-4.52, 4.72]
Canning 2012 1.5 4 8 0.7 5.2 9 4.3% 0.80 [-3.58, 5.18]
Fisher 2008 -2.8 9.72 10 -2.7 8.15 10 2.6% -0.10 [-7.96, 7.76]
Ganesan 2015 25.1 4.79 20 30.35 3.8 20 5.3% -5.25 [-7.93, -2.57]
Subtotal (95% CI) 47 48 16.5% -1.60 [-5.19, 2.00]
Heterogeneity: Tau² = 7.85; Chi² = 7.76, df = 3 (P = 0.05); I² = 61%
Test for overall effect: Z = 0.87 (P = 0.38)
4.4.4 Cueing v Control
Almeida 2012 1.4 8.57 28 2.1 9.28 14 3.5% -0.70 [-6.51, 5.11]
de Bruin 2010a -5.6 9.17 11 -1.8 6.53 11 3.1% -3.80 [-10.45, 2.85]
Shankar 2008 -4 8.29 14 1.21 7.75 14 3.5% -5.21 [-11.15, 0.73]
Subtotal (95% CI) 53 39 10.1% -3.15 [-6.68, 0.37]
Heterogeneity: Tau² = 0.00; Chi² = 1.18, df = 2 (P = 0.55); I² = 0%
Test for overall effect: Z = 1.75 (P = 0.08)
4.4.5 Dance v Control
Duncan 2012 -12.7 12.1 26 -3 9.4 26 3.5% -9.70 [-15.59, -3.81]
Hackney 2009 -2.1 10.96 31 5 10.33 17 3.3% -7.10 [-13.34, -0.86]
Subtotal (95% CI) 57 43 6.8% -8.48 [-12.76, -4.19]
Heterogeneity: Tau² = 0.00; Chi² = 0.35, df = 1 (P = 0.55); I² = 0%
Test for overall effect: Z = 3.88 (P = 0.0001)
4.4.6 Martial Arts v Control
Amano 2013 23.4 4.7 15 22 5.6 9 4.3% 1.40 [-2.96, 5.76]
Choi 2013 15.64 9.73 11 16.44 9.08 9 2.4% -0.80 [-9.06, 7.46]
Gao 2013 -8.05 5.5 37 -1.9 5.21 39 5.5% -6.15 [-8.56, -3.74]
Hackney 2009 -1.5 6.6 13 4.3 5.6 13 4.1% -5.80 [-10.51, -1.09]
Schmitz-Hubsch 2006 -0.32 10.9 31 5.54 14.77 21 2.8% -5.86 [-13.25, 1.53]
Subtotal (95% CI) 107 91 19.2% -3.70 [-7.11, -0.28]
Heterogeneity: Tau² = 8.36; Chi² = 10.01, df = 4 (P = 0.04); I² = 60%
Test for overall effect: Z = 2.12 (P = 0.03)
4.4.7 Nordic Walking v Control
Cugusi 2015 18.8 12.3 10 26 11.9 10 1.8% -7.20 [-17.81, 3.41]
Subtotal (95% CI) 10 10 1.8% -7.20 [-17.81, 3.41]
Heterogeneity: Not applicable
Test for overall effect: Z = 1.33 (P = 0.18)
Total (95% CI) 535 447 100.0% -4.24 [-5.90, -2.58]
Heterogeneity: Tau² = 11.55; Chi² = 85.63, df = 24 (P < 0.00001); I² = 72%
Test for overall effect: Z = 5.01 (P < 0.00001)
Test for subgroup differences: Chi² = 7.05, df = 6 (P = 0.32), I² = 14.9%
-20 -10 0 10 20
Favours Intervention Favours No Intervention

Clinical-Rated QoL

PDQ-39 Summary Index



PDQ-39 Mobility



E.5.2 Occupational therapy

Patient health related quality of life

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	OT	Control	Mean difference (MD) (95% CI)	Quality
Generic health related quality of life: EQ5D									
Sturkenboom 2014	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[4]	122	63	0.03; 95%CI -0.03 to 0.08	MODERATE
Parkinson's disease health related quality of life: PDQ 39									
Sturkenboom 2014	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[5]	122	63	-1.7; 95%CI -3.9 to 0.5	MODERATE

[1] Low risk of bias, as assessed by NICE RCT quality checklist
[2] N/A: Not applicable as only one study contributed to this analysis
[3] No serious indirectness; population was as described in review protocol
[4] No serious imprecision; confidence intervals are tight

Activities of daily living

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	OT	Control	Mean difference (MD) (95% CI)	Quality
Canadian participation 3 months									
Sturkenboom 2014	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	122	63	1.2; 95%C: I 0.8 to 1.6	HIGH
Canadian participation 6 months									

[1] Low risk of bias, as assessed by NICE RCT quality checklist
[2] NA: Not applicable as only one study contributed to this analysis
[3] No serious indirectness; population was as described in review protocol
[4] Serious imprecision: Non-significant results
[5] CI cross the MID of 1.6 points (Peto et al., 2001)

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	OT	Control	Mean difference (MD) (95% CI)	Quality
Sturkenboom 2014	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	122	63	0.9; 95%CI 0.5 to 1.3	HIGH
Canadian satisfaction 3 months									
Sturkenboom 2014	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	122	63	1.1; 95%CI 0. to 1.5	HIGH
Canadian satisfaction 6 months									
Sturkenboom 2014	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	122	63	0.9; 95%CI: 0.5 to 1.3	HIGH

[1] Low risk of bias, as assessed by NICE RCT quality checklist
[2] N/A: Not applicable as only one study contributed to this analysis
[3] No serious indirectness; population was as described in review protocol
[4] No serious imprecision; confidence intervals are tight

Recreation and leisure participation

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	OT	Control	Mean difference (MD) (95% CI)	Quality
Utrecht proactive coping competence scale									
Sturkenboom 2014	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[4]	122	63	0.09: 95%CI -0.02 to 1.21	MODERATE
Utrecht evaluation of rehabilitation participation satisfaction scale									
Sturkenboom	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[4]	122	63	3.2; 95%CI -0.6 to 6.8	MODERATE

[1] Low risk of bias, as assessed by NICE RCT quality checklist
[2] NA: Not applicable as only one study contributed to this analysis
[3] No serious indirectness; population was as described in review protocol
[4] Non-significant results

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	OT	Control	Mean difference (MD) (95% CI)	Quality
2014									

[1] Low risk of bias, as assessed by NICE RCT quality checklist
[2] N/A: Not applicable as only one study contributed to this analysis
[3] No serious indirectness; population was as described in review protocol
[4] No serious imprecision; confidence intervals are tight

Fatigue

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	OT	Control	Mean difference (MD) (95% CI)	Quality
Fatigue severity assessment									
Sturkenboom 2014	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[4]	122	63	0.1; 95%CI -0.2 to 0.4	MODERATE

[1] Low risk of bias, as assessed by NICE RCT quality checklist
[2] NA: Not applicable as only one study contributed to this analysis
[3] No serious indirectness; population was as described in review protocol
[4] No serious imprecision; confidence intervals are tight

Depression

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	OT	Control	Mean difference (MD) (95% CI)	Quality
Becks depression index									
Sturkenboom 2014	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[4]	121	62	-1.4; 95%CI -3.0 to 0.3	MODERATE

[1] Low risk of bias, as assessed by NICE RCT quality checklist

[1] Low risk of bias, as assessed by NICE RCT quality checklist
[2] NA: Not applicable as only one study contributed to this analysis
[3] No serious indirectness; population was as described in review protocol
[4] Serious imprecision; non-significant results

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	OT	Control	Mean difference (MD) (95% CI)	Quality
[2] N/A: Not applicable as only one study contributed to this analysis [3] No serious indirectness; population was as described in review protocol [4] No serious imprecision; confidence intervals are tight									

Carer quality of life

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	OT	Control	Mean difference (MD) (95% CI)	Quality
Carer quality of life: EQ5D 3 month follow-up									
Sturkenboom 2014	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	112	58	0.06; 95%CI: 0.02 to 0.11	HIGH
Carer quality of life EQ5D: 6 month follow up									
Sturkenboom 2014	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[4]	104	59	0.04; 95%CI -0.01 to 0.3	MODERATE
[1] Low risk of bias, as assessed by NICE RCT quality checklist [2] N/A: Not applicable as only one study contributed to this analysis [3] No serious indirectness; population was as described in review protocol [4] No serious imprecision; confidence intervals are tight									

E.5.3 Speech and language therapy

Speech impairment: Frenchay dysarthria score

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Therapy	control	Mean difference (95% CI)	Quality
Johnson (1990)	RCT	Serious[1]	N/A[2]	Serious[3]	Not serious	6	6	29 (13.66 to 44.34)	LOW

[1] Serious risk of bias: eligibility criteria, randomisation method, concealment of allocation, and placebo all inadequately described
[2] N/A: not applicable, as only one study contributed to analysis
[3] Serious indirectness: Eligibility criteria of population of interest ill-defined. It is only implied that all patients had PD.

Vocal loudness

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Therapy	control	Mean difference (95% CI)	Quality
Monologue reading									
2 studies: Johnson (1990) Ramig (2001)	RCT	Serious[1]	Serious[4]	Serious[3]	Not serious	29	21	6.17dB (3.57 to 8.77)	VERY LOW
Monologue reading - 6 month follow up									
Ramig (2001)	RCT	Serious[1]	N/A[2]	Serious[3]	Not serious	14	15	3.5dB (0.9 to 6.1)	LOW
Standard passage reading									
2 studies:	RCT	Serious[1]	Serious[5]	Serious[3]	Not serious	20	21	7.18dB (4.65 to 9.71).	VERY LOW

[1] Serious risk of bias: eligibility criteria, randomisation method, concealment of allocation, and placebo all inadequately described
[2] NA: not applicable, as only one study contributed to analysis
[3] Serious indirectness: Eligibility criteria of population of interest ill-defined. It is only implied that all patients had PD.
[4] Serious inconsistency: I^2 >40%
[5] Serious inconsistency: I^2 >40%

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Therapy	control	Mean difference (95% CI)	Quality
Johnson (1990) Ramig (2001)									
Standard passage reading - 6 month follow up									
Ramig (2001)	RCT	Serious[1]	N/A[2]	Serious[3]	Not serious	14	15	4.5dB (95%CI: 1.9 to 7.1)	LOW
Loudness of prolonged 'ah' sound									
Ramig (2001)	RCT	Serious[1]	N/A[2]	Serious[3]	Not serious	14	15	12.1 dB (8.9 to 15.4)	LOW
Loudness of prolonged 'ah' sound - 6 month follow up									
Ramig (2001)	RCT	Serious[1]	N/A[2]	Serious[3]	Not serious	14	15	9.4 dB (6.2 to 12.6)	LOW

[1] Serious risk of bias: eligibility criteria, randomisation method, concealment of allocation, and placebo all inadequately described
[2] N/A: not applicable, as only one study contributed to analysis
[3] Serious indirectness: Eligibility criteria of population of interest ill-defined. It is only implied that all patients had PD.
[4] Serious imprecision: confidence intervals are wide and cross line of no effect. Cochrane group cite 29 point change as potentially clinically meaningful.
[5] No serious inconsistency; confidence intervals of estimates overlap

Monotonicity

						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Therapy	control	Mean difference (95% CI)	Quality
Maximum pitch range									
Johnson (1990)	RCT	Serious[4]	N/A[2]	Serious[3]	Very serious[5]	6	6	66Hz (-4.4 to 136.6)	VERY LOW
Maximum volume range									

[1] Serious risk of bias; Poor randomisation method and poor concealment of allocation. Credibility of placebo condition not clear
[2] NA: not applicable, as only one study contributed to analysis
[3] Serious indirectness: Eligibility criteria of population of interest ill-defined. It is only implied that all patients had PD.
[4] Serious risk of bias: eligibility criteria, randomisation method, adequate concealment of allocation, and adequate placebo all inadequately described
[5] Very serious imprecision: Non-significant results and very wide CIs

						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Therapy	control	Mean difference (95% CI)	Quality
Johnson (1990)	RCT	Serious[4]	N/A[2]	Serious[3]	Not serious	6	6	23.7dB (9.3 to 38.1)	LOW

[1] Serious risk of bias: eligibility criteria, randomisation method, concealment of allocation, and placebo all inadequately described
[2] N/A: not applicable, as only one study contributed to analysis
[3] Serious indirectness: Eligibility criteria of population of interest ill-defined. It is only implied that all patients had PD.
[4] Serious imprecision: confidence intervals are wide and cross line of no effect. Cochrane group cite 29 point change as potentially clinically meaningful.

Swallowing safety: penetration aspiration

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	EMST	sham	Mean difference (95% CI)	Quality
Troche (2010)	RCT	Not serious[1]	N/A[4]	Not serious[2]	Not serious	30	30	-1.23 (-2.23 to -0.23)	HIGH

[1] Serious risk of bias: eligibility criteria, randomisation method, concealment of allocation, and placebo all inadequately described
[2] NA: not applicable, as only one study contributed to analysis
[3] Serious indirectness: Eligibility criteria of population of interest ill-defined. It is only implied that all patients had PD.
[4] Serious imprecision: confidence intervals are wide and cross line of no effect. Cochrane group cite 29 point change as potentially clinically meaningful.

Swallowing mechanism: duration of hyoid elevation (s)

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	EMST	Sham	Mean difference (95% CI)	Quality
Troche (2010)	RCT	Not serious[1]	N/A[4]	Not serious[2]	serious[3]	30	30	0.07s (-4.69 to 4.83)	MODERATE

[1] Serious risk of bias: eligibility criteria, randomisation method, concealment of allocation, and placebo all inadequately described
[2] N/A: not applicable, as only one study contributed to analysis
[3] Serious indirectness: Eligibility criteria of population of interest ill-defined. It is only implied that all patients had PD.
[4] Serious imprecision: confidence intervals are wide and cross line of no effect. Cochrane group cite 29 point change as potentially clinically meaningful.

[1] Low risk of bias, as assessed by NICE RCT quality checklist
[2] No serious indirectness: population clearly defined and match that outlined in review protocol
[3] Serious imprecision: non-significant results

Health related quality of life

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	EMST	sham	ANOVA F score, p value	Quality
Troche	RCT	Not serious[1]	N/A[1]	Not serious[2]	Not serious	30	30	F=3.007 (p=0.007)	LOW

[1] Serious risk of bias: eligibility criteria, randomisation method, concealment of allocation, and placebo all inadequately described
[2] N/A: not applicable, as only one study contributed to analysis

[1] NA: not applicable, as only one study contributed to analysis

E.5.4 Nutrition

Low protein diet

Quality assessment						No of patients			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Effect size (95% CI)	Quality
Total "on" time (Barichella 2006)									
1 (Barichella 2006)	RCT	Very serious[1,2,3]	N/A	Not serious	Not serious	18	18	MD 114 (19.92 to 208.08)	LOW
Postprandial "on" time (Barichella 2006)									
1 (Barichella 2006)	RCT	Very serious[1,2,3]	N/A	Not serious	Serious[4]	18	18	MD 30 (-17.04 to 77.0)	VERY LOW
Total "off" time (Barichella 2006)									
1 (Barichella 2006)	RCT	Very serious[1,2,3]	N/A	Not serious	Not serious	18	18	MD -107 (-212.53 to -1.47)	LOW
Postprandial "off" time (Barichella 2006)									
1 (Barichella 2006)	RCT	Very serious[1,2,3]	N/A	Not serious	Serious[4]	18	18	MD -30 (-77.37 to 17.37)	VERY LOW

1. Unclear if allocation concealed
2. Inadequate blinding or no blinding
3. Outcomes self-reported
4. Serious imprecision: Non-significant result

Low protein redistribution diet

Quality assessment						No of patients		Effect size (95% CI)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control		
Percentage of "on" time									
1 (Tsul 1989)	RCT	Very serious[1,2,3,4]	N/A	Not serious	Serious[5]	10	10	MD 10.55 (-4.28, 25.58)	VERY LOW
Modified Columbia scores									
1 (Tsul 1989)	RCT	Very serious[1,2,3,4]	N/A	Not serious	Serious[5]	10	10	MD -3.98 (-14.82, 6.86)	VERY LOW
Total "off" time									
1 (Croxson 1991)	RCT	Very serious[2,6]	N/A	Not serious	Serious[5]	8	8	MD -0.81 (-6.23, 4.61)	VERY LOW

1. Unclear method of randomisation
2. Unclear if allocation concealed
3. No precise definition of outcome
4. Inappropriate length of follow-up
5. Serious imprecision: Non-significant result
6. Outcomes self-reported

Comparison of two low-protein diets

Quality assessment						No of patients		Effect size (95% CI)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control		
Time spent in physical activity									
1 (Barichella 2007)	RCT	Very serious[1,2,3,4]	N/A	Not serious	Serious[5]	6	6	MD 0.37 (-1.13, 1.87)	VERY LOW
Energy expenditure									
1 (Barichella 2007)	RCT	Very serious[1,2,3,4]	N/A	Not serious	Serious[5]	6	6	MD 172 (-128, 472)	VERY LOW

Quality assessment						No of patients			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Effect size (95% CI)	Quality
Patient global improvement (very much better/much better)									
1 (Barichella 2007)	RCT	Very serious[1,2,3,4]	N/A	Not serious	Serious[5]	6	6	RR 13.00 (0.89, 189.39)	VERY LOW
Patient global improvement (very much better/much better)									
1 (Barichella 2007)	RCT	Very serious[1,2,3,4]	N/A	Not serious	Serious[5]	6	6	RR 0.08 (0.01, 1.12)	VERY LOW

1. Unclear if allocation concealed
2. Inadequate blinding or no blinding
3. Inappropriate length of follow up
4. Outcomes self reported
5. Serious imprecision: Non-significant result

High fibre supplement

Quality assessment						No of patients			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Effect size (95% CI)	Quality
Absorption: area under the curve									
1 (Fernandez-Martinez 2014)	RCT	Serious[1]	N/A	Not serious	Serious[2]	18	18	MD -0.63 (-10.30, 9.04)	LOW
Absorption: peak plasma concentration									
1 (Fernandez-Martinez 2014)	RCT	Serious[1]	N/A	Not serious	Serious[2]	18	18	MD -64.20 (-184.92, 56.52)	LOW
Absorption: time to peak blood level									

Quality assessment						No of patients			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Effect size (95% CI)	Quality
1 (Fernandez-Martinez 2014)	RCT	Serious[1]	N/A	Not serious	Serious[2]	18	18	MD 3.55 (-10.96, 18.06)	LOW

1. Unclear if allocation concealed
2. Serious imprecision: non-significant results

Fasting diet

Quality assessment						No of patients			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Effect size (95% CI)	Quality
Absorption: area under the curve									
1 (Brefel 1998)	RCT	Very serious[1,2,3]	N/A	Not serious	Serious[4]	12	12	MD 3.20 (-4.93, 11.33)	VERY LOW
Absorption: peak plasma concentration									
1 (Brefel 1998)	RCT	Very serious[1,2,3]	N/A	Not serious	Serious[4]	12	12	MD 1.52 (-0.16, 3.20)	VERY LOW
Absorption: time to peak plasma concentration									
1 (Brefel 1998)	RCT	Very serious[1,2,3]	N/A	Not serious	Serious[5]	12	12	MD -2.12 (-2.81, -1.43)	VERY LOW

1. Unclear method of randomisation
2. Unclear if allocation concealed
3. Inadequate blinding or no blinding
4. Serious imprecision: non-significant results
5. Means and SD imputed from medians and ranges

Creatine supplementation

Quality assessment						No of patients		Effect size (95% CI)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control		
SF-36 General health perception (mean difference from baseline) - better indicated by lower values									
1 (Bender 2006)	RCT	Very serious[2]	N/A	Not serious	Serious[3]	40	20	MD 5.00 (-4.53, 14.53)	VERY LOW
SF-36 Vitality (mean difference from baseline) - better indicated by lower values									
1 (Bender 2006)	RCT	Very serious[2]	N/A	Not serious	Serious[3]	40	20	MD 7.00 (-1.43, 15.43)	VERY LOW
SF-36 Role limitations (mean difference from baseline) - better indicated by lower values									
1 (Bender 2006)	RCT	Very serious[2]	N/A	Not serious	Not serious	40	20	MD 21.00 (5.29, 36.70)	LOW
SF-36 Mental health (mean difference from baseline) - better indicated by lower values									
1 (Bender 2006)	RCT	Very serious[2]	N/A	Not serious	Not serious	40	20	MD 8.00 (0.03, 15.97)	LOW
SF-36 Social functioning (mean difference from baseline) - better indicated by lower values									
1 (Bender 2006)	RCT	Very serious[2]	N/A	Not serious	Serious[3]	40	20	MD 4.00 (-5.62, 13.62)	VERY LOW
SF-36 Bodily pain (mean difference from baseline) - better indicated by lower values									
1 (Bender 2006)	RCT	Very serious[2]	N/A	Not serious	Serious[3]	40	20	MD -6.00 (-21.12, 9.12)	VERY LOW
SF-36 Role limitations (mean difference from baseline) - better indicated by lower values									
1 (Bender 2006)	RCT	Very serious[2]	N/A	Not serious	Serious[3]	40	20	MD -10.00 (-30.32, 10.32)	VERY LOW
SF-36 Physical functioning (mean difference from baseline) - better indicated by lower values									
1 (Bender 2006)	RCT	Very serious[2]	N/A	Not serious	Serious[3]	40	20	MD -4.00 (-14.08, 6.08)	VERY LOW
EQ-5D (mean difference from baseline)									
1 (Kieburtz 2015)	RCT	Not serious	N/A	Not serious	Serious[3]	334	342	MD 0.005 (-0.03, 0.04)	MODERATE

Quality assessment						No of patients			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Effect size (95% CI)	Quality
PDQ-39 Summary Index (mean difference from baseline)									
1 (Kieburtz 2015)	RCT	Not serious	N/A	Not serious	Very serious[6]	477	478	MD -1.2 (-4.2, 1.7)	LOW
Total UPDRS score (mean difference from baseline) - better indicated by lower values									
2 (Bender 2006, Kieburtz 2015)	RCT	Serious[1]	Not serious	Not serious	Serious[3]	370	356	MD 1.02 (-1.11, 3.15)	LOW
Total UPDRS score (mean difference from baseline) - better indicated by lower values									
1 (Haas 2007)	RCT	Serious[1]	N/A	Not serious	Serious[3]	10	10	MD -1.70 (-7.08, 3.68)	LOW
UPDRS complications (mean difference from baseline) - better indicated by lower values									
1 (Bender 2006)	RCT	Very serious[2]	N/A	Not serious	Serious[3]	40	20	MD 0.20 (-0.55, 0.95)	VERY LOW
UPDRS motor score (mean difference from baseline) - better indicated by lower values									
2 (Bender 2006, Kieburtz 2015)	RCT	Serious[1]	Not serious	Not serious	Not serious	370	356	MD 0.44 (-1.02, 1.90)	MODERATE
UPDRS motor score (mean difference from baseline) - better indicated by lower values									
1 (Haas 2007)	RCT	Serious[1]	N/A	Not serious	Serious[6]	10	10	MD -3.90 (-8.03, 0.23)	LOW
UPDRS ADL score (mean difference from baseline) - better indicated by lower values									
2 (Bender 2006, Kieburtz 2015)	RCT	Serious[1]	Not serious	Not serious	Serious[4]	373	359	MD 1.30 (-1.12, 3.72)	LOW
UPDRS ADL score (mean difference from baseline) - better indicated by lower values									
1 (Haas 2007)	RCT	Serious[1]	N/A	Not serious	Not serious	10	10	MD -0.20 (-2.20, 1.80)	MODERATE

Quality assessment						No of patients			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Effect size (95% CI)	Quality
UPDRS mental health score (mean difference from baseline) - better indicated by lower values									
2 (Bender 2006, Kieburtz 2015)	RCT	Serious[1]	Serious[5]	Not serious	Serious[3]	373	359	MD -0.0 (-0.27, 0.26)	VERY LOW
UPDRS mental health score (mean difference from baseline) - better indicated by lower values									
1 (Haas 2007)	RCT	Serious[1]	N/A	Not serious	Serious[3]	10	10	MD 0.40 (-0.08, 0.88)	LOW
Hoehn and Yahr score (mean difference from baseline) - better indicated by lower values									
1 (Haas 2007)	RCT	Serious[1]	N/A	Not serious	Not serious	10	10	MD -0.40 (-0.58, -0.22)	MODERATE
Mass, kg (mean difference from baseline) - better indicated by lower values									
1 (Haas 2007)	RCT	Serious[1]	N/A	Not serious	Serious[3]	10	10	MD 0.40 (-4.74, 5.54)	LOW
BMI (mean difference from baseline)									
1 (Kieburtz 2015)	RCT	Not serious	N/A	Not serious	Serious[3]	338	341	MD -0.3 (-0.8, 0.2)	MODERATE
Levodopa dose (mean difference from baseline) - better indicated by lower values									
1 (Bender 2006)	RCT	Very serious[2]	N/A	Not serious	Serious[3]	40	20	MD -57.0 (-145.3, 31.3)	VERY LOW
Dopamine agonist dose (mean difference from baseline) - better indicated by lower values									
1 (Bender 2006)	RCT	Very serious[2]	N/A	Not serious	Not serious	40	20	MD -132.0 (-195.8, -68.3)	LOW

