Depression in adults: treatment and management

Appendix L: GRADE profiles

Contents

Organisation and service delivery (chapter 5)

Service delivery

Collaborative care versus control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	COLLABORATIVE CARE	CONTROL	Relative (95% CI)	Absolute		
Depression symptoms- 6 months (follow-up mean 6; Better indicated by lower values)												
47	randomised trials	very serious[1]	serious[2]	no serious indirectness	no serious imprecision	none	0	-	-	SMD 0.31 lower (0.39 to 0.23 lower)	⊕OOO VERY LOW	CRITICAL
Depression symptoms- Simple collaborative care (follow-up mean 6 months; Better indicated by lower values)												
36	randomised trials	very serious[1]	serious[2]	no serious indirectness	no serious imprecision	none	0	-	-	SMD 0.32 lower (0.41 to 0.22 lower)	⊕OOO VERY LOW	CRITICAL
Depression symptoms- Complex collaborative care (follow-up mean 6 months; Better indicated by lower values)												
11	randomised trials	very serious[1]	serious[2]	no serious indirectness	no serious imprecision	none	0	-	-	SMD 0.28 lower (0.43 to 0.13 lower)	⊕OOO VERY LOW	CRITICAL
Depression symptoms at follow-up (follow-up mean 12 months; Better indicated by lower values)												
9	randomised trials	very serious[1]	very serious[3]	no serious indirectness	no serious imprecision	none	2259	2280	-	SMD 0.23 lower (0.4 to 0.07 lower)	⊕OOO VERY LOW	CRITICAL
Depression symptoms at follow-up - Simple collaborative care (follow-up mean 12 months; Better indicated by lower values)												

6	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	1264	1304	-	SMD 0.21 lower (0.3 to 0.12 lower)	⊕⊕OO LOW	CRITICAL
Depression symptoms at follow-up - Complex collaborative care (follow-up mean 12 months; Better indicated by lower values)												
3	randomised trials	very serious[1]	very serious[3]	no serious indirectness	serious[4]	none	995	976	-	SMD 0.27 lower (0.72 lower to 0.17 higher)	⊕OOO VERY LOW	CRITICAL
Non-response at follow-up (follow-up mean 12 months)												
10	randomised trials	very serious[1]	serious[2]	no serious indirectness	no serious imprecision	none	872/1732 (50.3%)	1156/1546 (74.8%)	RR 0.72 (0.63 to 0.81)	209 fewer per 1000 (from 142 fewer to 277 fewer) 191 fewer per 1000 (from 129 fewer to 252 fewer)	⊕OOO VERY LOW	CRITICAL
								68.1%				
Non-response at follow-up- Simple collaborative care (follow-up mean 12 months)												
4	randomised trials	very serious[1]	serious[2]	no serious indirectness	serious[4]	none	181/482 (37.6%)	247/413 (59.8%)	RR 0.66 (0.47 to 0.92)	203 fewer per 1000 (from 48 fewer to 317 fewer) 134 fewer per 1000 (from 32 fewer to 209 fewer)	⊕OOO VERY LOW	CRITICAL
								39.4%				
Non-response at follow-up - Complex collaborative care (follow-up mean 12 months)												
6	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	691/1250 (55.3%)	909/1133 (80.2%)	RR 0.75 (0.66 to 0.85)	201 fewer per 1000 (from 120 fewer to 273 fewer) 188 fewer per 1000 (from 112 fewer to 255 fewer)	⊕⊕OO LOW	CRITICAL
								75%				
Antidepressant use- 6 months (follow-up mean 6 months)												
31	randomised trials	very serious[1]	serious[2]	no serious indirectness	no serious imprecision	none	-	-	RR 1.39 (1.26 to 1.52)	- -	⊕OOO VERY LOW	CRITICAL
								0%				
Antidepressant use- 6 months - Simple collaborative care												

22	randomised trials	very serious[1]	serious[2]	no serious indirectness	no serious imprecision	none	-	-	RR 1.45 (1.26 to 1.66)	-	⊕OOO VERY LOW	CRITICAL
								0%		-		
Antidepressant use- 6 months - Complex collaborative care												
10	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	-	-	RR 1.29 (1.2 to 1.38)	-	⊕OOO VERY LOW	CRITICAL
								0%		-		
Antidepressant use at follow-up (follow-up mean 12 months)												
10	randomised trials	very serious[1]	serious[2]	no serious indirectness	serious[4]	none	1156/1799 (64.3%)	972/1819 (53.4%)	RR 1.18 (1.03 to 1.35)	96 more per 1000 (from 16 more to 187 