1. Serious risk of bias detected
2. Very serious risk of bias detected
3. Non-significant result
4. CI cross the MID of 3 (Schrag et al., 2006)
5. $I^2 > 40\%$
6. CI cross the MID of 1.6 (Peto 2001)

Amino acid supplementation

Quality assessment						No of patients		Effect size (95% CI)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control		
Body weight (kg) (mean difference from baseline) - better indicated by lower values									
1 (Cucca 2015)	RCT	Serious[1]	N/A	Not serious	Serious[2]	7	7	MD -6.50 (-13.71, 0.71)	LOW
Body motor score (kg) (mean difference from baseline) - better indicated by lower values									
1 (Cucca 2015)	RCT	Serious[1]	N/A	Not serious	Very serious[3]	7	7	MD 3.20 (-3.60, 10.00)	VERY LOW

1. Serious risk of bias detected
2. Non-significant result
3. CI cross the MID of 3.25 (Horvath et al., 2015)

Co-enzyme Q10 supplemention

Quality assessment						No of patients		Effect size (95% CI)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control		
Total UPDRS (mean difference from baseline) - better indicated by lower values									
3 (Negida 2016)	SR	Not serious	Not serious	Not serious	Serious[1]	475	468	MD -0.05 (-0.25, 0.15)	MODERATE
UPDRS motor score (mean difference from baseline) - better indicated by lower values									
4 (Negida 2016)	SR	Not serious	Not serious	Not serious	Not serious	546	539	MD 0.05 (-0.07, 0.17)	HIGH
UPDRS ADL score (mean difference from baseline) - better indicated by lower values									
4 (Negida 2016)	SR	Not serious	Serious[2]	Not serious	Not serious	546	539	MD -0.10 (-0.35, 0.15)	MODERATE
UPDRS mental health score (mean difference from baseline) - better indicated by lower values									
4 (Negida 2016)	SR	Not serious	Not serious	Not serious	Serious[1]	546	539	MD -0.03 (-0.23, 0.17)	MODERATE
Schwab and England modified score (mean difference from baseline) - better indicated by lower values									

Quality assessment						No of patients		Effect size (95% CI)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control		
3 (Negida 2016)	SR	Not serious	Serious[2]	Not serious	Serious[1]	546	539	MD 0.08 (-0.17, 0.29)	LOW
UPDRS ADL/motor score (mean difference from baseline) - better indicated by lower values									
1 (Storch 2007)	RCT	Not serious	N/A	Not serious	Serious[1]	64	67	MD 2.15 (-1.08, 5.38)	MODERATE

1. Non-significant result
2. Considerable between study heterogeneity

Trigonella foenum-gracum l seeds

Quality assessment						No of patients		Effect size (95% CI)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control		
Total UPDRS (mean difference from baseline) - better indicated by lower values									
1 (Nathan 2014)	RCT	Not serious	N/A	Not serious	Serious[1]	23	19	MD -5.36 (-13.70, 2.98)	MODERATE
UPDRS motor score (mean difference from baseline) - better indicated by lower values									
1 (Nathan 2014)	RCT	Not serious	N/A	Not serious	Serious[2]	23	19	MD -4.76 (-11.82, 2.30)	MODERATE
UPDRS ADL score (mean difference from baseline) - better indicated by lower values									
1 (Nathan 2014)	RCT	Not serious	N/A	Not serious	Very serious[2]	23	19	MD 0.07 (-3.66, 3.80)	LOW
UPDRS mental health score (mean difference from baseline) - better indicated by lower values									
1 (Nathan 2014)	RCT	Not serious	N/A	Not serious	Serious[1]	23	19	MD -0.65 (-2.03, 0.73)	MODERATE
Hoehn and Yahr Stage reversal									
1 (Nathan 2014)	RCT	Not serious	N/A	Not serious	Serious[1]	23	19	RR 4.13 (0.53, 32.38)	MODERATE
Hoehn and Yahr Stage unchanged									

Quality assessment						No of patients		Effect size (95% CI)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control		
1 (Nathan 2014)	RCT	Not serious	N/A	Not serious	Serious[1]	23	19	RR 0.83 (0.57, 1.21)	MODERATE
Hoehn and Yahr Stage advancement									
1 (Nathan 2014)	RCT	Not serious	N/A	Not serious	Serious[1]	23	19	RR 0.83 (0.19, 3.63)	MODERATE

1. Non-significant result
2. CI cross the MID of 3.25 (Horvath et al., 2015)

Vitamin D supplementation

Quality assessment						No of patients		Effect size (95% CI)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control		
PDQ39 total (mean difference from baseline) - better indicated by lower values									
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Serious[1]	55	57	MD 2.26 (-8.72, 4.20)	MODERATE
PDQ39 cognitive impairment (mean difference from baseline) - better indicated by lower values									
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Serious[2]	55	57	MD -1.50 (-8.08, 5.08)	MODERATE
PDQ39 social support (mean difference from baseline) - better indicated by lower values									
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Serious[2]	55	57	MD -3.65 (-10.53, 3.23)	MODERATE
PDQ39 bodily support (mean difference from baseline) - better indicated by lower values									
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Serious[2]	55	57	MD -5.67 (-13.63, 2.29)	MODERATE
PDQ39 communication (mean difference from baseline) - better indicated by lower values									
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Serious[2]	55	57	MD -2.17 (-9.70, 5.36)	MODERATE
PDQ39 stigma (mean difference from baseline) - better indicated by lower values									

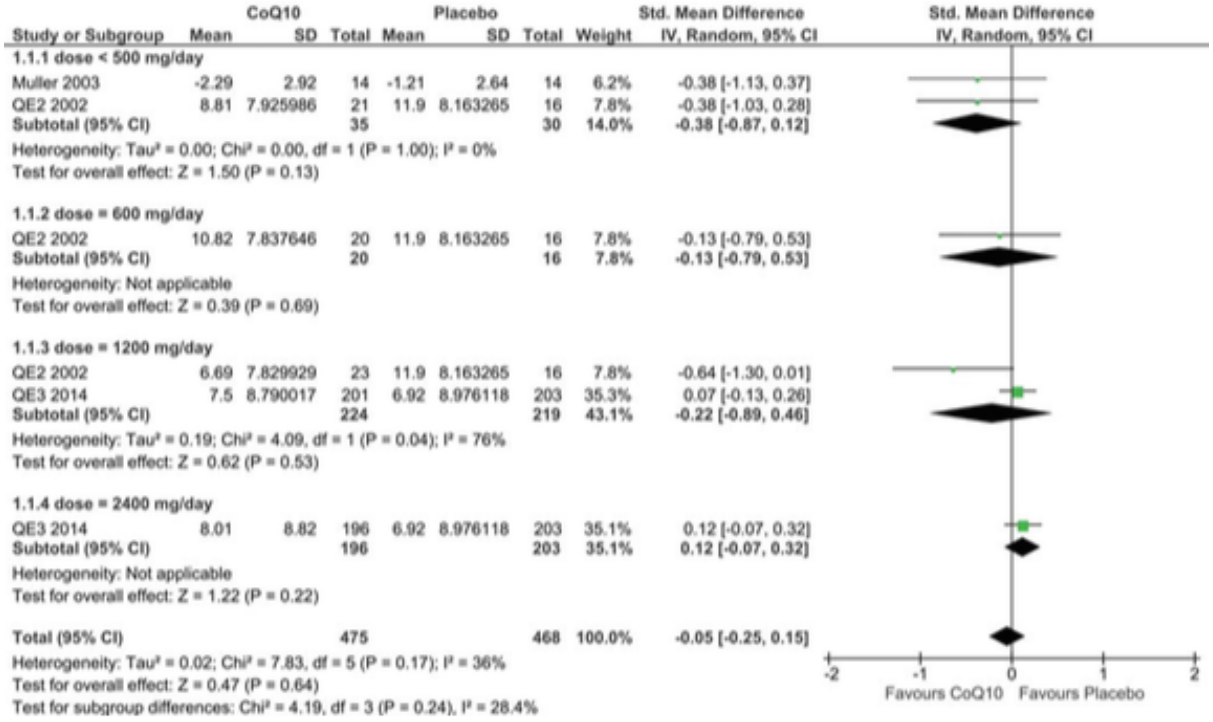
Quality assessment						No of patients			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control	Effect size (95% CI)	Quality
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Serious[2]	55	57	MD 5.75 (-1.88, 13.38)	MODERATE
PDQ39 emotional wellbeing (mean difference from baseline) - better indicated by lower values									
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Serious[2]	55	57	MD -1.71 (-9.94, 6.52)	MODERATE
PDQ39 activities of daily living (mean difference from baseline) - better indicated by lower values									
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Serious[2]	55	57	MD -1.64 (-10.64, 7.36)	MODERATE
PDQ39 mobility (mean difference from baseline) - better indicated by lower values									
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Serious[2]	55	57	MD -3.03 (-12.62, 6.56)	MODERATE
EQ-5D (mean difference from baseline) - better indicated by lower values									
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Serious[2]	55	57	MD 0.05 (-0.05, 0.15)	MODERATE
MMSE (Stage 1-5) - better indicated by lower values									
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Serious[2]	55	57	MD -0.60 (-1.33, 0.13)	MODERATE
Total UPDRS (mean difference from baseline) - better indicated by lower values									
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Not serious	55	57	MD -5.07 (-10.13, -0.01)	HIGH
UPDRS complications (mean difference from baseline) - better indicated by lower values									
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Serious[2]	55	57	MD -0.09 (-0.62, 0.44)	MODERATE
UPDRS motor score (mean difference from baseline) - better indicated by lower values									
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Serious[3]	55	57	MD -2.10 (-5.64, 1.44)	MODERATE
UPDRS ADL score (mean difference from baseline) - better indicated by lower values									
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Serious[3]	55	57	MD -5.24 (-10.32, -0.16)	MODERATE

Quality assessment						No of patients		Effect size (95% CI)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Intervention	Control		
UPDRS mental health score (mean difference from baseline) - better indicated by lower values									
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Serious[2]	55	57	MD -0.38 (-0.93, 0.17)	MODERATE
Hoehn & Yahr score (mean difference from baseline) - better indicated by lower values									
1 (Suzuki 2013)	RCT	Not serious	N/A	Not serious	Not serious	55	57	MD -0.31 (-0.55, 0.07)	HIGH

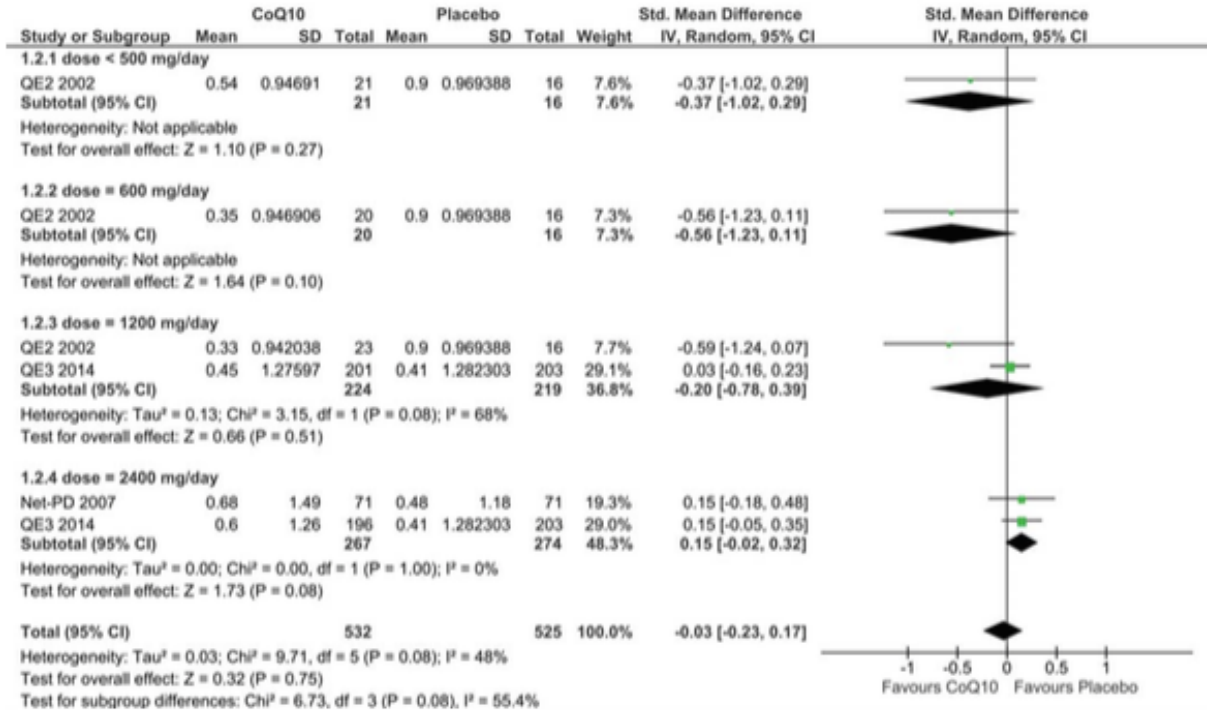
1. CI cross the MID of 1.6 (Peto et al., 2001)
2. Non-significant result
3. CI cross the MID of 3.25 (Horvath et al., 2015)
4. CI cross the MID of 3 (Schrag et al., 2006)

Forest plots (CoQ10 supplementation)

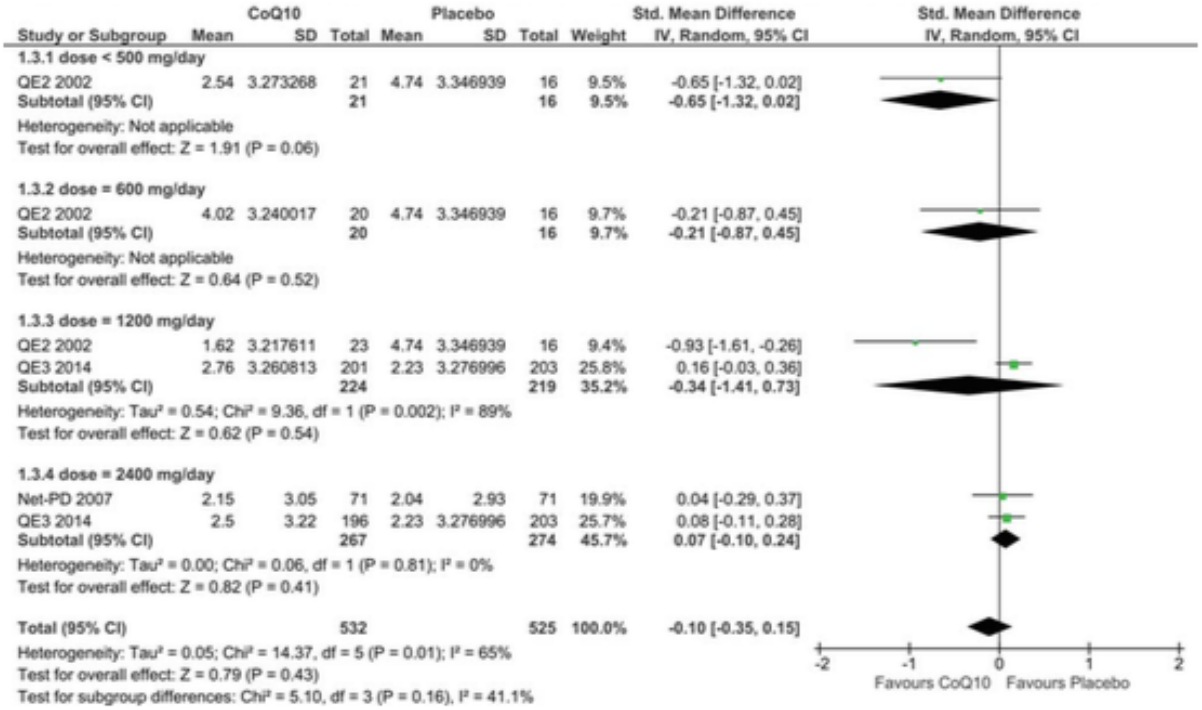
UPDRS Total



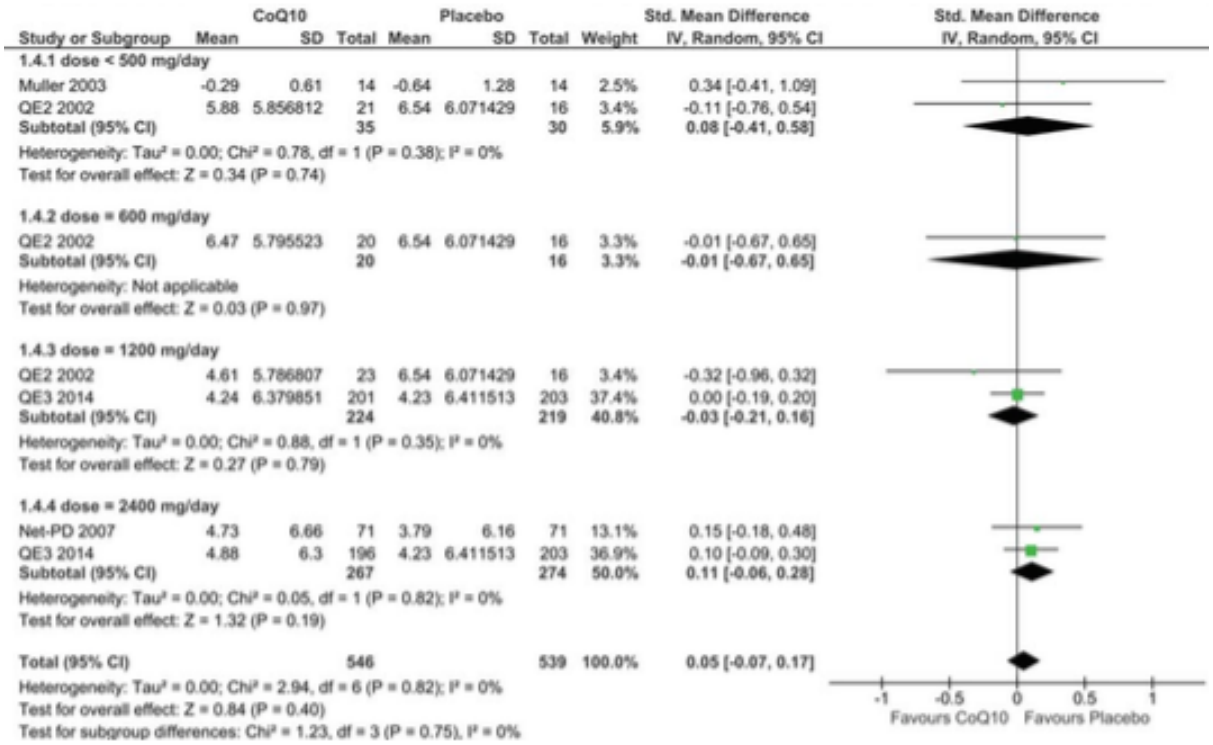
UPDRS I (mental)



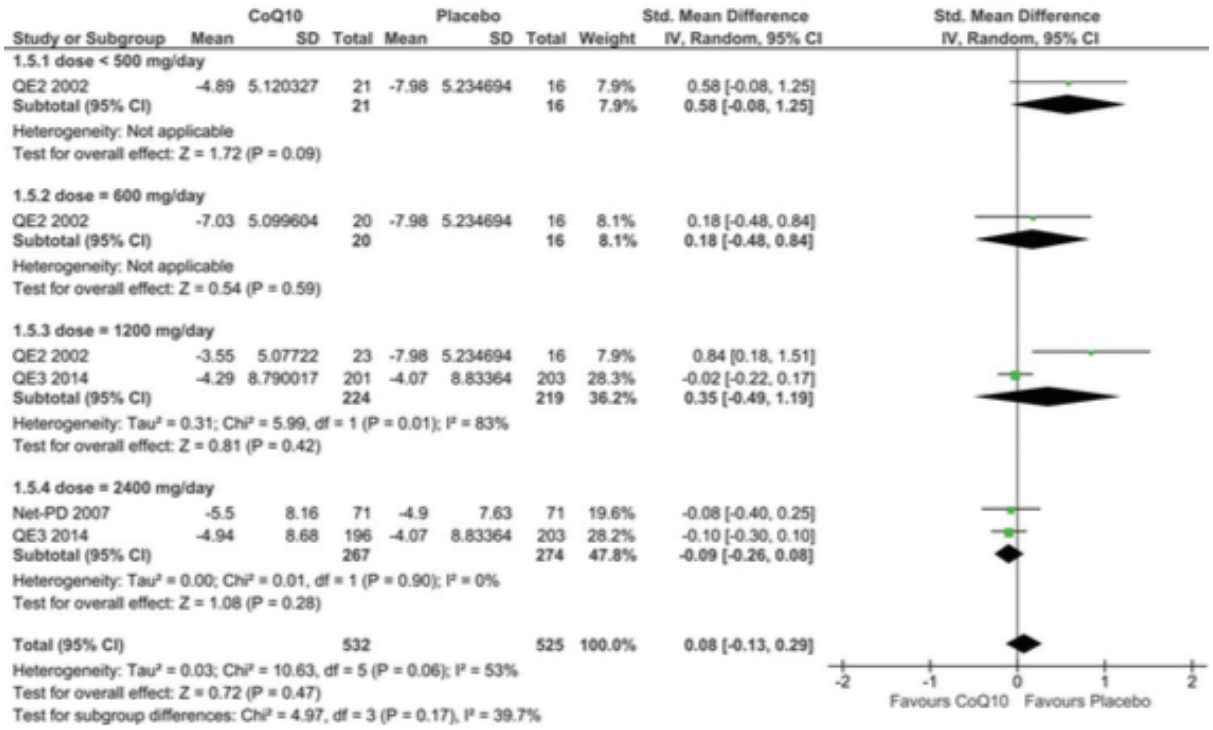
UPDRS II (ADL)


CoQ10
Placebo
Std. Mean Difference
Std. Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Random, 95% CI
IV, Random, 95% CI
1.3.1 dose < 500 mg/day
QE2 2002 2.54 3.273268 21 4.74 3.346939 16 9.5% -0.65 [-1.32, 0.02]
Subtotal (95% CI) 21 16 9.5% -0.65 [-1.32, 0.02]
Heterogeneity: Not applicable
Test for overall effect: Z = 1.91 (P = 0.06)
1.3.2 dose = 600 mg/day
QE2 2002 4.02 3.240017 20 4.74 3.346939 16 9.7% -0.21 [-0.87, 0.45]
Subtotal (95% CI) 20 16 9.7% -0.21 [-0.87, 0.45]
Heterogeneity: Not applicable
Test for overall effect: Z = 0.64 (P = 0.52)
1.3.3 dose = 1200 mg/day
QE2 2002 1.62 3.217611 23 4.74 3.346939 16 9.4% -0.93 [-1.61, -0.26]
QE3 2014 2.76 3.260813 201 2.23 3.276996 203 25.8% 0.16 [-0.03, 0.36]
Subtotal (95% CI) 224 219 35.2% -0.34 [-1.41, 0.73]
Heterogeneity: Tau² = 0.54; Chi² = 9.36, df = 1 (P = 0.002); I² = 89%
Test for overall effect: Z = 0.62 (P = 0.54)
1.3.4 dose = 2400 mg/day
Net-PD 2007 2.15 3.05 71 2.04 2.93 71 19.9% 0.04 [-0.29, 0.37]
QE3 2014 2.5 3.22 196 2.23 3.276996 203 25.7% 0.08 [-0.11, 0.28]
Subtotal (95% CI) 267 274 45.7% 0.07 [-0.10, 0.24]
Heterogeneity: Tau² = 0.00; Chi² = 0.06, df = 1 (P = 0.81); I² = 0%
Test for overall effect: Z = 0.82 (P = 0.41)
Total (95% CI) 532 525 100.0% -0.10 [-0.35, 0.15]
Heterogeneity: Tau² = 0.05; Chi² = 14.37, df = 5 (P = 0.01); I² = 65%
Test for overall effect: Z = 0.79 (P = 0.43)
Test for subgroup differences: Chi² = 5.10, df = 3 (P = 0.16), I² = 41.1%
-2
-1
0
1
2
Favours CoQ10
Favours Placebo

UPDRS III (motor)

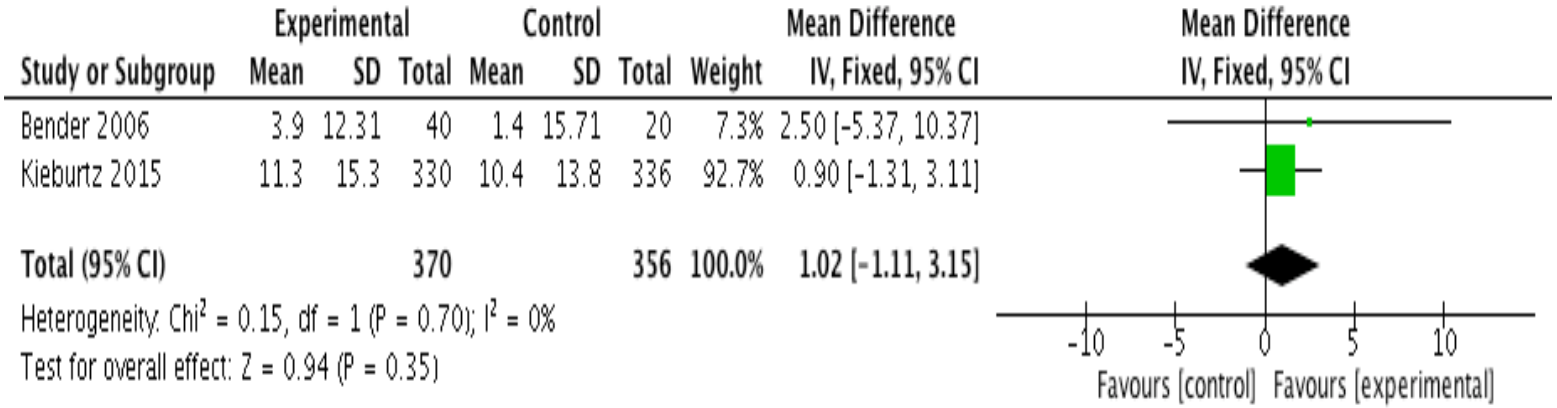


Schwab and England modified score ("for examiner")

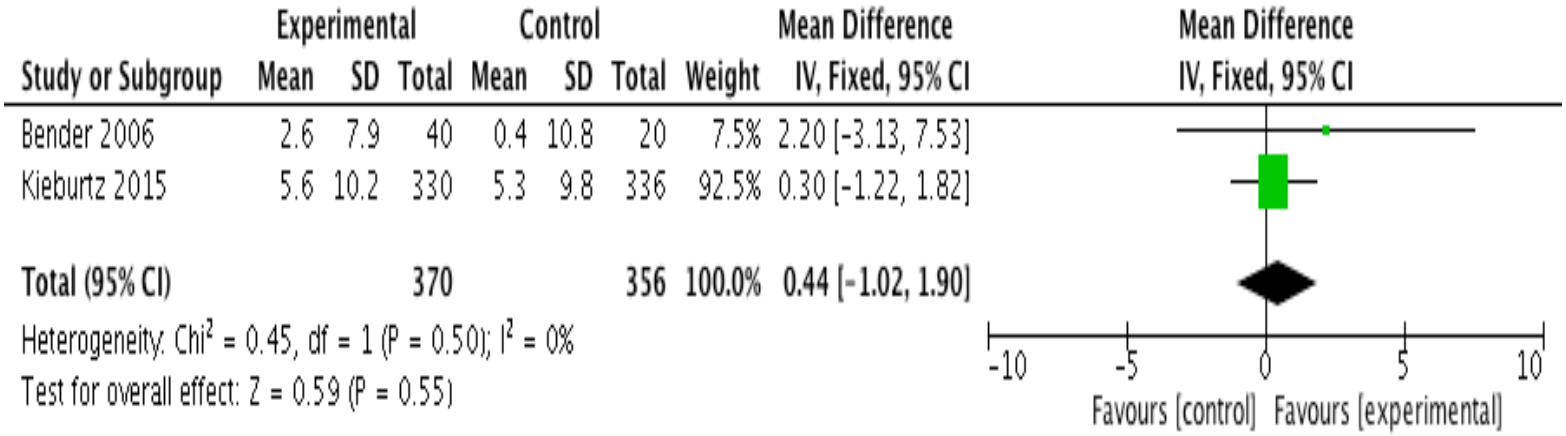


Forest plots (Creatine supplementation)

UPDRS Total



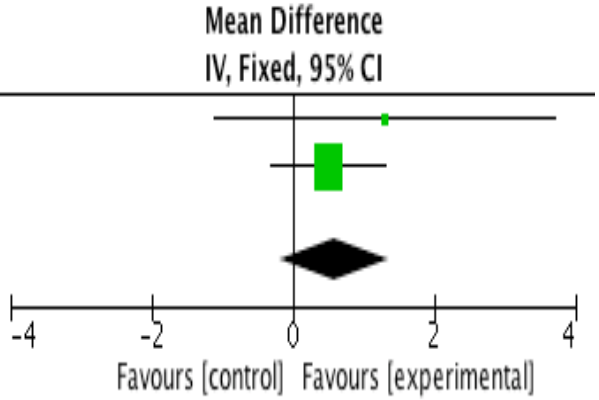
UPDRS Motor



UPDRS ADL

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Bender 2006	1.4	4.5	40	0.1	4.51	20	10.3%	1.30 [-1.12, 3.72]
Kieburtz 2015	4.5	5.7	333	4	5.1	339	89.7%	0.50 [-0.32, 1.32]
Total (95% CI)			373			359	100.0%	**0.58 [-0.19, 1.36]**

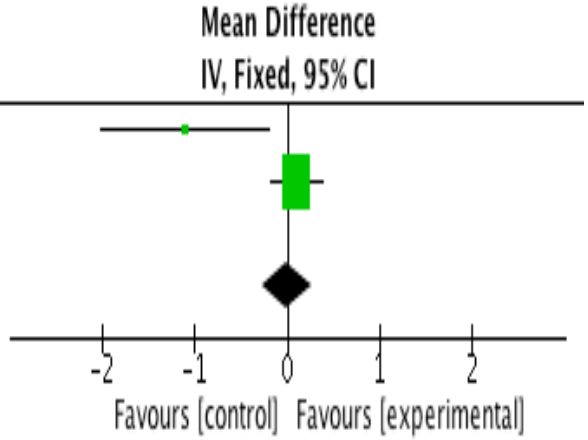
Heterogeneity: $Chi^2 = 0.38$, df = 1 (P = 0.54); $I^2 = 0\%$
Test for overall effect: Z = 1.47 (P = 0.14)



UPDRS Mental

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Bender 2006	-0.3	1.76	40	0.8	1.66	20	8.7%	-1.10 [-2.01, -0.19]
Kieburtz 2015	1.2	1.9	333	1.1	1.8	339	91.3%	0.10 [-0.18, 0.38]
Total (95% CI)			373			359	100.0%	**-0.00 [-0.27, 0.26]**

Heterogeneity: $Chi^2 = 6.11$, df = 1 (P = 0.01); $I^2 = 84\%$
Test for overall effect: Z = 0.03 (P = 0.98)



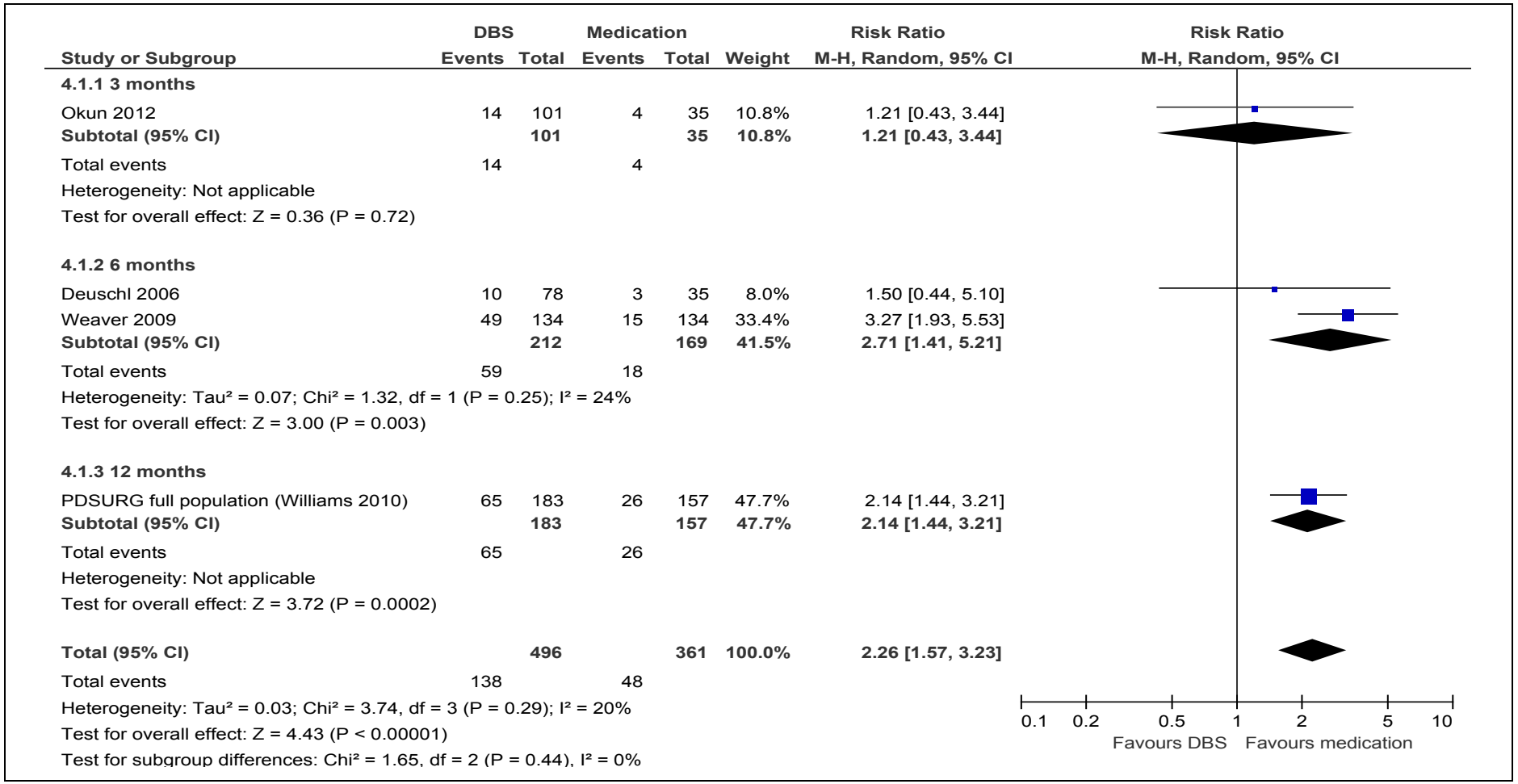
E.6 Advanced therapies: deep brain stimulation and levodopa–carbidopa intestinal gel

E.6.1 Deep brain stimulation compared with best medical treatment for advanced Parkinson's disease

E.6.1.1 Adverse events

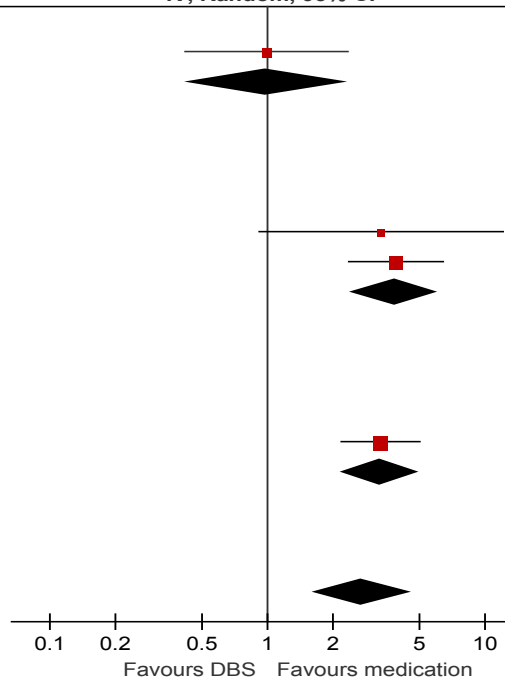
No. of studies	Design	Quality assessment				No. of events / no. of patients or patient-years		Effect (95% CI)	Quality
		Risk of bias	Inconsistency	Indirectness	Imprecision	DBS	Control		
Serious adverse events (probability of experiencing ≥1)									
4[1,2,3,4]	RCT	No serious	Not serious[5]	Serious[6]	No serious	138/496	48/361	RR = 2.26 (1.57 to 3.23)	MODERATE
Serious adverse events (rate per patient-year)									
4[1,2,3,4]	RCT	No serious	Not serious[5]	Serious[6]	No serious	208 per 314.25pt-yrs	58 per 291.25pt-yrs	IRR = 2.72 (1.60 to 4.64)	MODERATE
Falls (probability of experiencing ≥1)									
4[1,2,3,4]	RCT	No serious	Serious[7]	Serious[6]	Serious[8]	29/496	14/361	RR = 1.24 (0.32 to 4.83)	VERY LOW
Falls (rate per patient-year)									
4[1,2,3,4]	RCT	No serious	Serious[7]	Serious[6]	Serious[8]	30 per 314.25pt-yrs	14 per 291.25pt-yrs	IRR = 1.44 (0.45 to 4.62)	VERY LOW

1 *Okun 2012*
2 *Deuschl 2006*
3 *Weaver 2009*
4 *Williams 2010 (main PDSURG publication [all participants regardless of HY score]; no subgroup data available for this outcome)*
5 *Statistical heterogeneity observed; however, this is almost wholly ascribable to differences between Okun 2012 and other studies, and this is explicable on the grounds that participants in the control arm of Okun 2012 underwent surgical implantation of inert device, so not downgraded*
6 *Some RCTs include a nontrivial proportion of participants with less advanced PD than the specified population for this question*
7 *Marked statistical heterogeneity and inconsistency in definition of events (some RCTs report all recorded falls; some falls leading to fracture only)*
8 *At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit and no difference*

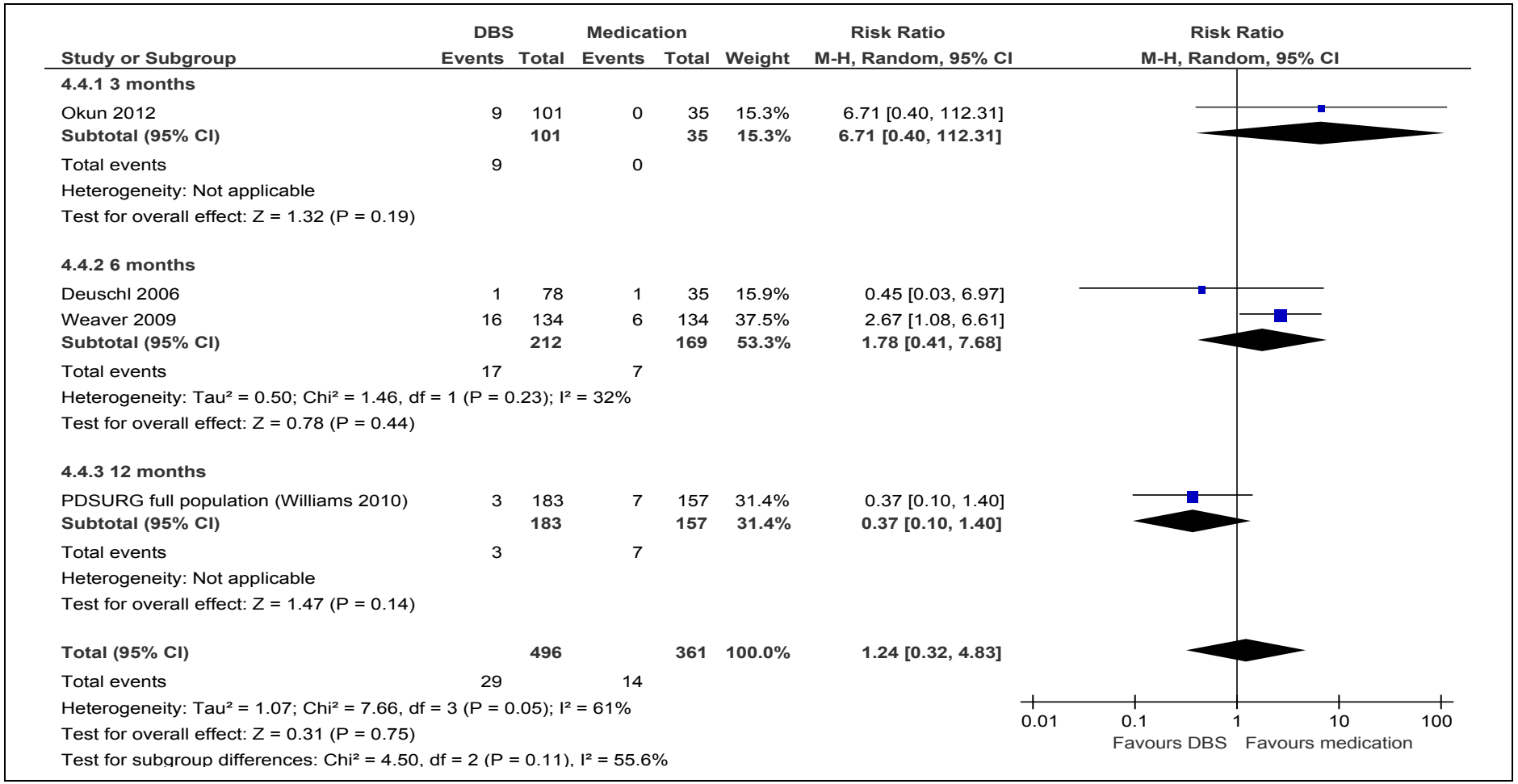


DBS -v- medication alone: serious adverse events (proportion of participants experiencing ≥1) – forest plot

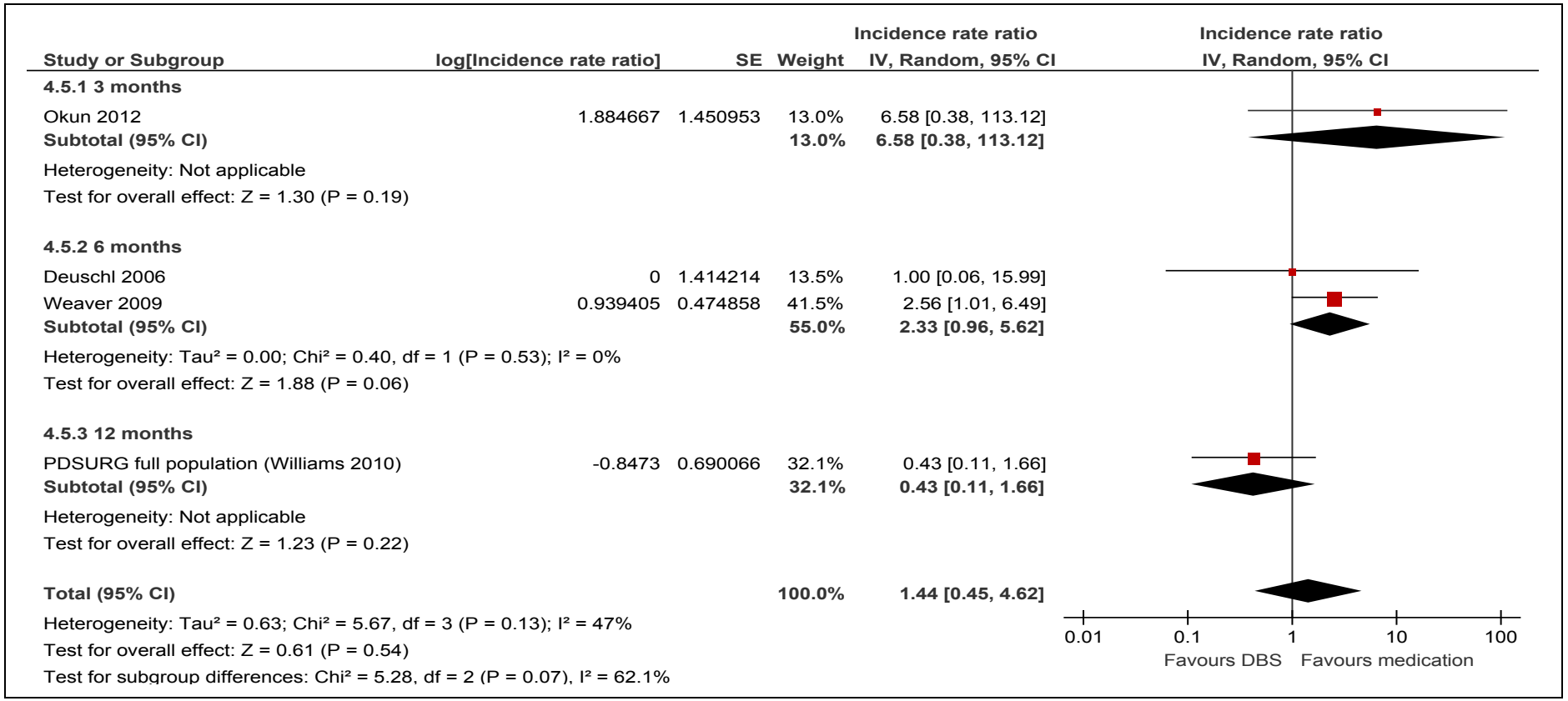
Study or Subgroup	log[Incidence rate ratio]	SE	DBS Total	Medication Total	Weight	Incidence rate ratio IV, Random, 95% CI
4.2.1 3 months						
Okun 2012	-0.00995	0.439155	101	35	20.6%	0.99 [0.42, 2.34]
Subtotal (95% CI)			**101**	**35**	**20.6%**	**0.99 [0.42, 2.34]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.02 (P = 0.98)						
4.2.2 6 months						
Deuschl 2006	1.203973	0.658281	78	35	12.3%	3.33 [0.92, 12.11]
Weaver 2009	1.360231	0.254611	134	134	32.0%	3.90 [2.37, 6.42]
Subtotal (95% CI)			**212**	**169**	**44.3%**	**3.82 [2.40, 6.08]**
Heterogeneity: Tau² = 0.00; Chi² = 0.05, df = 1 (P = 0.82); I² = 0%						
Test for overall effect: Z = 5.64 (P < 0.00001)						
4.2.3 12 months						
PDSURG full population (Williams 2010)	1.197052	0.211895	183	157	35.1%	3.31 [2.19, 5.01]
Subtotal (95% CI)			**183**	**157**	**35.1%**	**3.31 [2.19, 5.01]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 5.65 (P < 0.00001)						
Total (95% CI)			**496**	**361**	**100.0%**	**2.72 [1.60, 4.64]**
Heterogeneity: Tau² = 0.16; Chi² = 7.66, df = 3 (P = 0.05); I² = 61%						
Test for overall effect: Z = 3.69 (P = 0.0002)						
Test for subgroup differences: Chi² = 7.61, df = 2 (P = 0.02), I² = 73.7%						



DBS -v- medication alone: serious adverse events (rate per patient-year) – forest plot



DBS -v- medication alone: falls (proportion of participants experiencing ≥1) – forest plot

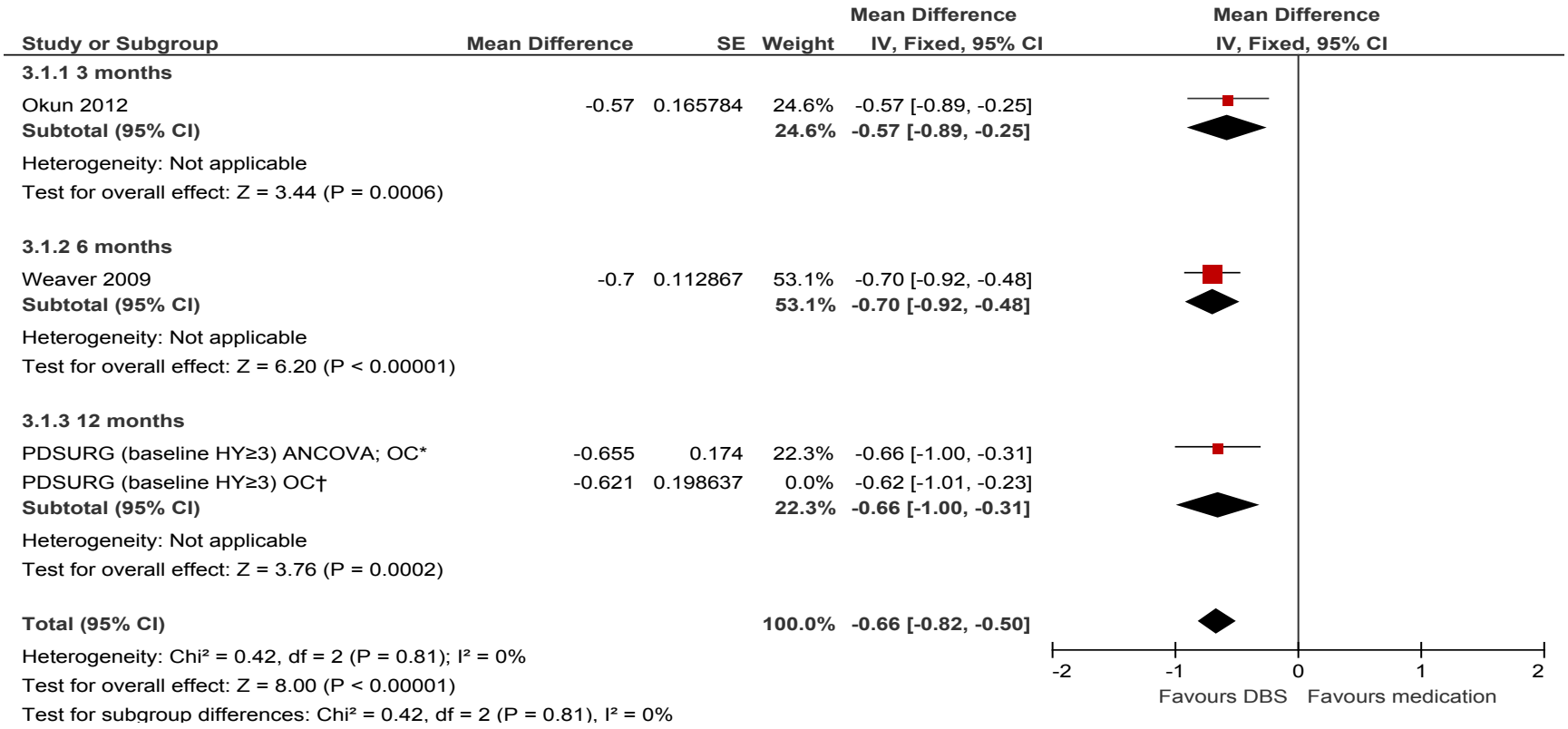


DBS -v- medication alone: falls (rate per patient-year) – forest plot

E.6.1.2 Symptom severity

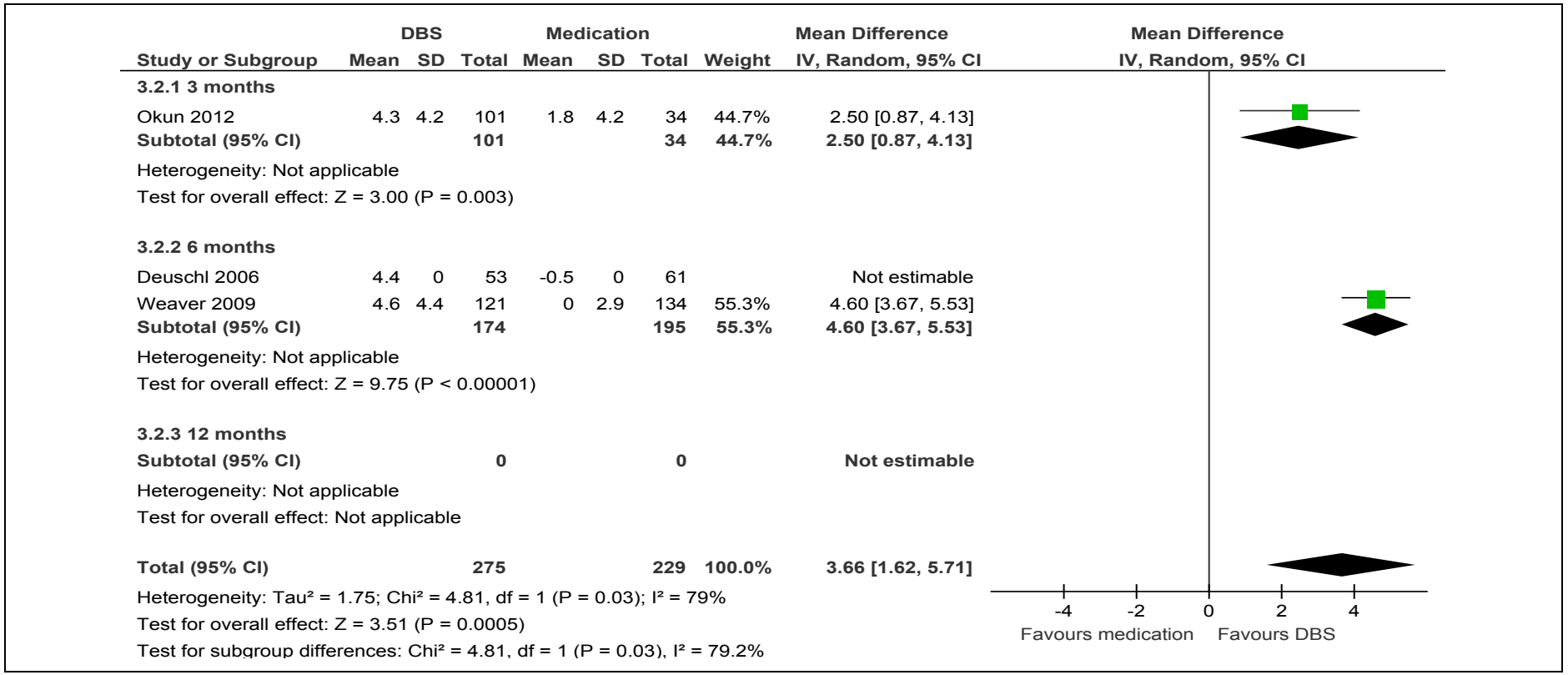
Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	DBS	Control	Mean difference (MD) (95% CI)	Quality
Hoehn and Yahr score (off medication) (lower is better); 3–12 months									
3[1,3,4]	RCT	No serious	No serious	Serious[5]	No serious	261	215	-0.66 (-0.82 to -0.50)	MODERATE
Daily 'on' time without troublesome dysinkesias (higher is better); 3–6 months									
2[1,3]	RCT	No serious	Serious[8]	Serious[5]	No serious	275	229	3.66 (1.62 to 5.71)	LOW
Daily 'off' time (lower is better); 6–12 months									
2[3,4]	RCT	No serious	No serious	Very serious[5,9]	No serious	169	185	-2.48 (-3.10 to -1.86)	LOW
UPDRS I (lower is better); 3–12 months									
4[1,2,3,4]	RCT	No serious	No serious	Serious[5]	No serious	323	281	-0.29 (-0.60 to 0.02)	MODERATE
UPDRS II on (lower is better); 3–12 months									
4[1,2,3,10]	RCT	No serious	No serious[7]	Serious[5]	No serious	352	276	-2.98 (-4.50 to -1.46)	MODERATE
UPDRS III on (lower is better); 3–12 months									
4[1,2,3,10]	RCT	No serious	Serious[8]	Serious[5]	No serious	331	280	-4.93 (-7.52 to -2.34)	LOW
UPDRS IV (lower is better); 3–12 months									
3[1,3,4]	RCT	No serious	Serious[8]	Serious[5]	No serious	243	204	-4.05 (-5.83 to -2.28)	LOW

1 *Okun 2012*
2 *Deuschl 2006*
3 *Weaver 2009*
4 *PDSURG observed cases; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score used to estimate treatment effect in people with Hoehn and Yahr score ≥3 at baseline; calculated by guideline developers from patient-level data supplied by investigators (NB HY score ≥3 was a prespecified subgroup in the trial protocol and a randomisation stratification variable)*
5 *Some RCTs include a nontrivial proportion of participants with less advanced PD than the specified population for this question*
6 *At a 95% confidence level, data are only consistent with no meaningful effect*
7 *Some heterogeneity between 3-month and 6–12-month results; however direction of effect modification appears consistent and plausible, so not downgraded*
8 *I^2 greater than 40% with no obvious explanation for heterogeneity*
9 *PDSURG off time estimate approximated from answer to UPDRS Q39 (categorical proportion of waking day spent 'off')*
10 *PDSURG multiply imputed data; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score used to estimate treatment effect in people with Hoehn and Yahr score ≥3 at baseline; calculated by guideline developers from patient-level data supplied by investigators (NB HY score ≥3 was a prespecified subgroup in the trial protocol and a randomisation stratification variable)*

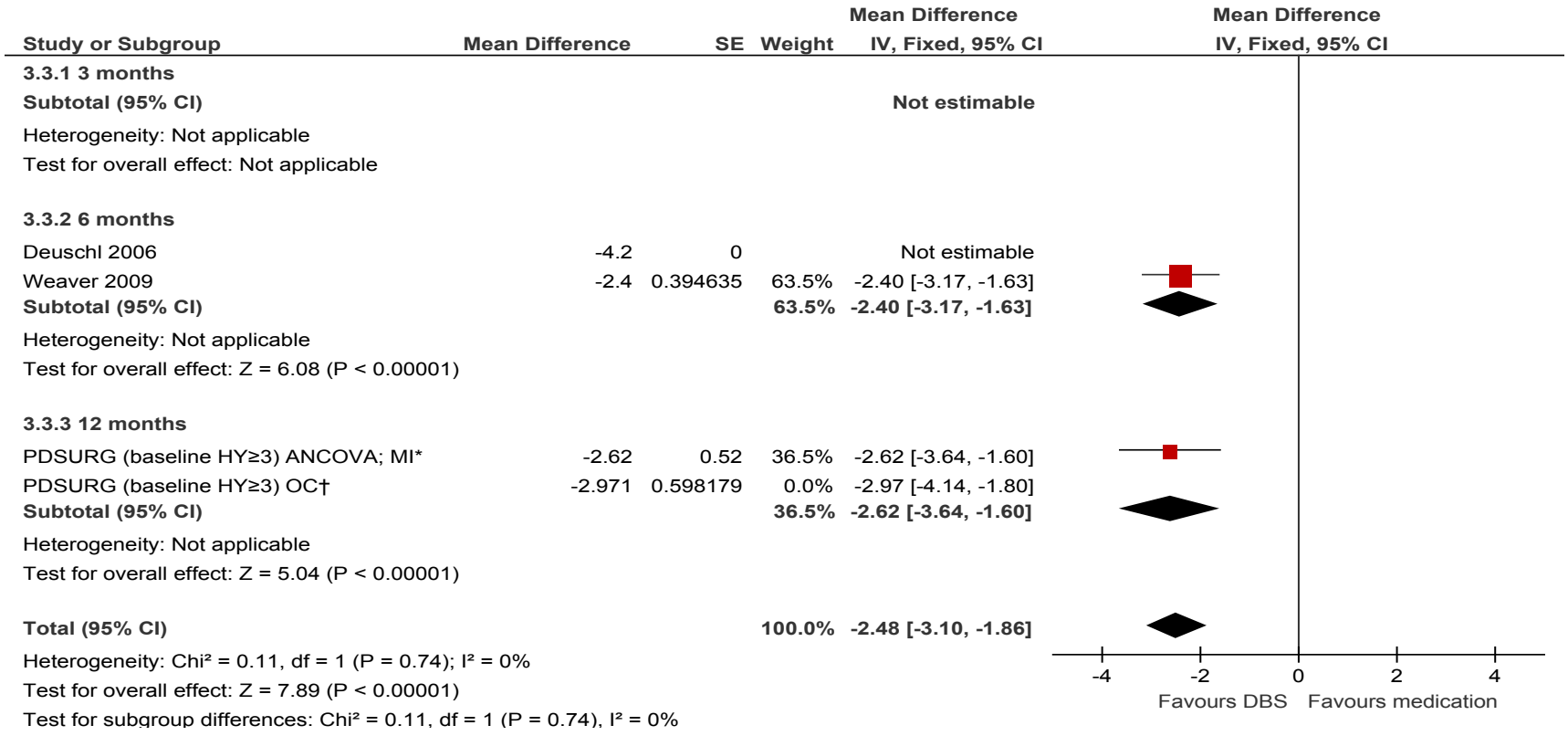


** observed cases; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score*
† observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only

DBS -v- medication alone: Hoehn and Yahr score (off medication) – forest plot

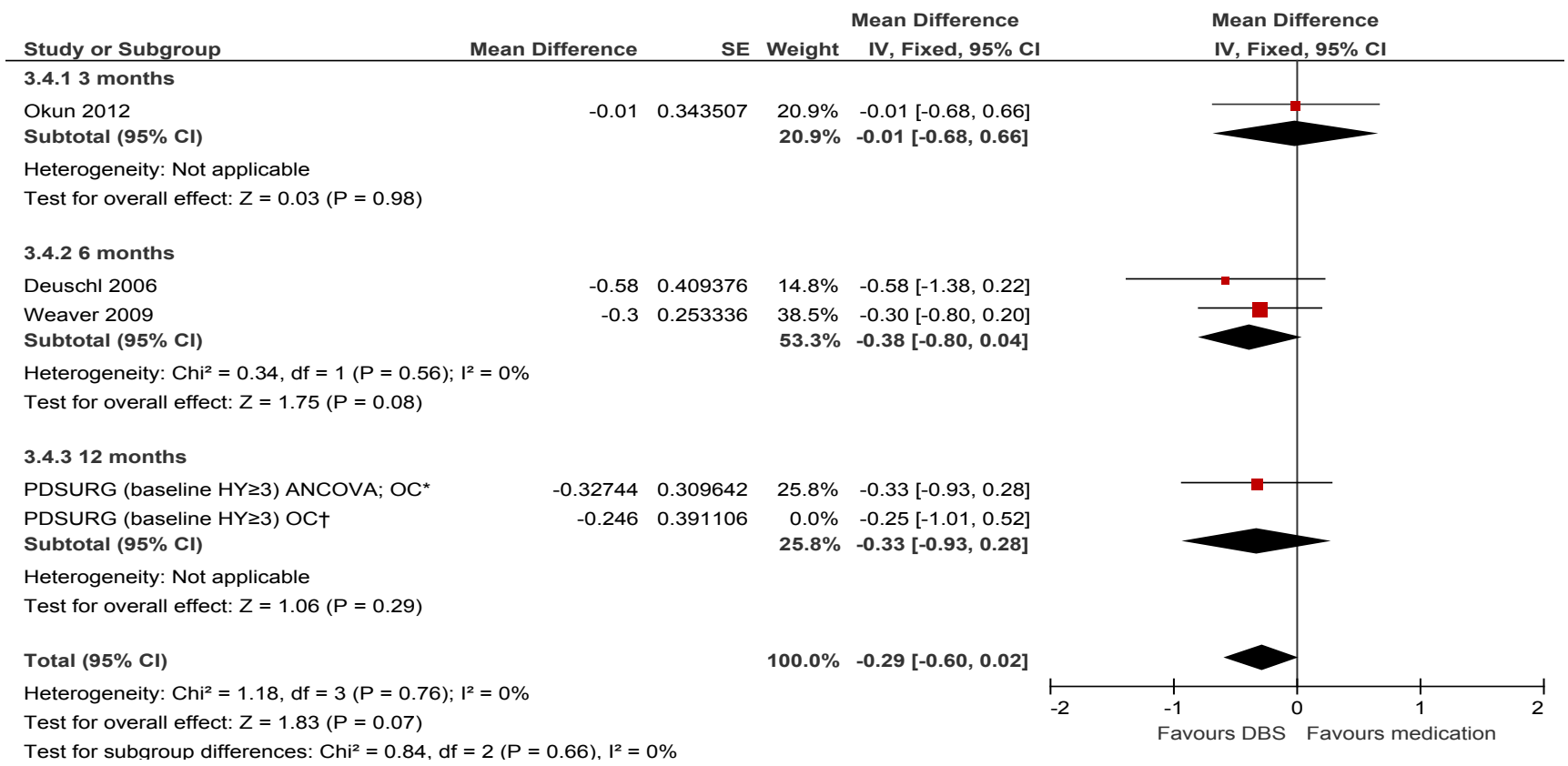


DBS -v- medication alone: mean daily 'on' time without troublesome dysinkesias – forest plot



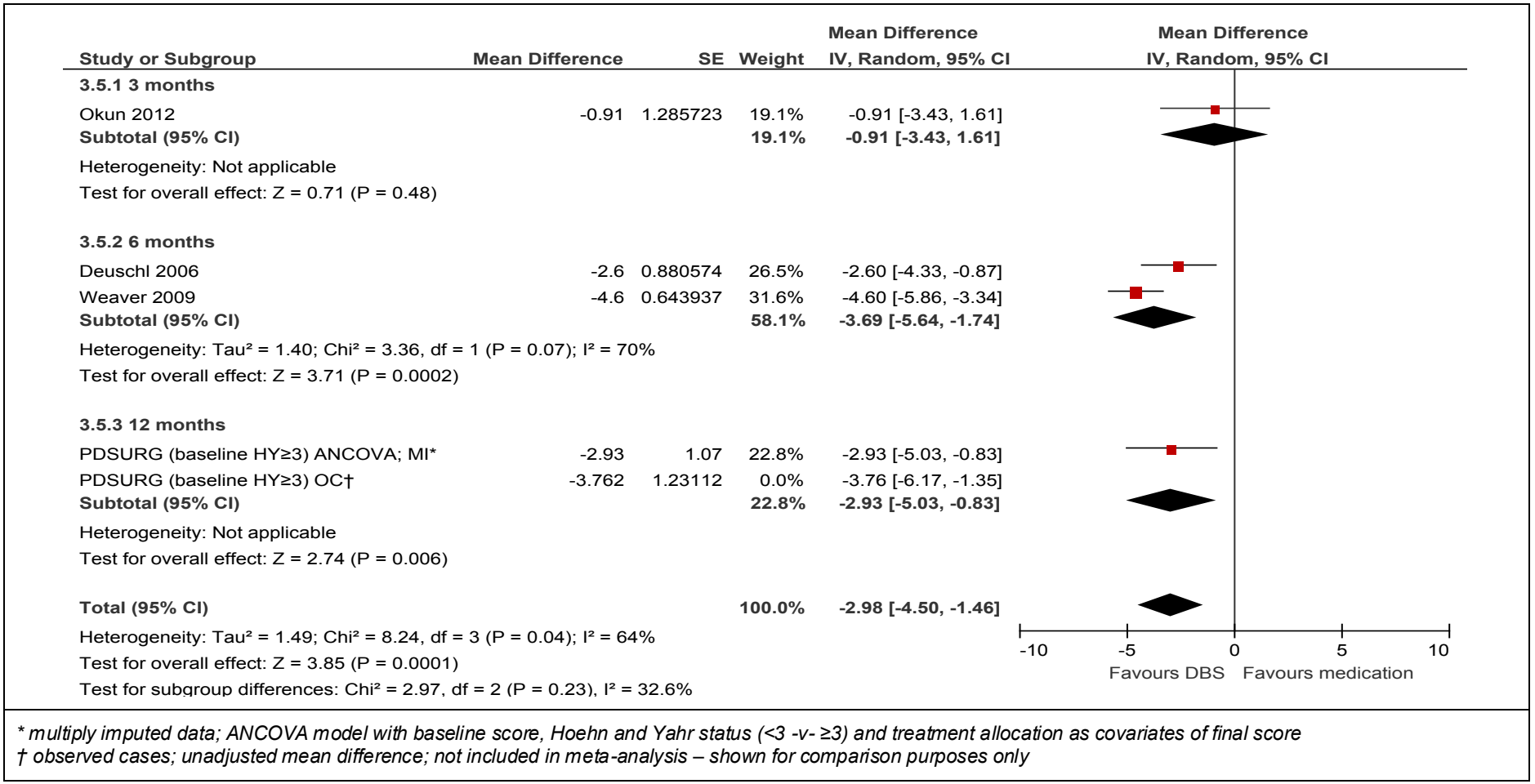
** multiply imputed data; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score*
† observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only

DBS -v- medication alone: mean daily 'off' time – forest plot



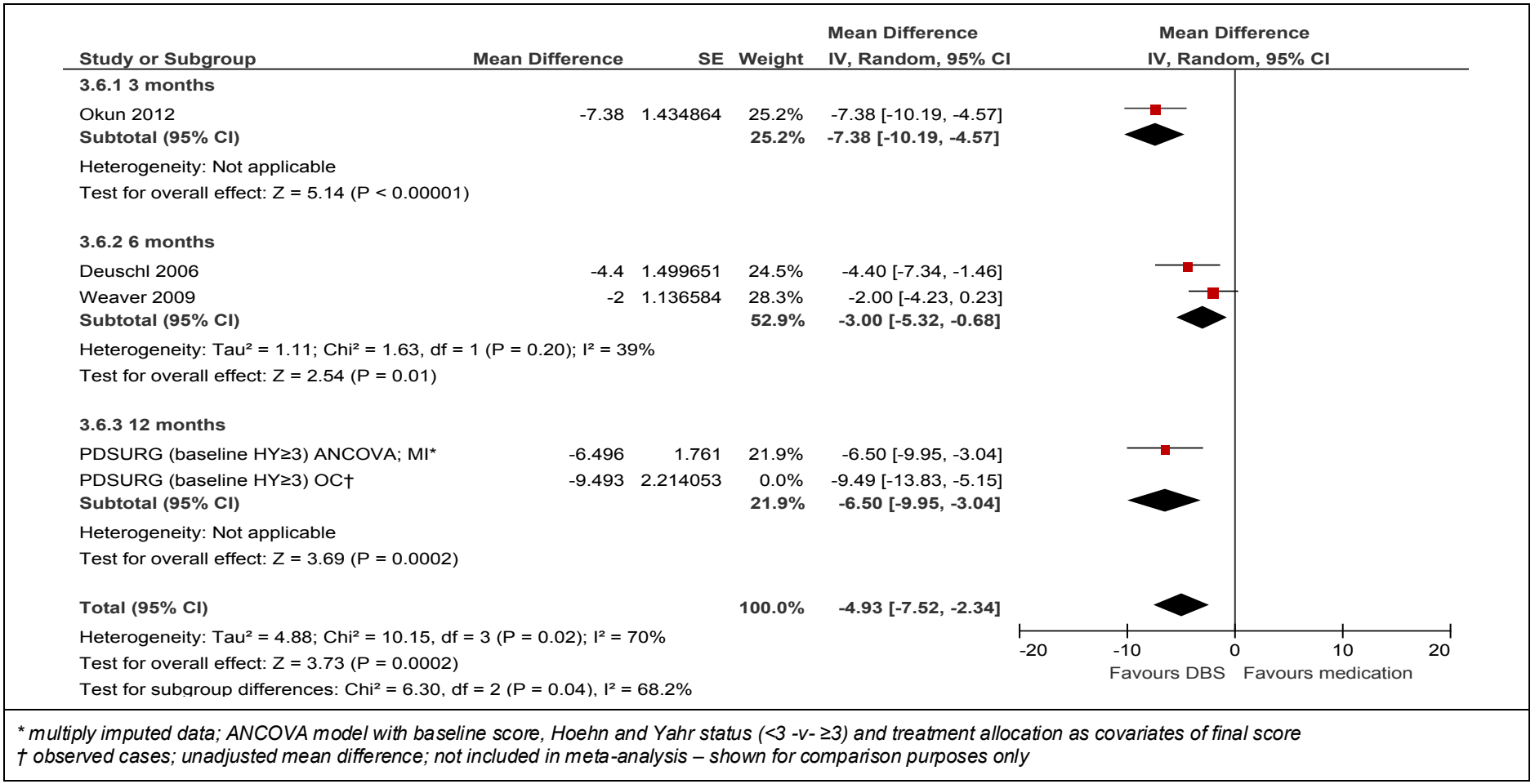
** observed cases; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score*
† observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only

DBS -v- medication alone: UPDRS I – forest plot



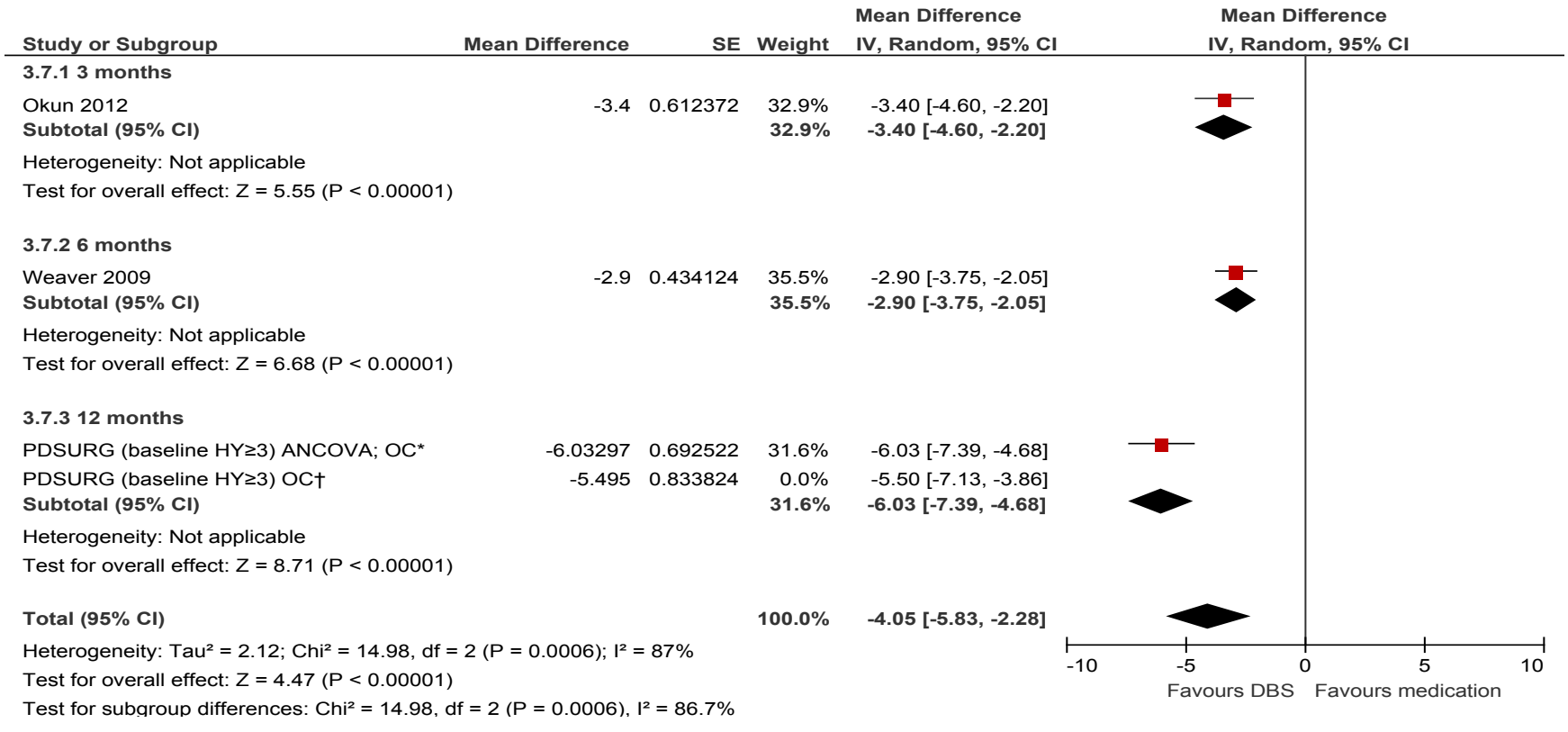
** multiply imputed data; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score*
† observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only

DBS -v- medication alone: UPDRS II (on) – forest plot



** multiply imputed data; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score*
† observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only

DBS -v- medication alone: UPDRS III (on) – forest plot



** observed cases; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score*
† observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only

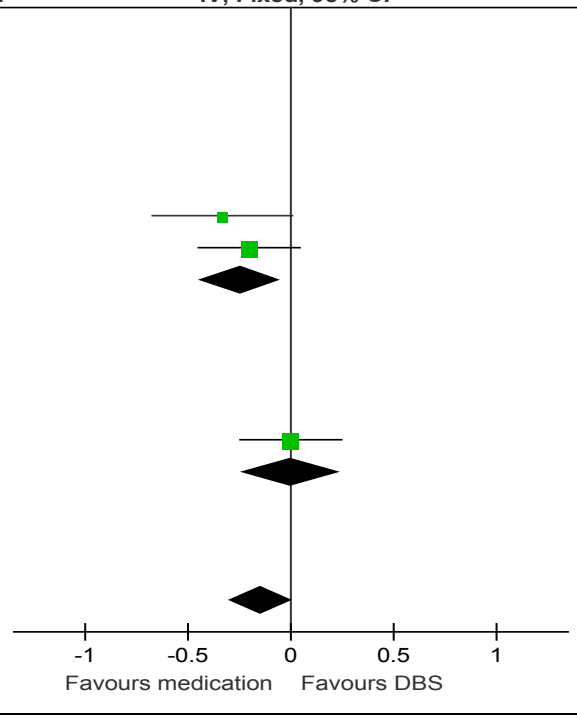
DBS -v- medication alone: UPDRS IV – forest plot

E.6.1.3 Neuropsychological outcomes

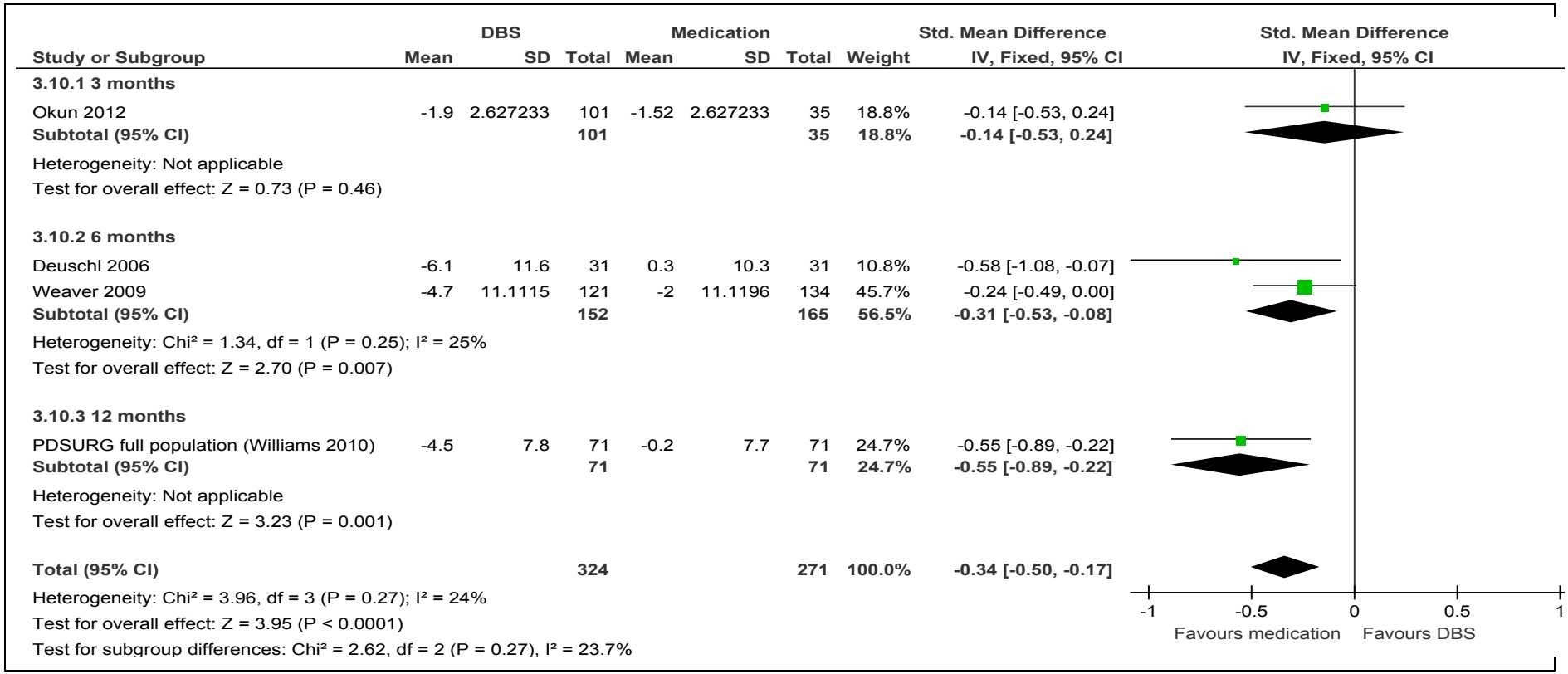
Quality assessment						Number of patients			
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	DBS	Control	Effect (95% CI)	Quality
Cognitive function (different measures pooled [standardised mean difference]) (higher is better); 6–12 months									
3[2,3,4]	RCT	No serious	Serious[5]	Serious[6]	Serious[7]	310	334	SMD = -0.16 (-0.34 to 0.03)	VERY LOW
Semantic fluency (higher is better); 3–12 months									
4[1,2,3,4]	RCT	No serious	No serious	Serious[6]	Serious[7]	324	271	SMD = -0.34 (-0.50 to -0.17)	LOW
Phonemic fluency (higher is better); 6–12 months									
3[2,3,4]	RCT	No serious	No serious	Serious[6]	No serious	222	235	SMD = -0.52 (-0.71 to -0.33)	MODERATE
Depression (different measures pooled [standardised mean difference]) (lower is better); 3–6 months									
3[1,2,3]	RCT	No serious	Serious[5]	Serious[6]	Very serious[8]	274	233	SMD = -0.17 (-0.58 to 0.25)	VERY LOW

1 *Okun 2012*
2 *Deuschl 2006 (semantic fluency and phonemic fluency reported for a subgroup of participants in Witt 2009)*
3 *Weaver 2009*
4 *Williams 2010 (main PDSURG publication [all participants regardless of HY score]; no subgroup data available for this outcome)*
5 *I^2 greater than 40% with no obvious explanation for heterogeneity*
6 *Some RCTs include a nontrivial proportion of participants with less advanced PD than the specified population for this question*
7 *At a 95% confidence level, data are consistent with appreciable harm and no meaningful effect*
8 *At a 95% confidence level, data are consistent with appreciable benefit, appreciable harm and no meaningful effect*

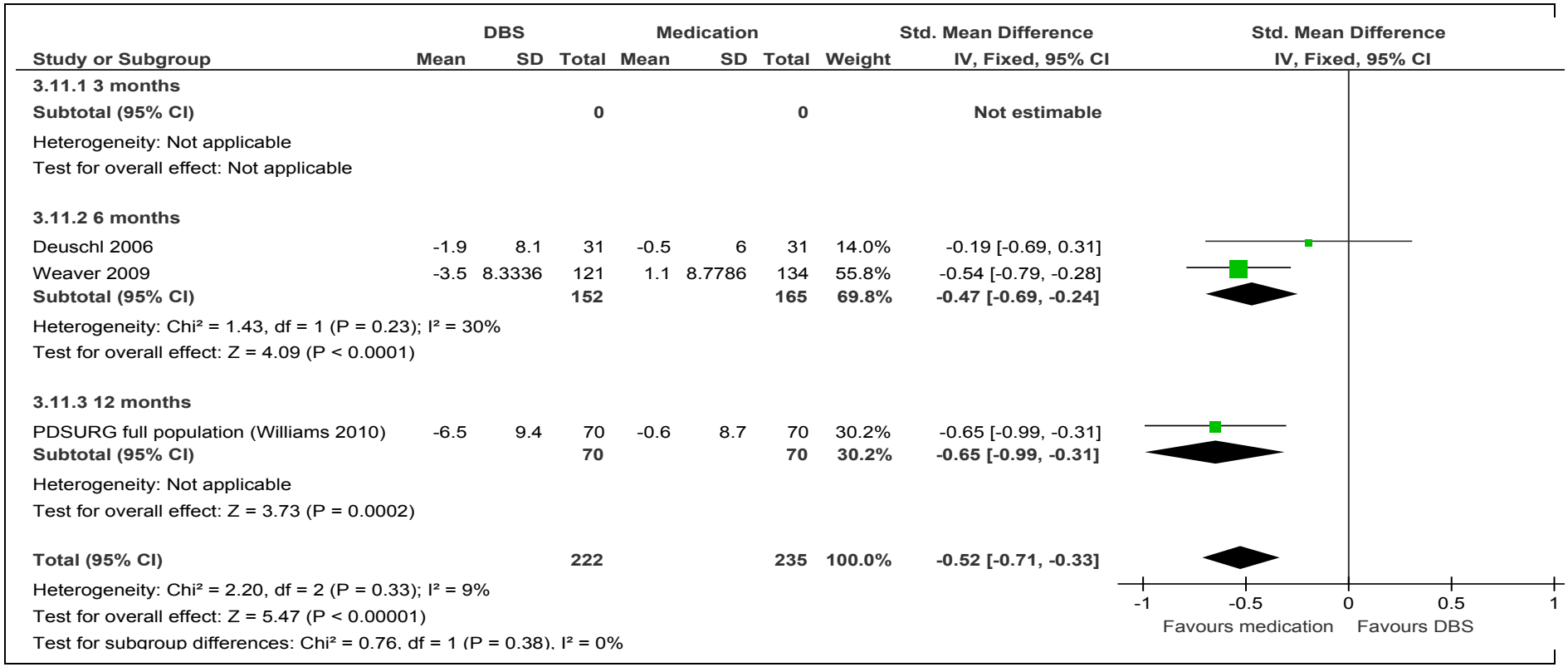
Study or Subgroup	DBS Mean	DBS SD	DBS Total	Medication Mean	Medication SD	Medication Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
3.9.1 3 months								
Subtotal (95% CI)			**0**			**0**		**Not estimable**
Heterogeneity: Not applicable								
Test for overall effect: Not applicable								
3.9.2 6 months								
Deuschl 2006	-2	4.9	68	-0.5	4	67	20.8%	-0.33 [-0.67, 0.01]
Weaver 2009	-0.2	6.1	121	0.9	4.7	134	39.6%	-0.20 [-0.45, 0.04]
Subtotal (95% CI)			**189**			**201**	**60.4%**	**-0.25 [-0.45, -0.05]**
Heterogeneity: Chi² = 0.37, df = 1 (P = 0.54); I² = 0%								
Test for overall effect: Z = 2.43 (P = 0.01)								
3.9.3 12 months								
PDSURG full population (Williams 2010)	-0.4	3.5	121	-0.4	2.9	133	39.6%	0.00 [-0.25, 0.25]
Subtotal (95% CI)			**121**			**133**	**39.6%**	**0.00 [-0.25, 0.25]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.00 (P = 1.00)								
Total (95% CI)			**310**			**334**	**100.0%**	**-0.15 [-0.30, 0.01]**
Heterogeneity: Chi² = 2.72, df = 2 (P = 0.26); I² = 26%								
Test for overall effect: Z = 1.89 (P = 0.06)								
Test for subgroup differences: Chi² = 2.35, df = 1 (P = 0.13), I² = 57.4%								



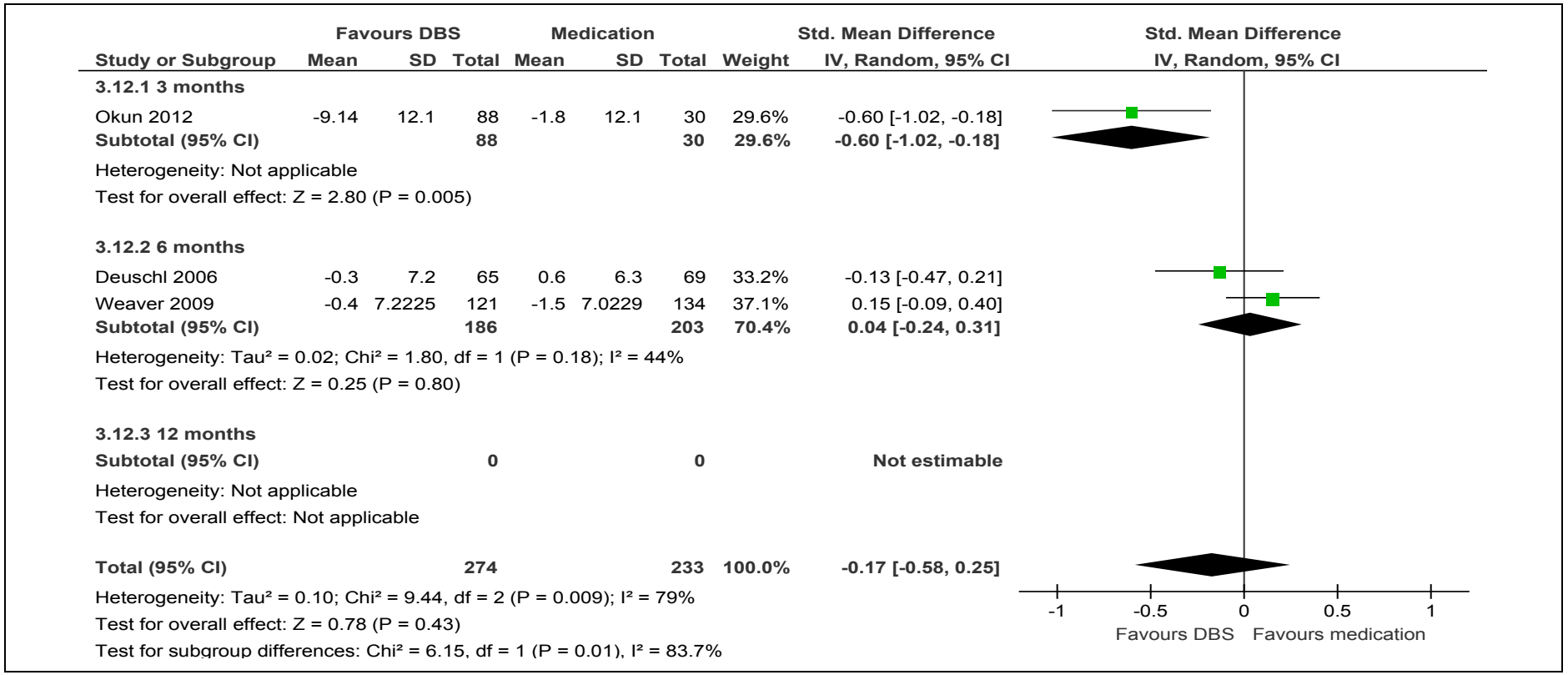
DBS -v- medication alone: cognitive function (different measures pooled [standardised mean difference]) – forest plot



DBS -v- medication alone: semantic fluency – forest plot



DBS -v- medication alone: phonemic fluency – forest plot

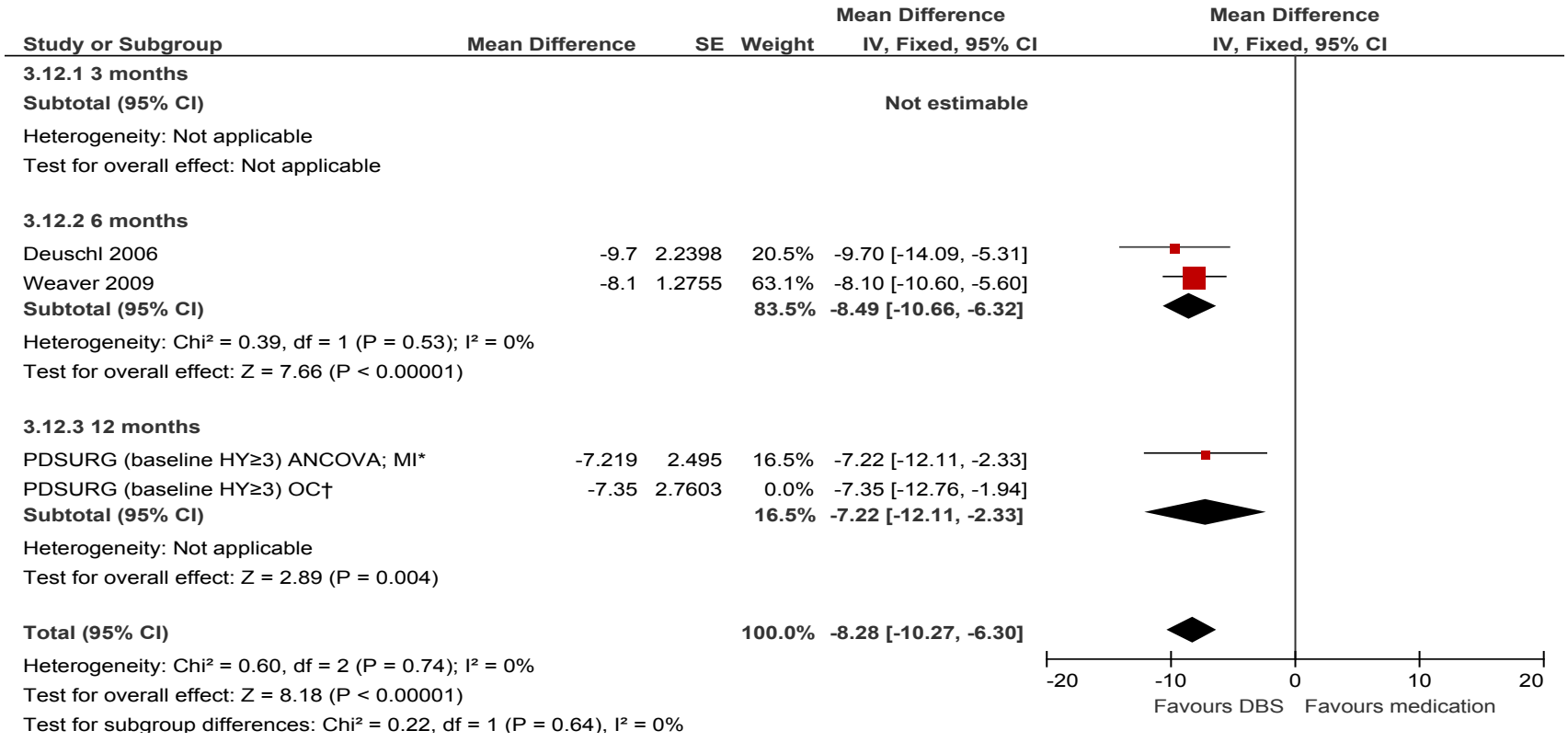


DBS -v- medication alone: depression (different measures pooled [standardised mean difference]) – forest plot

E.6.1.4 Health related quality of life – patient

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	DBS	Control	Mean difference (MD) (95% CI)	Quality
EQ-5D (higher is better); 12 months									
1[4]	RCT	No serious	No serious	No serious	No serious	50	50	0.123 (0.022 to 0.225)	HIGH
PDQ-39 (lower is better); 6–12 months									
3[2,3,4]	RCT	No serious	No serious	Serious[5]	No serious	243	258	-8.28 (-10.27 to -6.30)	MODERATE

1 *Okun 2012*
2 *Deuschl 2006*
3 *Weaver 2009*
4 *PDSURG multiply imputed data; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score used to estimate treatment effect in people with Hoehn and Yahr score ≥3 at baseline; calculated by guideline developers from patient-level data supplied by investigators (NB HY score ≥3 was a prespecified subgroup in the trial protocol and a randomisation stratification variable)*
5 *Some RCTs include a nontrivial proportion of participants with less advanced PD than the specified population for this question*



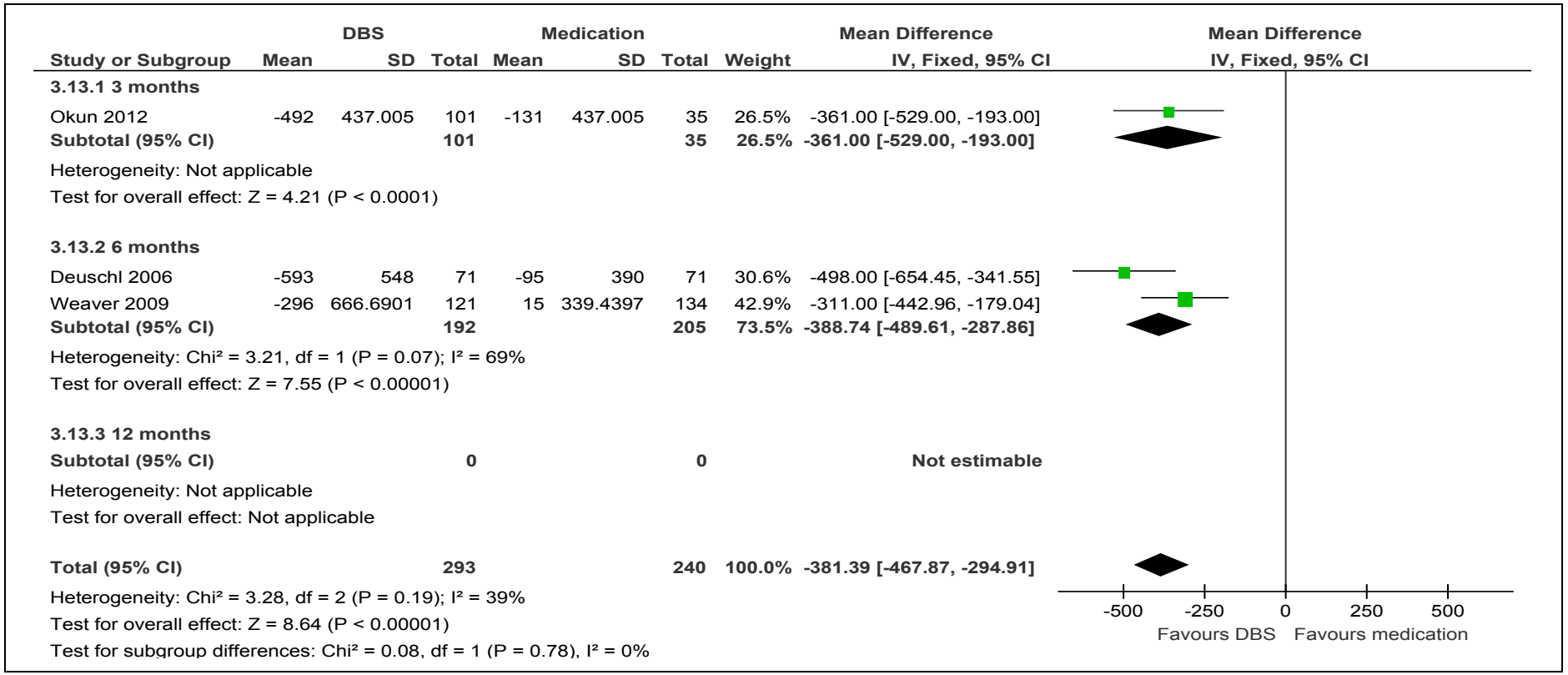
** multiply imputed data; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score*
† observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only

DBS -v- medication alone: PDQ-39 – forest plot

E.6.1.5 Medication load

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	DBS	Control	Mean difference (MD) (95% CI)	Quality
Daily dosage of anti-Parkinson's medication (levodopa mg equivalent) (lower is better); 3–6 months									
3[1,2,3]	RCT	No serious	No serious	Serious[5]	No serious	293	240	-381 (-468 to -295)	MODERATE

1 *Okun 2012*
2 *Deuschl 2006*
3 *Weaver 2009*
4 *Some RCTs include a nontrivial proportion of participants with less advanced PD than the specified population for this question*



DBS -v- medication alone: change in mean daily dose of anti-Parkinson's medication (levodopa mg equivalent) – forest plot

E.6.2 Levodopa–carbidopa intestinal gel compared with best medical treatment for advanced Parkinson's disease

E.6.2.1 Adverse events

No. of studies	Design	Quality assessment				No. of events / no. of patients		Effect (95% CI)	Quality
		Risk of bias	Inconsistency	Indirectness	Imprecision	LCIG	Placebo-LCIG		
Serious adverse events (probability of experiencing ≥1)									
Olanow et. al. (2014)	RCT	High[1]	NA[2]	Serious[3]	Serious[4]	5/37	7/34	RR = 0.66 (0.23 to 1.87)	VERY LOW
Any adverse events (probability of experiencing ≥1)									
Olanow et. al. (2014)	RCT	High[1]	NA[2]	Serious[3]	None	35/37	34/34	RR = 0.95 (0.86 to 1.04)	LOW
Device complications (probability of experiencing ≥1)									
Olanow et. al. (2014)	RCT	High[1]	NA[2]	Serious[3]	Serious[4]	34/37	29/34	RR = 1.08 (0.91 to 1.28)	VERY LOW
Falls (probability of experiencing ≥1)									
Olanow et. al. (2014)	RCT	High[1]	NA[2]	Serious[3]	Serious[4]	4/37	4/34	RR = 0.92 (0.25 to 3.39)	VERY LOW

[1] High risk of bias, due to device implantation in both trial arms
[2] NA: Not applicable as only 1 study contributed to this analysis
[3] Serious indirectness, due to device implantation in both trial arms
[4] At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit and no difference

E.6.2.2 Symptom severity

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	LCIG	Control	Mean difference (MD) (95% CI)	Quality
On time without dyskinesias (hrs, increase is good)									
Olanow et. al. (2014)	RCT	Low[1]	NA[2]	None[3]	None	35	31	2.28 (0.4 to 4.09)	HIGH
Off time per day (hrs, reduction is good)									
Olanow et. al. (2014)	RCT	Low[1]	NA[2]	None[3]	None	35	31	-1.91 (-3.03 to -0.79)	HIGH
UPDRS II (on) (lower is better)									
Olanow et. al. (2014)	RCT	Low[1]	NA[2]	None[3]	None	35	31	-3.00 (-5.16 to -0.84)	HIGH
UPDRS III (on) (lower is better)									
Olanow et. al. (2014)	RCT	Low[1]	NA[2]	None[3]	Serious[4]	35	31	1.40 (-2.72 to 5.52)	MODERATE
Clinical global impression of change score (lower is better)									
Olanow et. al. (2014)	RCT	Low[1]	NA[2]	None[3]	None	35	31	-0.7 (-1.4 to -0.1)	HIGH

[1] Low risk of bias, as assessed by NICE RCT quality checklist
[2] NA: Not applicable as only 1 study contributed to this analysis
[3] No serious indirectness; population was as described in review protocol
[4] At a 95% confidence level, data are consistent with appreciable harm, appreciable benefit and no difference

E.6.2.3 Health-related quality of life – patient

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	LCIG	Control	Mean difference (MD) (95% CI)	Quality
Generic health-related quality of life: EQ-5D									
Olanow et. al. (2014)	RCT	Low[1]	NA[2]	None[3]	Serious[4]	35	31	0.07 (-0.01 to 0.15)	MODERATE
Parkinson's disease-related quality of life: PDQ 39									
Olanow et. al. (2014)	RCT	Low[1]	NA[2]	None[3]	None	35	31	-7.00 (-12.49 to -1.51)	HIGH

[1] Low risk of bias, as assessed by NICE RCT quality checklist
[2] NA: Not applicable as only one study contributed to this analysis
[3] No serious indirectness; population was as described in review protocol
[4] At a 95% confidence level, data are consistent with appreciable benefit and no difference

E.6.2.4 Health-related quality of life – carer

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	LCIG	Control	Mean difference (MD) (95% CI)	Quality
Zarit carer burden interview									
Olanow et. al. (2014)	RCT	Low[1]	NA[2]	None[3]	Serious[4]	35	31	-4.5 (-10.58 to 1.58)	MODERATE

[1] Low risk of bias, as assessed by NICE RCT quality checklist
[2] NA: Not applicable as only one study contributed to this analysis
[3] No serious indirectness; population was as described in review protocol
[4] At a 95% confidence level, data are consistent with appreciable benefit and no difference

E.6.2.5 Medication load

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	LCIG	Control	Mean difference (MD) (95% CI)	Quality
Levodopa daily dosage (mg)									
Olanow et. al. (2014)	RCT	Low[1]	NA[2]	None[3]	Serious[4]	35	31	-158.0 (-324.5 to 8.5)	MODERATE

[1] Low risk of bias, as assessed by NICE RCT quality checklist
[2] NA: Not applicable as only one study contributed to this analysis
[3] No serious indirectness; population was as described in review protocol
[4] At a 95% confidence level, data are consistent with appreciable benefit and no difference

E.6.3 Indirect comparison of DBS and LCIG

E.6.3.1 Symptom severity

Comparison	Studies	Timepoint	Pairwise data	Direct evidence		Indirect evidence	
				Effect measure (95%CI)	Quality of evidence	Effect measure (95%CI)	Quality of evidence
UPDRS II (lower is better)							
DBS (n=45) -v- BMT (n=47)	PDSURG (HY≥3)[6]	52wk	E.6.1.2	-2.92 (-5.02 to -0.82)	HIGH	–	–
LCIG (n=35) -v- BMT (n=31)	Olanow et al. (2014)	12wk	E.6.2.2	-3.00 (-5.16 to -0.84)	HIGH	–	–
DBS -v- LCIG	–	52wk[1]	–	–	–	0.08 (-3.14 to 3.29)	LOW[2,3]
UPDRS III (lower is better)							
DBS (n=40) -v- BMT (n=38)	PDSURG (HY≥3)[6]	52wk	E.6.1.2	-6.48 (-9.93 to -3.03)	HIGH	–	–
LCIG (n=35) -v- BMT (n=31)	Olanow et al. (2014)	12wk	E.6.2.2	1.40 (-2.72 to 5.52)	MODERATE[4]	–	–
DBS -v- LCIG	–	52wk[1]	–	–	–	-7.88 (-13.63 to -2.14)	MODERATE[2]
Off time (lower is better)							
DBS (n=48) -v- BMT (n=51)	PDSURG (HY≥3)[6]	52wk	E.6.1.2	-2.62 (-3.65 to -1.60)	MODERATE[5]	–	–
LCIG (n=35) -v- BMT (n=31)	Olanow et al. (2014)	12wk	E.6.2.2	-1.91 (-3.03 to -0.79)	HIGH	–	–
DBS -v- LCIG	–	52wk[1]	–	–	–	-0.71 (-2.29, 0.87)	VERY LOW[2,3,5]

[1] *Incorporating increased uncertainty for LCIG -v- BMT due to unknown 'drift' from 12wk to 52wk timepoints (parameterised using Fernandez et al. 2015)*
[2] *Downgraded for indirectness (12wk estimate used to estimate 52wk effects)*
[3] *Downgraded for imprecision (at a 95% confidence level, data are consistent with appreciable benefit with DBS, appreciable benefit with LCIG and no meaningful difference)*
[4] *Downgraded for imprecision*
[5] *Downgraded for indirectness (off time estimate approximated from answer to UPDRS Q39 [categorical proportion of waking day spent 'off'])*
[6] *PDSURG multiply imputed data; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score used to estimate treatment effect in people with Hoehn and Yahr score ≥3 at baseline; calculated by guideline developers from patient-level data supplied by investigators (NB HY score ≥3 was a prespecified subgroup in the trial protocol and a randomisation stratification variable)*

E.6.3.2 Health-related quality of life – patient

Comparison	Studies	Timepoint	Pairwise data	Direct evidence		Indirect evidence	
				Effect measure (95%CI)	Quality of evidence	Effect measure (95%CI)	Quality of evidence
EQ-5D (higher is better)							
DBS (n=50) -v- BMT (n=50)	PDSURG (HY≥3)[5]	52wk	E.6.1.4	0.12 (0.02 to 0.22)	HIGH	–	–
LCIG (n=35) -v- BMT (n=31)	Olanow et al. (2014)	12wk	E.6.2.3	0.07 (-0.01 to 0.15)	MODERATE[4]	–	–
DBS -v- LCIG	–	52wk[1]	–	–	–	0.05 (-0.08 to 0.19)	LOW[2,3]
PDQ-39 (lower is better)							
DBS (n=51) -v- BMT (n=51)	PDSURG (HY≥3)[5]	52wk	E.6.1.4	-7.21 (-12.10 to -2.32)	HIGH	–	–
LCIG (n=35) -v- BMT (n=31)	Olanow et al. (2014)	12wk	E.6.2.3	-7.00 (-12.49 to -1.51)	HIGH	–	–
DBS -v- LCIG	–	52wk[1]	–	–	–	-0.21 (-7.92 to 7.50)	LOW[2,3]

[1] *Incorporating increased uncertainty for LCIG -v- BMT due to unknown 'drift' from 12wk to 52wk timepoints (parameterised using Fernandez et al. 2015)*
[2] *Downgraded for indirectness (12wk estimate used to estimate 52wk effects)*
[3] *Downgraded for imprecision (at a 95% confidence level, data are consistent with appreciable benefit with DBS, appreciable benefit with LCIG and no meaningful difference)*
[4] *Downgraded for imprecision*
[5] *PDSURG multiply imputed data; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score used to estimate treatment effect in people with Hoehn and Yahr score ≥3 at baseline; calculated by guideline developers from patient-level data supplied by investigators (NB HY score ≥3 was a prespecified subgroup in the trial protocol and a randomisation stratification variable)*

E.6.4 Deep brain stimulation compared with best medical treatment for earlier Parkinson's disease

E.6.4.1 Adverse events

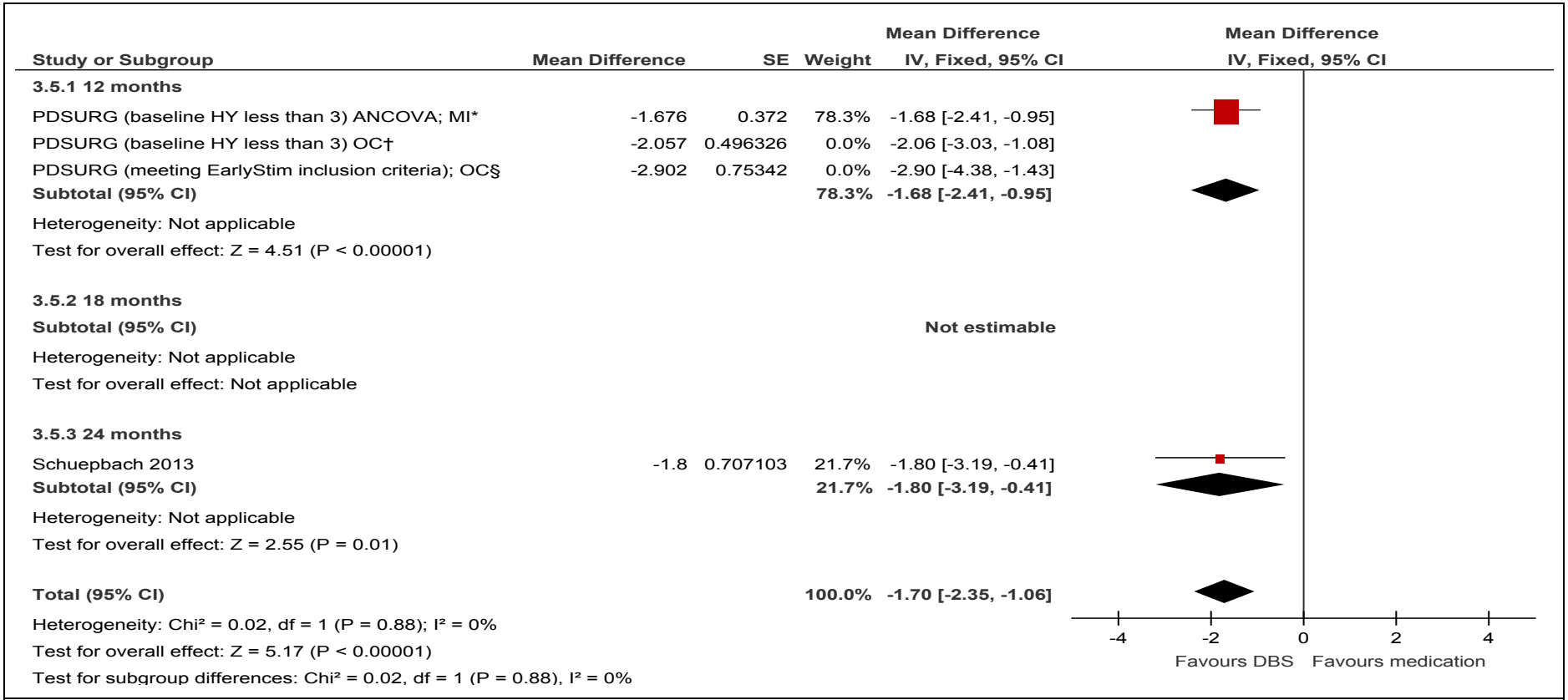
No. of studies	Design	Quality assessment				No. of events / no. of patients or patient-years		Effect (95% CI)	Quality
		Risk of bias	Incons-istency	Indirectness	Imprecision	DBS	Control		
Serious adverse events (probability of experiencing ≥1); 24 months									
1[2]	RCT	No serious	N/A	Not serious	Serious[3]	68/124	56/127	RR = 1.24 (0.97 to 1.60)	MODERATE
Serious adverse events (rate per patient-year); 24 months									
1[2]	RCT	No serious	N/A	Not serious	Serious[4]	123 per 246pt-yrs[5]	128 per 249pt-yrs[5]	IRR = 0.97 (0.76 to 1.25)	MODERATE
Falls (probability of experiencing ≥1); 24 months									
1[2]	RCT	No serious	N/A	Not serious	Serious[4]	8/124	5/127	RR = 1.64 (0.55 to 4.87)	MODERATE
Falls (rate per patient-year); 24 months									
1[2]	RCT	No serious	N/A	Not serious	Serious[4]	11 per 246pt-yrs[5]	5 per 249pt-yrs[5]	IRR = 2.23 (0.77 to 6.41)	MODERATE

1 *Schüpbach 2007*
2 *Schüpbach 2013*
3 *at a 95% confidence level, data are consistent with appreciable harm and no meaningful effect*
4 *at a 95% confidence level, data are consistent with appreciable benefit, appreciable harm and no meaningful effect*
5 *assuming dropouts withdrew at 1 year (i.e. halfway through 2-year follow-up)*

E.6.4.2 Symptom severity

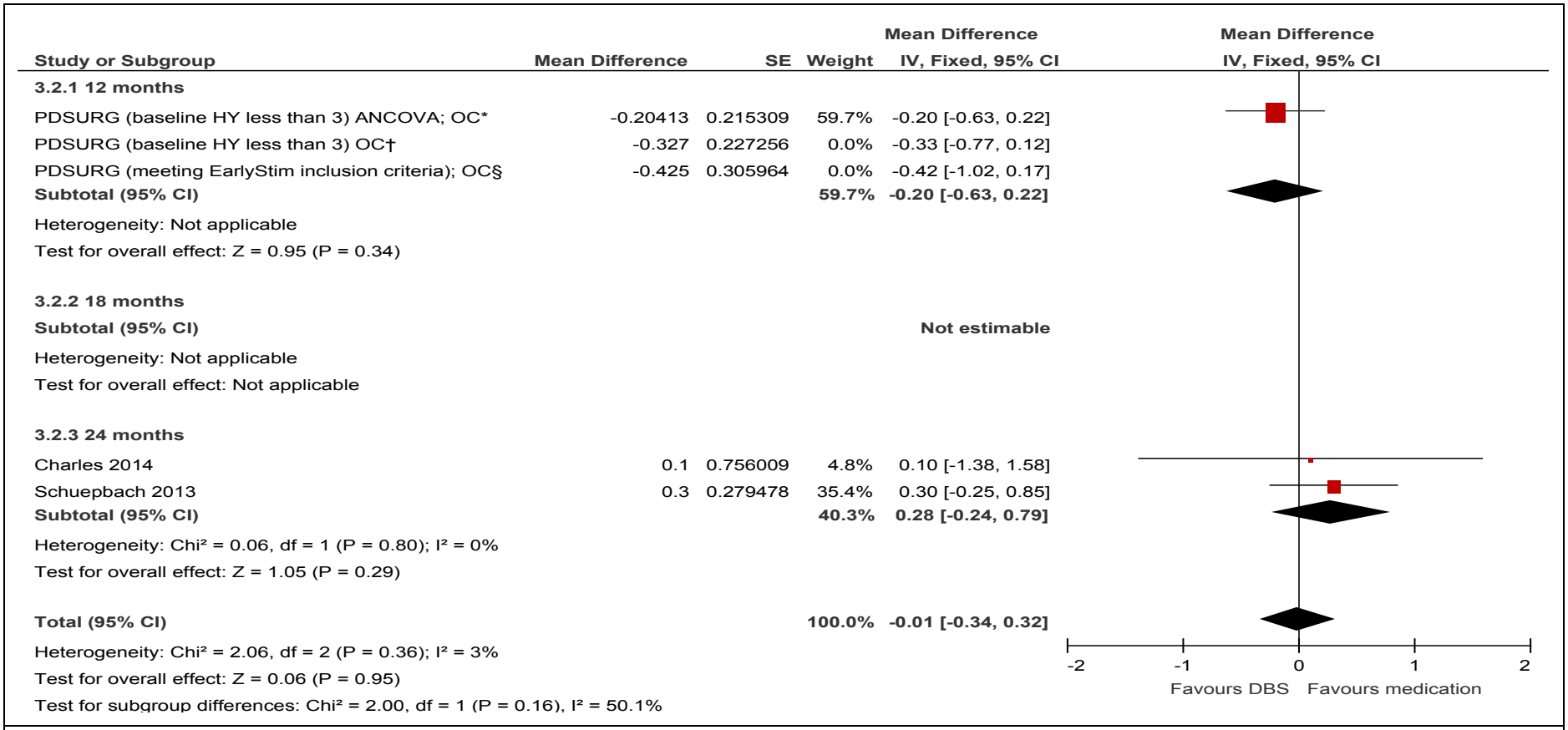
Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	DBS	Control	Mean difference (MD) (95% CI)	Quality
Hoehn and Yahr score (off medication) (lower is better); 3–12 months									
1[4]	RCT	No serious	N/A	No serious	No serious	85	95	-0.32 (-0.56 to -0.09)	HIGH
Daily 'on' time without troublesome dysinkesias (higher is better); 24 months									
1[2]	RCT	No serious	N/A	No serious	No serious	105	110	1.90 (0.51 to 3.29)	HIGH
Daily 'off' time (lower is better); 12–24 months									
2[2,3]	RCT	No serious	No serious	No serious	No serious	209	212	-1.70 (-2.35 to -1.06)	HIGH
UPDRS I (lower is better); 12–24 months									
3[2,4,5]	RCT	No serious	No serious	No serious	Serious[7]	233	225	-0.01 (-0.34 to 0.32)	MODERATE
UPDRS II on (lower is better); 12–24 months									
4[1,2,3,5]	RCT	No serious	No serious	No serious	Serious[7]	246	244	0.48 (-0.40 to 1.37)	MODERATE
UPDRS III on (lower is better); 12–24 months									
4[1,2,3,5]	RCT	No serious	No serious	No serious	No serious	243	241	-3.21 (-4.49 to -1.93)	HIGH
UPDRS IV (lower is better); 12–24 months									
4[1,2,4,5]	RCT	No serious	Serious[6]	No serious	No serious	214	212	-4.68 (-6.75 to -2.61)	MODERATE

1 *Schüpbach 2007*
2 *Schüpbach 2013*
3 *PDSURG (subgroup with baseline HY<3); multiply imputed data; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score; calculated by guideline developers from patient-level data supplied by investigators*
4 *PDSURG (subgroup with baseline HY<3); observed cases; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score; calculated by guideline developers from patient-level data supplied by investigators*
5 *Charles 2014*
6 *I^2 greater than 40% with no obvious explanation for heterogeneity*
7 *at a 95% confidence level, data are consistent with appreciable benefit and no effect*



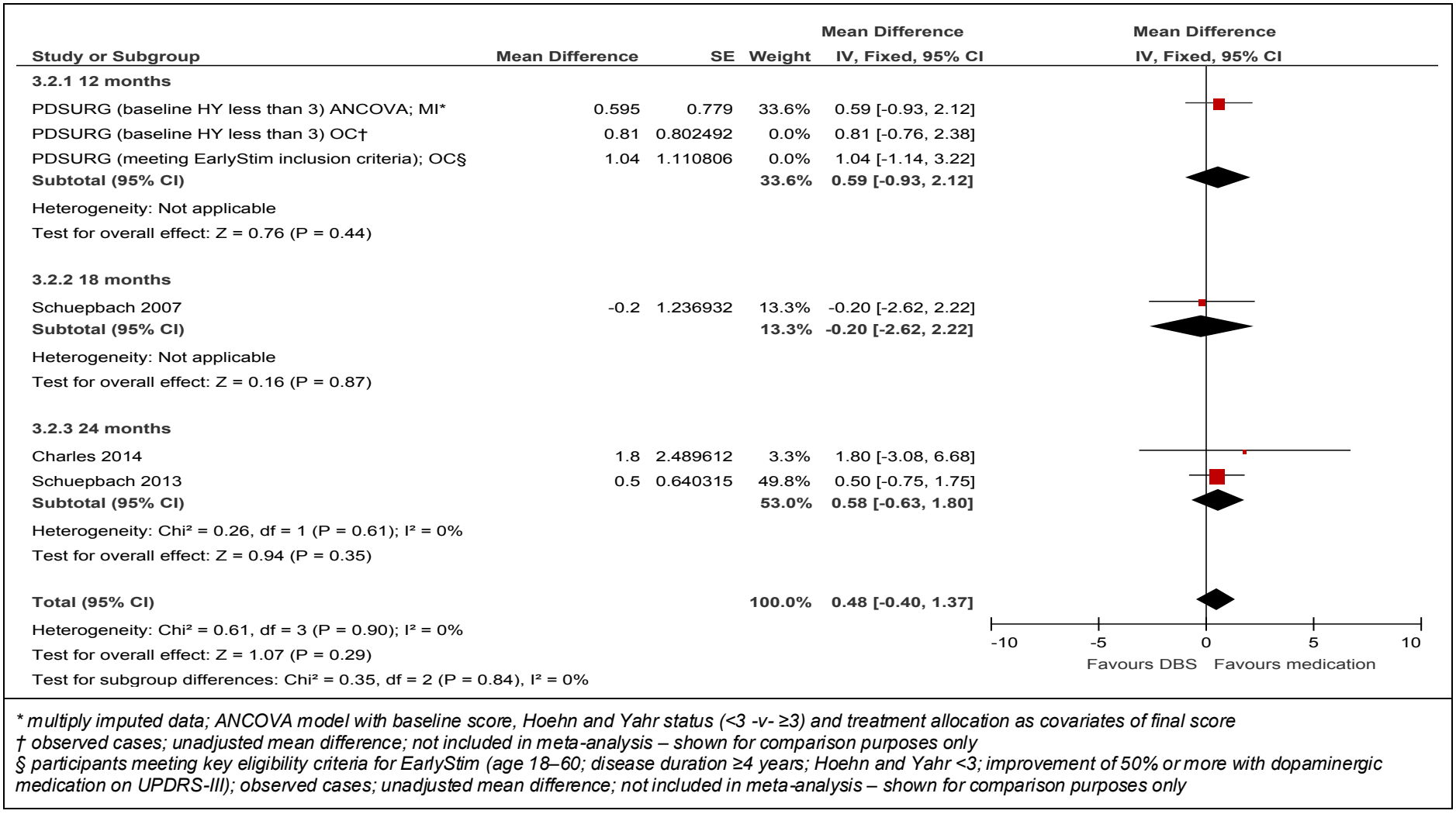
** multiply imputed data; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score*
† observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only
§ participants meeting key eligibility criteria for EarlyStim (age 18–60; disease duration ≥4 years; Hoehn and Yahr <3; improvement of 50% or more with dopaminergic medication on UPDRS-III); observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only

DBS -v- medication alone: mean daily 'off' time – forest plot



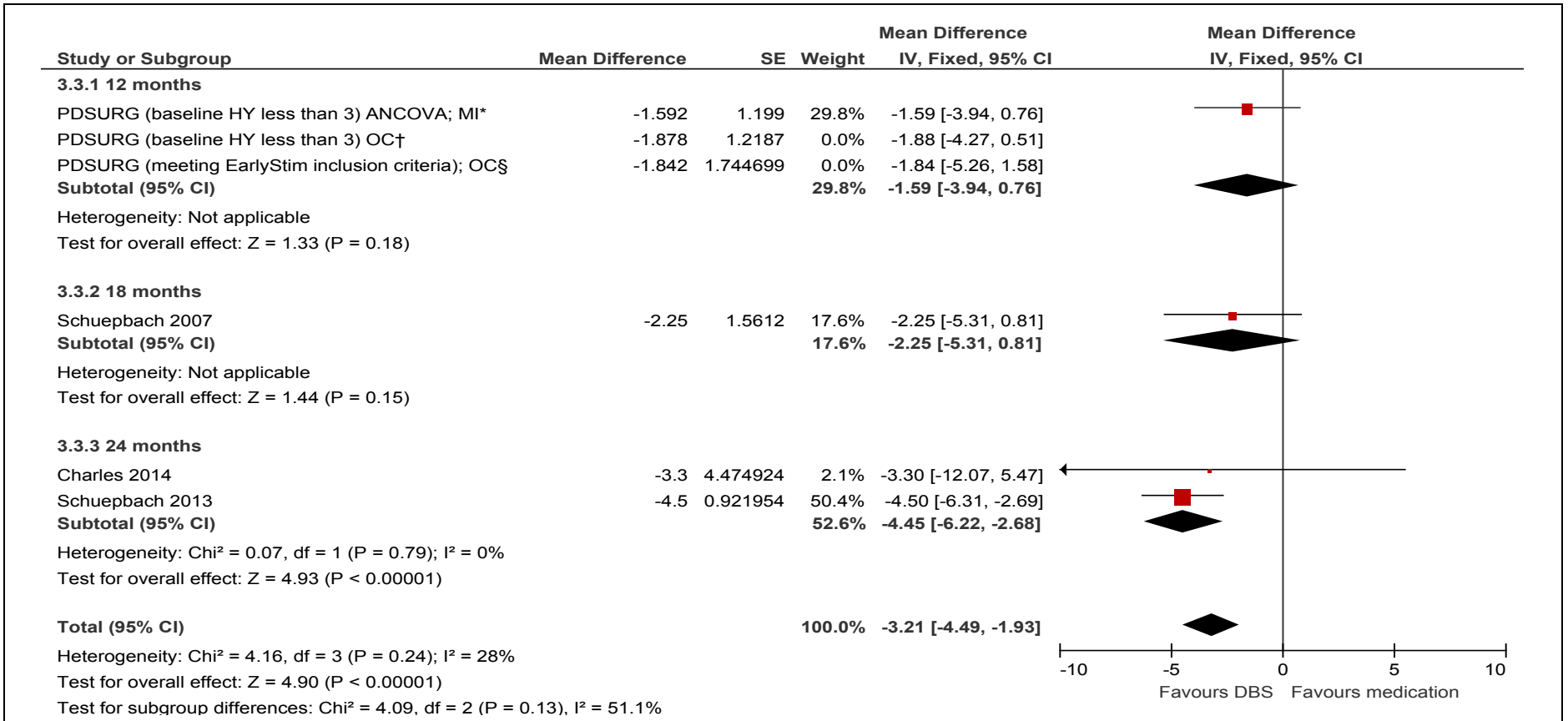
** observed cases; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score*
† observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only
§ participants meeting key eligibility criteria for EarlyStim (age 18–60; disease duration ≥4 years; Hoehn and Yahr <3; improvement of 50% or more with dopaminergic medication on UPDRS-III); observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only

DBS -v- medication alone: UPDRS I – forest plot



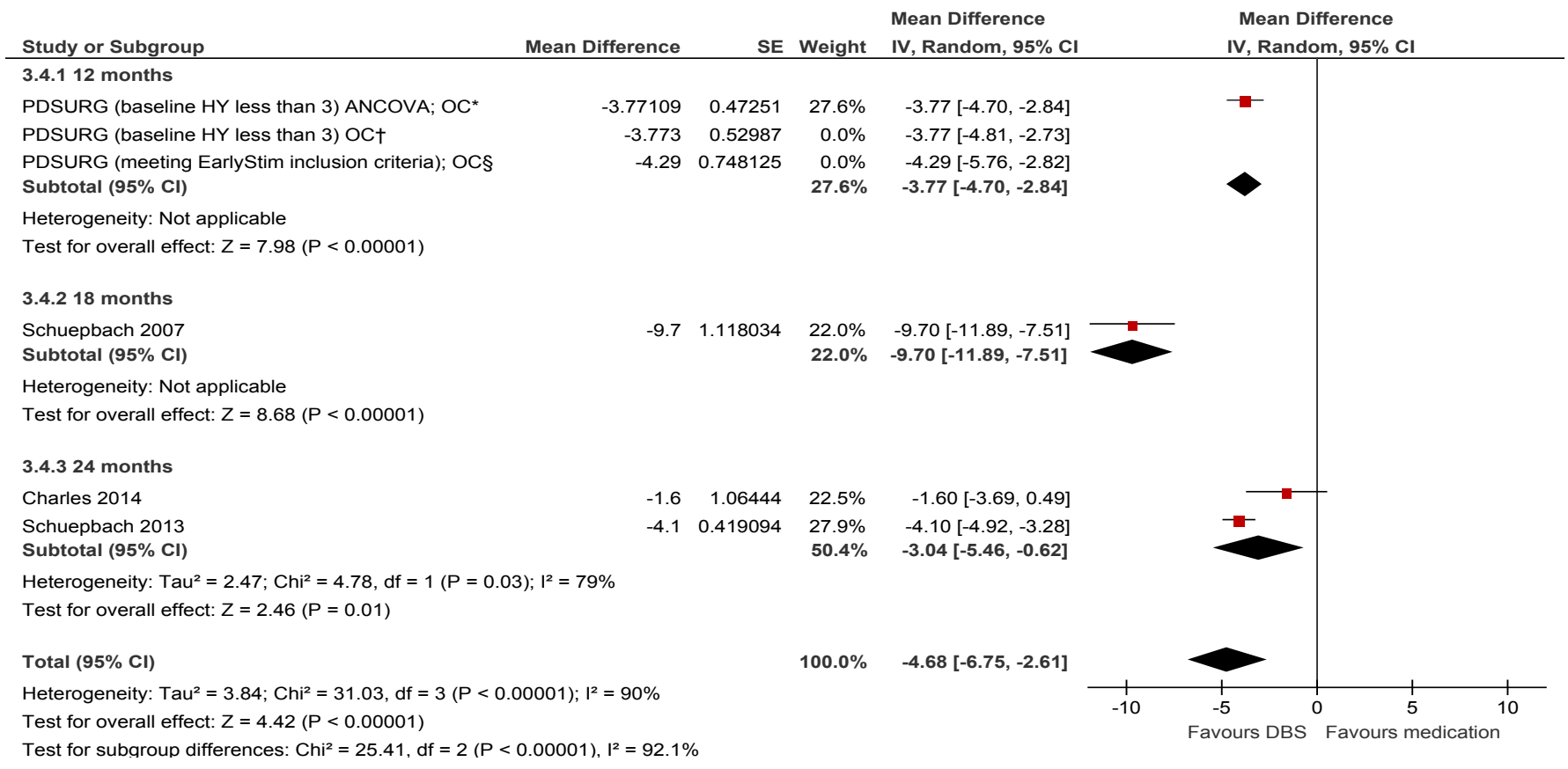
** multiply imputed data; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score*
† observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only
§ participants meeting key eligibility criteria for EarlyStim (age 18–60; disease duration ≥4 years; Hoehn and Yahr <3; improvement of 50% or more with dopaminergic medication on UPDRS-III); observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only

DBS -v- medication alone: UPDRS II (on) – forest plot



** multiply imputed data; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score*
† observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only
§ participants meeting key eligibility criteria for EarlyStim (age 18–60; disease duration ≥4 years; Hoehn and Yahr <3; improvement of 50% or more with dopaminergic medication on UPDRS-III); observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only

DBS -v- medication alone: UPDRS III (on) – forest plot



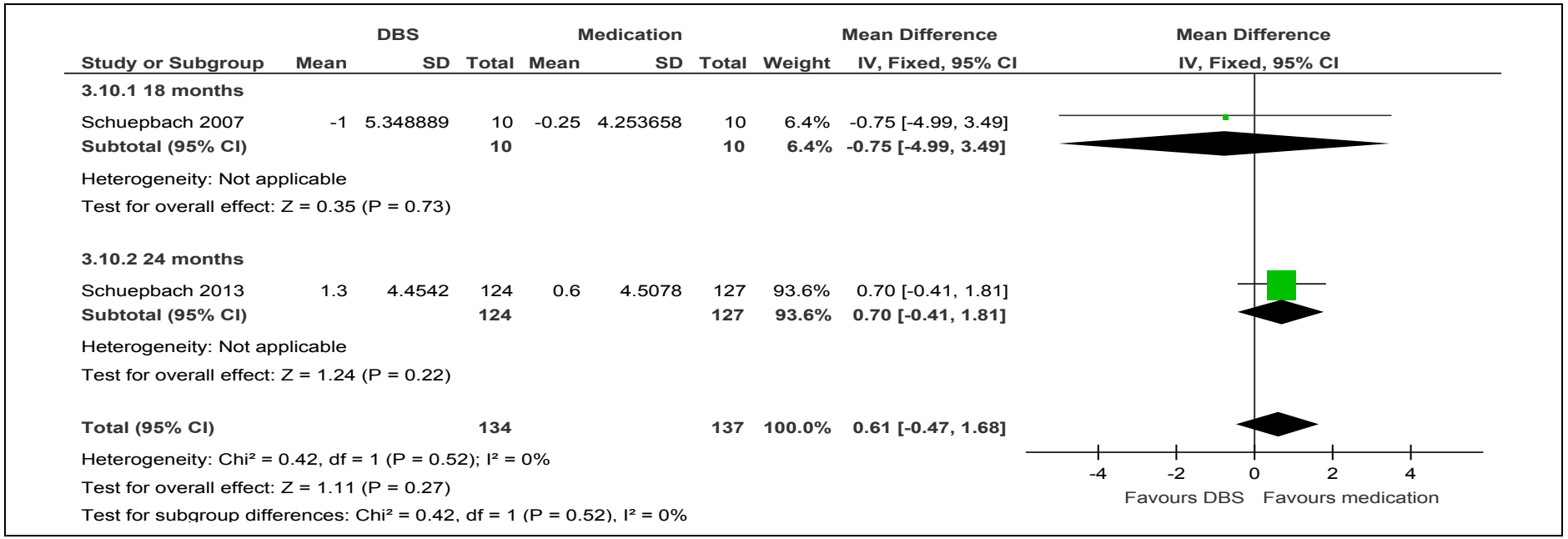
** observed cases; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score*
† observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only
§ participants meeting key eligibility criteria for EarlyStim (age 18–60; disease duration ≥4 years; Hoehn and Yahr <3; improvement of 50% or more with dopaminergic medication on UPDRS-III); observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only

DBS -v- medication alone: UPDRS IV – forest plot

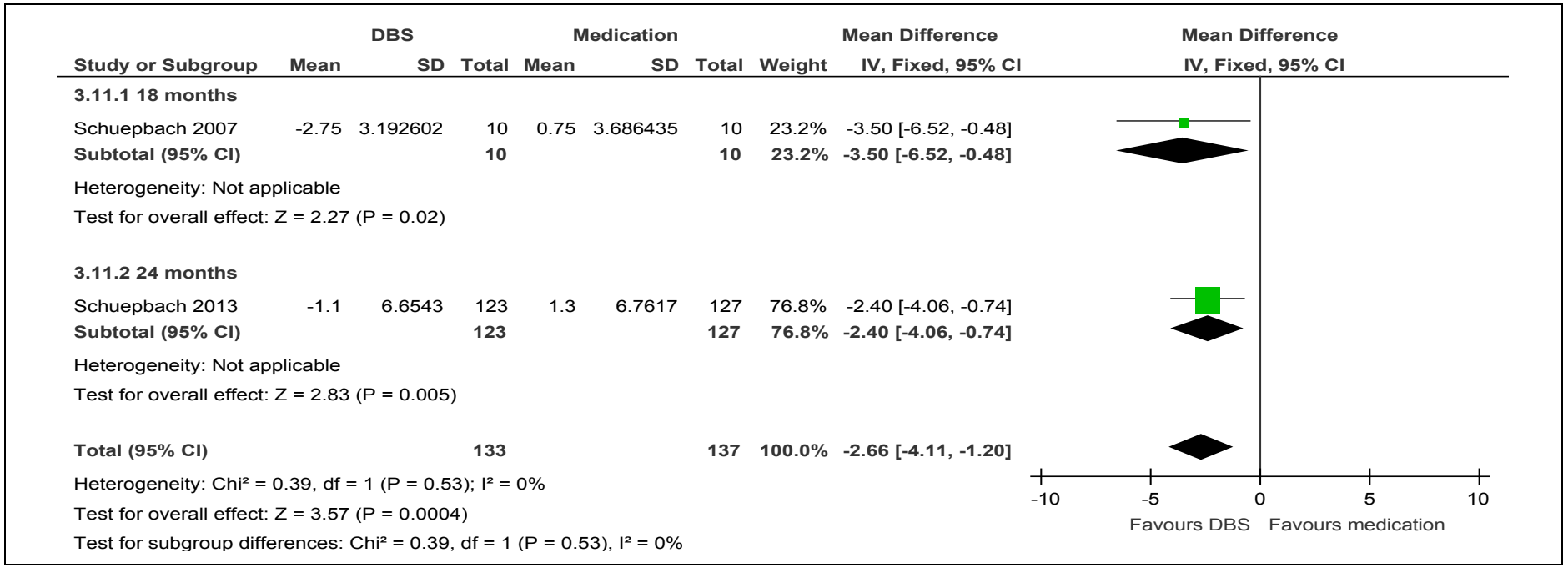
E.6.4.3 Neuropsychological outcomes

Quality assessment						Number of patients		Mean difference (MD) (95% CI)	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	DBS	Control		
Cognitive function (MDRS) (higher is better); 18–24 months									
2[1,2]	RCT	No serious	Not serious	Not serious	Serious[3]	134	137	0.61 (-0.47 to 1.68)	MODERATE
Depression (Montgomery–Åsberg depression scale) (lower is better); 18–24 months									
2[1,2]	RCT	No serious	Not serious	Not serious	Not serious	133	137	-2.66 (-4.11 to -1.20)	HIGH

1 *Schüpbach 2007*
2 *Schüpbach 2013*
3 *at a 95% confidence level, data are consistent with appreciable benefit, appreciable harm and no meaningful effect*



DBS -v- medication alone: cognitive function (MDRS) – forest plot



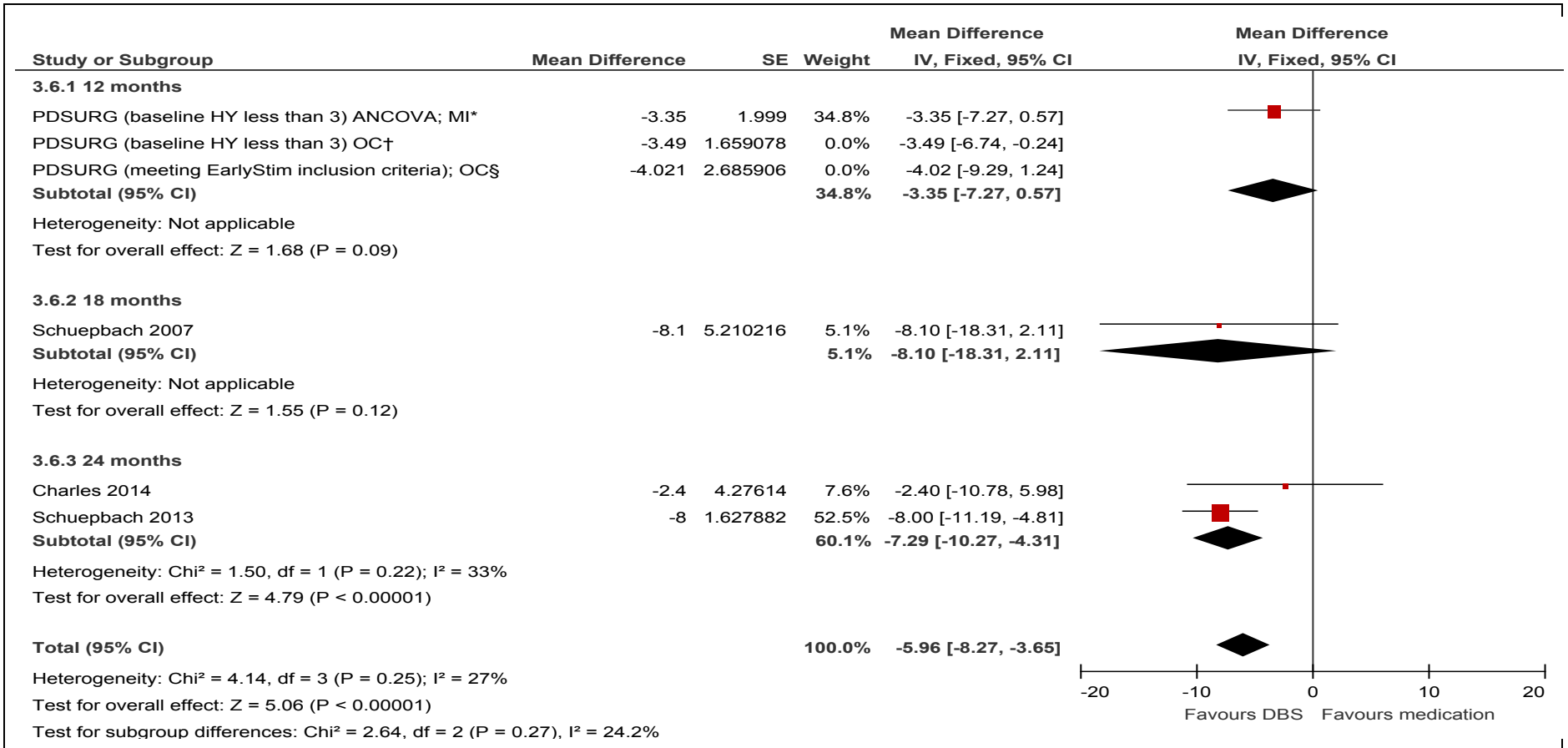
DBS -v- medication alone: depression (MADRS) – forest plot

E.6.4.4 Health related quality of life – patient

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	DBS	Control	Mean difference (MD) (95% CI)	Quality
EQ-5D (higher is better); 12 months									
1[3]	RCT	No serious	NA	No serious	Very serious[4]	104	99	0.00 (-0.08 to 0.07)	LOW
PDQ-39 (lower is better); 12–24 months									
4[1,2,3,5]	RCT	No serious	No serious	No serious	No serious	306	288	-5.96 (-8.27 to -3.65)	HIGH

1 *Schüpbach 2007*
2 *Schüpbach 2013*
3 *PDSURG (subgroup with baseline HY<3); multiply imputed data; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score; calculated by guideline developers from patient-level data supplied by investigators*
4 *at a 95% confidence level, data are consistent with appreciable benefit, appreciable harm and no meaningful effect*

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	DBS	Control	Mean difference (MD) (95% CI)	Quality
5	*Charles 2014*								



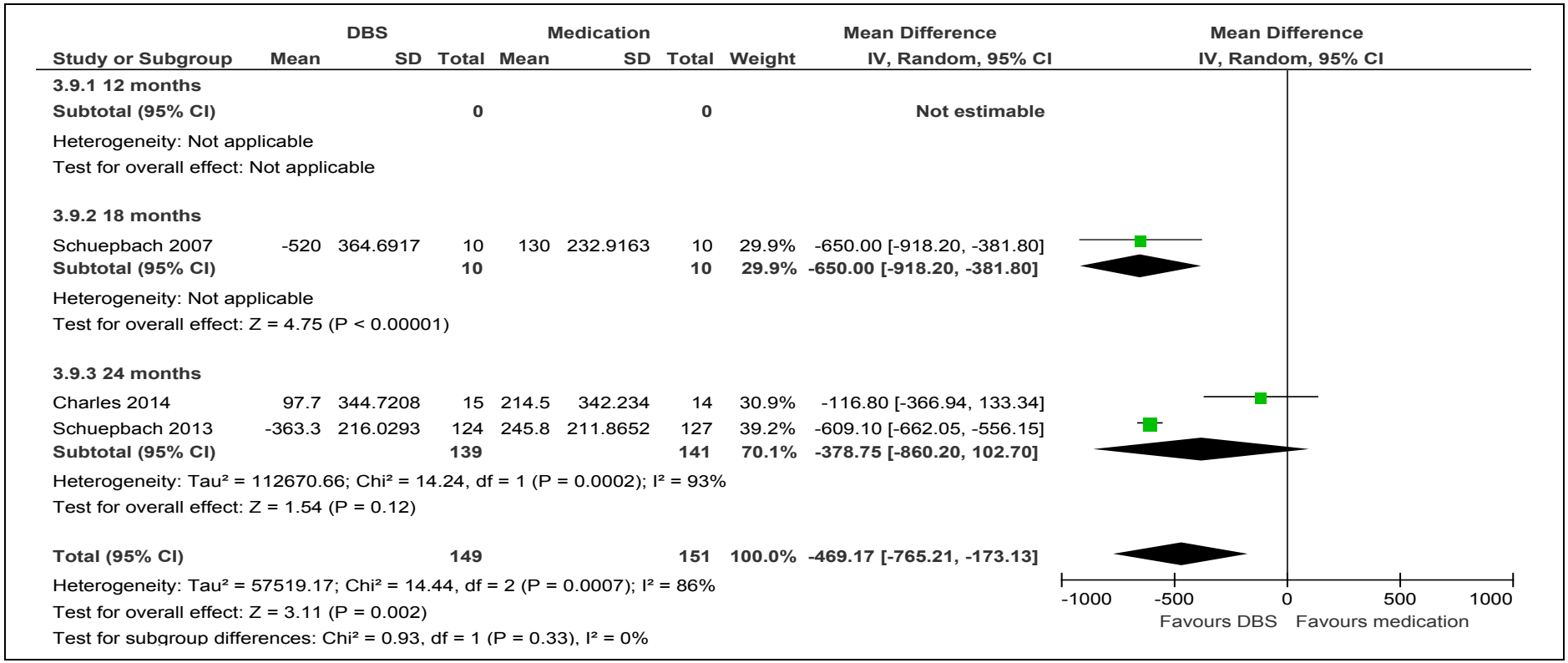
** multiply imputed data; ANCOVA model with baseline score, Hoehn and Yahr status (<3 -v- ≥3) and treatment allocation as covariates of final score*
† observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only
§ participants meeting key eligibility criteria for EarlyStim (age 18–60; disease duration ≥4 years; Hoehn and Yahr <3; improvement of 50% or more with dopaminergic medication on UPDRS-III); observed cases; unadjusted mean difference; not included in meta-analysis – shown for comparison purposes only

DBS -v- medication alone: PDQ-39 – forest plot

E.6.4.5 Medication load

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	DBS	Control	Mean difference (MD) (95% CI)	Quality
Daily dosage of anti-Parkinson's medication (levodopa mg equivalent) (lower is better); 24 months									
3[1,2,3]	RCT	No serious	Serious[4]	No serious	No serious	149	151	-469 (-765 to -173)	MODERATE

1 *Schüpbach 2007*
2 *Schüpbach 2013*
3 *Charles 2014*
4 *I^2 greater than 40% with no obvious explanation for heterogeneity*



DBS -v- medication alone: medication load (levodopa equivalent mg/day)

E.7 Managing and monitoring impulse control disorder as an adverse effect of dopaminergic treatment

E.7.1 Predictors for the development of impulse control disorders

Predictive factors for the development of ICD - unadjusted odds ratios (OR)

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	With ICD	No ICD	OR: 95%CI	Quality
Male gender									
Joutsa 2012 Antonini 2016	Cohort	Serious[1]	Serious[2]	Not serious[3]	Not serious	93	963	6.10 (2.16 to 17.18) 1.14 (0.68, 1.92)	LOW
Comorbid anxiety or depression									
Pontone 2006	Cohort	Very serious[4]	N/A[6]	Not serious[3]	Serious[5]	9	100	2.54 (0.6 to 10.15)	VERY LOW
DA use									
Pontone 2006 Voon 2007	Cohort	Very serious[4]	Not serious	Not serious[3]	Not serious	30	386	10.46 (3.13 to 34.91)	LOW
Pramipexole use									
Imamura 2008 Pontone 2006 Sharma 2015	Cohort	Very serious[4]	Not serious	Not serious[3]	Not serious	20	137	3.26 (1.99 to 5.35)	LOW

[1] Unadjusted odds ratio
[2] N/A; not applicable as only 1 study contributed to this analysis
[3] No serious indirectness; population is as described in review protocol
[4] Serious risk of bias, as assessed by NICE or CASP quality assessment checklist and unadjusted odds ratios
[5] Non-significant results
[6] Serious inconsistency; confidence intervals around point estimates do not overlap

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	With ICD	No ICD	OR: 95%CI	Quality
Amantadine use									
Weintraub 2010b Sharma 2015	Cohort	Serious[1]	Not serious	Not serious	Not serious	728	2357	1.68 (1.36 to 2.08)	MODERATE
Levodopa use									
Imamura 2008 Antonini 2016	Cohort	Serious[1]	Not serious	Serious[3]	Serious[5]	82	752	0.27 (0.05 to 1.29) 2.35 (0.83 to 6.61)	VERY LOW
Combination levodopa and pramipexole therapy									
Imamura 2008	Cohort	Serious[1]	N/A[2]	Serious[3]	Serious[5]	11	37	1.96 (0.3 to 8.79)	VERY LOW
Entacapone use									
Sharma 2015	Cohort	Serious[1]	N/A[2]	Not serious	Serious[5]	74	255	1.47 (0.75 to 2.9)	LOW
Rasagaline use									
Sharma 2015	Cohort	Serious[1]	N/A[2]	Not serious	Serious[5]	74	255	0.98 (0.5 to 1.9)	LOW
Marriage status (unmarried)									
Sharma 2015	Cohort	Serious[1]	N/A[2]	Not serious	Not serious	74	255	9.6 (2.9 to 31.3)	MODERATE
Alcohol intake (high alcohol consumption)									
Sharma 2015	Cohort	Serious[1]	N/A[2]	Not serious	Not serious	74	255	4.0 (2.0 to 8.05)	MODERATE
Smoker status (smoker)									
Imamura 2008	Cohort	Serious[1]	N/A[2]	Serious[3]	Not serious	11	37	7.5 (3.5 to 16.15)	LOW

[1] Unadjusted odds ratio
[2] N/A; not applicable as only 1 study contributed to this analysis
[3] Serious indirectness; population was comprised of only those with pathological gambling

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	With ICD	No ICD	OR: 95%CI	Quality
Family history of alcohol or gambling abuse									
Voon 2007	Cohort	Serious[1]	N/A	Not serious	Not serious	21	286	5.66 (1.78 to 18.03)	MODERATE
Rotigotine dose (12-16mg/day versus 2-10mg/day)									
Antonini 2016	Cohort	Serious[1]	N/A	Not serious	Serious[5]	71	715	0.66 (0.40 to 1.08)	LOW

Predictive factors for the development of ICD - Adjusted odds ratios (OR)

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	With ICD	No ICD	Adjusted OR (95%CI)	Quality
Younger age at onset of PD									
4 studies: Auyeung 2011 Gliadi 2007 Wentraub 2006 Sharma 2015	Cohort	Serious[1]	Serious[2]	Not serious	Not serious	844	2976	OR1: 4.1 (1.1 to 15.9) OR2: 0.99 (0.99 to 1.00) OR3: 2.40 (1.91 to 3.02) OR4: 0.96 (0.93 to 0.99)	LOW
Comorbid anxiety or depression									
Auyeung 2011	Cohort	Serious[3]	N/A	Not serious	Not serious	15	198	10.0 (2.0 to 50.8)	MODERATE
Joutsa 2012	Cohort	Not serious	N/A	Not serious	Not serious	22	248	1.095 (1.001 to 1.195)	HIGH
Gender male									
2 studies: Gliadi 2007 Weintraub 2006	Cohort	Serious[1]	N/A	Not serious	Serious[5]	782	2689	OR1: 1.10 (1.00 to 1.22) OR2: 4.34 (0.54 to 34.4871)	LOW
DA use									
2 studies: Weintraub 2006 Weintraub 2010a	Cohort	Not serious	Not serious	Not serious	Not serious	749	2608	OR1: 16.7 (2.61 to 100) OR2: 2.64 (2.01 to 3.46)	HIGH

[1] Serious risk of bias as assessed by CASP cohort study checklist. Due to the very tight confidence intervals, this Gliadi et al study is heavily weighing the overall estimate
[2] Serious inconsistency; confidence intervals around point estimates do not overlap
[3] Serious risk of bias: Study unclear as to how depression is retrospectively accounted for an in what subset of the study population

[1] Serious risk of bias, as assessed by CASP cohort study quality checklist
[2] Non-significant results

DA LEDD 60-160 mg/d									
Lee 2010	Cohort	Not serious[1]	Not serious	Not serious[2]	Not serious	118	1049	3.3 (1.3 to 9.1)	HIGH
DA LEDD > 150mg/day									
Lee 2010 Sharma 2015	Cohort	Not serious[1]	Serious[3]	Not serious[3]	Not serious	118	1049	OR1 = 4.3 (1.6 to 11.9) OR2 = 4.52 (1.6 to 12.5)	MODERATE
DA LEDD 400 - 800mg/day									
Lee 2010 Sharma 2015	Cohort	Not serious[1]	Serious[4]	Not serious[3]	Serious[8]	118	1049	OR1 = 0.8 (0.4 to 1.6) OR2 = 1.38 (0.5 to 3.82)	LOW
DA LEDD >750mg/day									
Lee 2010	Cohort	Not serious[1]	N/A[4]	Not serious[3]	Serious[8]	118	1049	1.0 (0.5 to 2.1)	MODERATE
DA treatment duration < 2 years									
Gliadi 2007	Cohort	Serious[5]	N/A[5]	Serious[6]	Serious[8]	27	166	0.95 (0.84 to 1.08)	VERY LOW
DA treatment duration 3 - 5 years									
Gliadi 2007	Cohort	Serious[6]	N/A[5]	Serious[7]	Serious[8]	27	166	1.04 (0.01 to 1.18)	VERY LOW
DA treatment duration > 6 years									
Gliadi 2007	Cohort	Serious[6]	N/A[5]	Serious [7]	Not serious	27	166	1.18 (1.00 to 1.39)	LOW
Amantadine use									
2 studies: Weintraub 2006/2010a	Cohort	Not serious[1]	Not serious	Not serious[3]	Not serious	749	2608	1.35 (1.07 to 1.70)	HIGH

[1] Low risk of bias, as assessed by CASP cohort study quality check list
[2] No serious indirectness; population was as described in review protocol
[3] Serious inconsistency: Lee and Sharma define drug dosage differently, whereby Lee defined >160mg and 540-750mg; Sharma defines as 150-300mg, and >300mg
[4] NA; not applicable as one only study contributed to this analysis
[5] Serious risk of bias, as assessed by CASP cohort study quality check list
[6] Serious indirectness; population was comprised of those with CGEC behaviours, not ICD diagnosis
[8] Non-significant results

Levodopa use									
Weintraub 2010a	Cohort	Not serious[1]	N/A[2]	Not serious[3]	Not serious	728	2357	1.51 (1.09 to 2.09)	HIGH
Prior history of ICD symptoms									
Weintraub 2006	Cohort	Not serious[1]	N/A[2]	Not serious[3]	Not serious	21	251	15.54 (2.83 to 76.16)	HIGH
Family history of alcohol abuse									
Weintraub 2010a	Cohort	Not serious[1]	N/A[2]	Not serious[3]	Not serious	728	2357	2.08 (1.33 to 3.25)	HIGH

Incidence of ICD

Quality assessment						Number of patients		
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**ICD**	**No ICD**	**Quality**
ICD rate with short- and long-acting DAs								
Rizos 2016	Survey based on medical records and clinical interviews	Not serious	N/A	Not serious	Serious[4]	57	368	MODERATE
Incidence of ICD and association with dopamine replacement therapy								
Wang 2016	Interviews	Not serious	N/A	Not serious	Serious[4]	9	208	MODERATE

[1] Low risk of bias, as assessed by CASP study quality checklist
[2] NA; not applicable as only one study contributed to the analysis
[3] No serious indirectness; population was as described in review protocol
[4] Serious imprecision: Low numbers of ICD vs no ICD

E.7.2 Managing dopaminergic treatment in people who have developed impulse control disorder

Adjustment of dopaminergic medication

Quality assessment						Number of patients	Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Patients with ICD (N=18)	n/N (%) resolution of symptoms	Quality
Discontinuation of dopaminergic therapy								
Bastiaens 2013	Cohort	Not serious[1]	N/A[2]	Not serious[3]	Very serious[4]	n=10	10/10 (100%)	LOW
Reduction of dopaminergic therapy								
Bastiaens 2013	Cohort	Not serious[1]	N/A[2]	Not serious[3]	Very Serious[4]	n=5	3/5 (60%)	LOW
Continue same dosage of dopaminergic therapy								
Bastiaens 2013	Cohort	Not serious[1]	N/A[2]	Not serious[3]	Very Serious[4]	n=3	0/3 (0%)	LOW
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Patients with ICD (N=18)**	**n/N with DAWS**	**Quality**
Development of DAWS in those who discontinued dopaminergic therapy								
Bastiaens 2013	Cohort	Not serious[1]	N/A[2]	Not serious[3]	Very serious[4]	10	4/10	LOW
Development of DAWS in those who reduced dopaminergic therapy								
Bastiaens 2013	Cohort	Not serious[1]	N/A[2]	Not serious[3]	Very Serious[4]	5	1/5	LOW
Development of DAWS in those who continued same dosage of dopaminergic therapy								
Bastiaens 2013	Cohort	Not serious[1]	N/A[2]	Not serious[3]	Very Serious[4]	3	1/3	LOW

[1] Low risk of bias, as assessed by CASP cohort study quality checklist
[2] NA; not applicable, only 1 study contributed to this analysis
[3] No serious indirectness, study population were as outlined in review protocol
[4] Very serious imprecision; very small sample size to derive meaningful population prevalence estimates

Cognitive behavioural therapy (CBT) for ICD

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	CBT	Control	MD: 95%CI	Quality
Resolution of ICD symptoms									
Okai 2013	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	27	17	-4.17 (-5.8 to -2.5)	HIGH
Effect of CBT on CGIC score									
Okai 2013	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	27	17	-0.8 (-5.6 to -0.3)	HIGH
Effect of CBT on general health (GHQ)									
Okai 2013	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	27	17	-3.8 (-5.6 to -2.0)	HIGH
Effect of CBT on mental health (NPI)									
Okai 2013	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	27	17	-4.7 (-9.1 to -0.3)	HIGH
Effect of CBT on social adjustment									
Okai 2013	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	27	17	-3.6 (-6 to -1.3)	HIGH
Effect of CBT on depression (BDI)									
Okai 2013	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[4]	27	17	-3.5 (-6.6 to 0.4)	MODERATE
Effect of CBT on anxiety (BAI)									
Okai 2013	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[4]	27	17	-1.8 (-5.4 to 1.8)	MODERATE
Effect of CBT on carers perception of the quality of their relationship with their partner (GRIMS marital state)									
Okai 2013	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[4]	27	17	-2.3 (-5.7 to 1.3)	MODERATE
Effect of CBT on carers general health (GHQ)									
Okai 2013	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[4]	27	17	-1.5 (-3.2 to 0.1)	MODERATE

[1] Low risk of bias, as assessed by NICE RCT study quality checklist
[2] NA; not applicable, only 1 study contributed to this analysis

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	CBT	Control	MD: 95%CI	Quality
[1] Low risk of bias, as assessed by NICE RCY study quality checklist [2] NA; not applicable, only 1 study contributed to this analysis [3] No serious indirectness, study population were as outlined in review protocol [4] Non-significant results									

Naltrexone therapy

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Naltrexone	placebo	MD: 95%CI	Quality
QUIP ICD score									
Papay 2014	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	24	26	7.37 (2.45 to 12.66)	HIGH
Change in CGIC score (change of 1 or 2 points from baseline)									
Papay 2014	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[4]	24	26	OR = 1.57 (0.47 to 5.23)	MODERATE
UPDRS motor sore									
Papay 2014	RCT	Not serious[1]	N/A[2]	Not serious[3]	Serious[5]	24	26	−3.70 (−9.24 to 1.84)	MODERATE
Adverse events that lead to study discontinuation									
Papay 2014	RCT	Not serious[1]	N/A[2]	Not serious[3]	Not serious	24	26	0	LOW
[1] Low risk of bias, as assessed by NICE RCT study quality checklist [2] N/A; not applicable, only 1 study contributed to this analysis [3] No serious indirectness, study population were as outlined in review protocol [4] Non-significant result [5] CI cross the MID between 3.25 (Horvath et al., 2015) and 5 points (Schrag et al., 2006)									

Amantadine therapy

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Amantadine	placebo	MD (95% CI)	Quality
Symptom assessment scale (SAS)									
Thomas 2010	Cross-over RCT	Serious[1]	N/A[2]	Serious[3]	Not serious	12	5	-9.6 (-10.12 to-9.08)	LOW
Yale-Brown obsessive compulsive scale (Y-BOCS)									
Thomas 2010	Cross-over RCT	Serious[1]	N/A[2]	Serious[3]	Not serious	12	5	-9.17 (-11.1 to -10.3)	LOW
Resolution of PG spending behaviour									
Thomas 2010	Cross-over RCT	Serious[1]	N/A[2]	Serious[3]	Not serious	12	5	-16.40 (-18.73 to -14.27)	LOW
Adverse events									
Thomas 2010	Cross-over RCT	Serious[1]	N/A[2]	Serious[3]	Not serious	12	5	5 patients dropped out of the amantadine group	LOW

[1] Serious risk of bias, as assessed by NICE RCT quality checklist
[2] N/A; not applicable as only 1 study contributed to this analysis
[3] Serious indirectness; population was composed of those with pathological gambling only

[1] Serious risk of bias, as assessed by NICE RCT quality checklist
[2] NA; not applicable as only 1 study contributed to this analysis
[3] Serious indirectness; population was composed of those with pathological gambling only

E.8 Palliative Care

Patient support needs

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Score on support need survey; 0 (no need) to 5 (serious need)**	**Quality**
Highest self-rated support needs of patients with PD (mean score >2.5)								
Information about PD	Kirstjanson (2006)	Survey	Serious[1]	N/A[2]	Serious[3]	174	3.5	LOW
Equipment for daily living	Kirstjanson (2006)	Survey	Serious[1]	N/A[2]	Serious [3]	174	2.62	LOW

Need for open discussion concerning treatment and care

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Supporting statement**	**Quality**
Open dialogue between patient and clinician								
Discussion of medication	Giles (2009)	Interview	Very serious[4]	N/A[2]	Serious[5]	2	*"I'm the type of woman, I'm afraid to ask too many questions because sometimes I feel like they would say, like you're asking too many questions, just take the pills" (from Giles et al., 2009)*	VERY LOW

[1] Serious risk of bias (CASP cohort quality check list): Methodology not clear, not clear whether all survey material was standardised or validated
[2] N/A; not applicable, single study
[3] Serious indirectness - population was restricted to moderate disease; no advanced or newly diagnosed participants
[4] Very serious risk of bias (CASP qualitative check list): study methodology unclear, interview open to researcher interpretation, role of interviewer in shaping response unclear
[5] Serious indirectness; very small number of patients,

Advance care directives

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Supporting statement**	**Quality**
Advanced care directives								
Input from healthcare team to inform planning	2: Giles (2009) Hasson (2010)	Interview	Very serious[1]	Not serious[2]	Not serious[3]	22	*"To help the family or as a group decide what would be the best care situation for the person, and you know what to expect" (from Giles et al., 2009)*	LOW

Advance care planning

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Percentage(%) of patients who completed action**	**Quality**
Advanced planning of legal will								
Complete will	Kwak (2014)	Survey	Serious[4]	N/A[5]	Not serious[3]	64	93.7%	MODERATE
Share will with spouse	Kwak (2014)	Survey	Serious[4]	N/A[5]	Not serious[3]	64	90.6%	MODERATE
Share will with physician	Kwak (2014)	Survey	Serious[4]	N/A[5]	Not serious[3]	64	37.5%	MODERATE
Preferences for communication about advance care planning								
Advance care	Tuck (2015)	Survey	Serious[4]	N/A[5]	Not serious[3]	267	68.5% (with any kind of advance care planning documents)	MODERATE

[1] Very serious risk of bias (CASP qualitative check list); Hasson (2010) study was retrospective and open to memory bias; methodology very open to researcher interpretation and unclear in Giles (2009)
[2] No serious inconsistency, both studies share similar message
[3] No serious indirectness, all participants were carers of a person with PD and therefore matched protocol
[4] Serious risk of bias (CASP cohort quality check list): Methodology not clear, not clear whether all survey/questionnaire material was standardised or validated
[5] N/A, not applicable; single study

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Percentage(%) of patients who completed action**	**Quality**
planning documents								
When should your doctor discuss advance care planning	Tuck (2015)	Survey	Serious[1]	N/A[2]	Not serious	267	-	MODERATE
Who should ideally raise issues regarding advance care planning to discuss	Tuck (2015)	Survey	Serious[1]	N/A[2]	Not serious	267	94.4% responded	MODERATE

Support needs

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Score on support need survey; 0 (no need) to 5 (serious need)**	**Quality**
Greatest support needs identified by carers (mean score >2.5)								
Information: how to provide care	Kirstjanson (2006)	Survey	Serious[1]	N/A[2]	Serious [3]	141	3.31	LOW

[1] Serious risk of bias (CASP cohort quality check list): Methodology not clear, not clear whether all survey material was standardised or validated
[2] N/A; not applicable, single study
[3] Serious indirectness - population was restricted to moderate disease; no advanced or newly diagnosed participants

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Score on support need survey; 0 (no need) to 5 (serious need)**	**Quality**
Reliable support workers	Kirstjanson (2006)	Survey	Serious[1]	N/A[2]	Serious[3]	141	2.84	LOW
Financial assistance for care	Kirstjanson (2006)	Survey	Serious[1]	N/A[2]	Serious[3]	141	2.72	LOW
Flexible home support program access	Kirstjanson (2006)	Survey	Serious[1]	N/A[2]	Serious[3]	141	2.52	LOW

Multidisciplinary care

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Supporting statement**	**Quality**
Multidisciplinary care need								
Need for coordinated care	2: Hasson (2010) Giles (2009)	Interview	Very serious[1]	Not serious[2]	Not serious[3]	22	*"There seems to be a vague boundary between the responsibilities that one person has and the responsibilities another has. They just don't seem to work as a team or have any team effort as such. You are nearly taking pot luck with each one in turn" (Giles et al., 2009)*	LOW

[1] Very serious risk of bias (CASP qualitative check list); Hasson study was retrospective and open to memory bias; methodology very open to researcher interpretation and unclear in Giles (09)
[2] No serious inconsistency, both studies share similar message
[3] No serious indirectness, all participants were carers of a person with PD and therefore matched protocol

Decision making

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Percentage(%) of carers who elected care goal**	**Quality**
End of life care goals								
Several people discuss; 1 person decide on action	Kwak (2014)	Survey	Serious[1]	N/A[2]	Not serious[3]	64	53%	MODERATE
One person decide alone	Kwak (2014)	Survey	Serious[1]	N/A[2]	Not serious[3]	64	28%	MODERATE
Several people decide on action together	Kwak (2014)	Survey	Serious[1]	N/A[2]	Not serious[3]	64	14%	MODERATE
Carer to be involved in decision making	Kwak (2014)	Survey	Serious[1]	N/A[2]	Not serious[3]	64	92%	MODERATE
Other family members to be involved in decision making	Kwak (2014)	Survey	Serious[1]	N/A[2]	Not serious[3]	64	72%	MODERATE
Physician to be involved in decision making	Kwak (2014)	Survey	Serious[1]	N/A[2]	Not serious[3]	64	70%	MODERATE
Carer, family, and	Kwak (2014)	Survey	Serious[1]	N/A[2]	Not serious[3]	64	52%	MODERATE

[1] Serious risk of bias: Methodology not clear, not clear whether all survey material was standardised or validated
[2] N/A, single study

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Percentage(%) of carers who elected care goal**	**Quality**
physician to be involved in decision making								

Information needs

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Supporting statement**	**Quality**
Information at diagnosis about Parkinson's disease								
understanding the disease	Giles (2009)	Interview	Very serious [1]	N/A[2]	Serious[3]	5	*"you have to be prepared and understand it's just kind of a shocker and no one really explained to us what all of this meant" (from Giles et al., 2009)*	VERY LOW
Information to help carers prepare to advancement of disease								
Preparation for end of life	Hasson (2010)	Interview	Serious[4]	N/A[2]	Not serious[5]	15	*"I knew he was deteriorating but I didn't expect him to die so soon" (Hasson et al., 2010)"*	MODERATE

[1] Very serious risk of bias (CASP qualitative check list): methodology unclear and open to researcher interpretation
[2] N/A, not applicable, single study
[3] Serious indirectness, very small sample size, may be unrepresentative of general population
[4] Serious bias (CASP qualitative check list), retrospective perspective may bias responses
[5] No serious indirectness; carers of patient with PD as specified in protocol

Satisfaction with care

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Percentage (%) of carers who reported satisfaction (FAMCARE assessment)**	**Quality**
Satisfaction with care received								
Information giving	Kirstjanson (2006)	Survey	Serious[1]	N/A	Serious[2]	141	69%	LOW
Physical care	Kirstjanson (2006)	Survey	Serious[1]	N/A	Serious[2]	141	80%	LOW
Phycosocial care	Kirstjanson (2006)	Survey	Serious[1]	N/A	Serious[2]	141	63%	LOW
Availability of care	Kirstjanson (2006)	Survey	Serious[1]	N/A	Serious[2]	141	71%	LOW

Respite opportunities and availability of care

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Supporting statement**	**Quality**
Respite opportunities								
Access to respite	2: Hasson (2010) Giles (2009)	Interview	Very serious[3]	Not serious[4]	Not serious[5]	22	*"they (government homecare) still haven't called us …so we're lucky that, you know, we finally made the decision to move on. Because I don't know what we would have done… I don't think my mom would have lasted" (from Giles et al., 2009)*	LOW

[1] Serious risk of bias: Methodology not clear, not clear whether all survey material was standardised or validated
[2] Serious indirectness - population was restricted to moderate disease; no advanced or newly diagnosed participants
[3] Very serious risk of bias (CASP qualitative check list); Hasson (2010) study was retrospective and open to memory bias; methodology very open to researcher interpretation and unclear in Giles (2009)
[4] No serious inconsistency, both studies share similar message
[5] No serious indirectness, all participants were carers of a person with PD and therefore matched protocol

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Supporting statement**	**Quality**

Access to domiciliary palliative care services

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Supporting statement**	**Quality**
Access to domiciliary palliative care services								
Access to palliative care services	2: Hasson (2010) Giles (2009)	Interview	Very serious[4]	Not serious[2]	Not serious	22	*"that (home care services) is something that you know somebody should tell those people". (from Giles et al., 2009)*	LOW

Patient and carer quality of life

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Mean score (SD) on self-rated QoL scale (0 = very poor, 10 = excellent\)**	**Quality**
Patient quality of life (QoL)								
Patient-rated QoL	Kirstjanson (2006)	Survey	Serious[1]	N/A[2]	Serious[3]	174	6.87 (2.29)	LOW
Satisfaction with QoL	Kirstjanson (2006)	Survey	Serious[1]	N/A[2]	Serious[3]	174	5.55 (2.68)	LOW

[1] Very serious risk of bias: Methodology not clear, not clear whether all survey material was standardised or validated
[2] N/A, not applicable, single study
[3] Serious indirectness - population was restricted to moderate disease; no advanced or newly diagnosed participants
[4] Very serious risk of bias (CASP qualitative check list); Hasson (2010) study was retrospective and open to memory bias; methodology very open to researcher interpretation and unclear in Giles (2009)

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Mean score (SD) on self-rated QoL scale (0 = very poor, 10 = excellent\)**	**Quality**
Carer quality of life (QoL)								
carer-rated QoL	Kirstjanson (2006)	Survey	Serious[1]	N/A[2]	Serious[3]	141	6.59 (2.27)	LOW
Satisfaction with QoL	Kirstjanson (2006)	Survey	Serious[1]	N/A[2]	Serious[3]	141	6.35 (2.58)	LOW

Symptom severity experience in patients

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Mean score (SD) on symptom assessment scale (SAS; 0 = no problem, 10=worst problem)**	**Quality**
Worst experienced symptoms								
Fatigue and tiredness	Kirstjanson (2006)	Survey	Serious[1]	NA[2]	Serious[3]	174	5.1 (2.9)	LOW
concentration	Kirstjanson (2006)	Survey	Serious[1]	NA[2]	Serious[3]	174	3.9 (3.1)	LOW
sleeping	Kirstjanson (2006)	Survey	Serious[1]	NA[2]	Serious[3]	174	4.1 (3.3)	LOW

Incidence of anxiety and depression in patients and carers

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Percentage (%) of patients/carers experiencing anxiety and/or depression assessed by Hospital Anxiety Depression Scale (HADS) in patients and General health questionnaire (QHQ) in carers**	**Quality**
Patient self-reported moderate-to severe experience								
Anxiety	Kirstjanson (2006)	Survey	Serious[1]	N/A[2]	Serious[3]	174	20%	LOW
Depression	Kirstjanson (2006)	Survey	Serious[1]	N/A[2]	Serious[3]	174	30%	LOW
Carer self-reported moderate-to severe experience								
Anxiety and depression	Kirstjanson (2006)	Survey	Serious[1]	N/A[2]	Serious[3]	141	19%	LOW

[1] Serious risk of bias: Methodology not clear, not clear whether all survey material was standardised or validated
[2] N/A, not applicable, single study
[3] Serious indirectness - population was restricted to moderate disease; no advanced or newly diagnosed participants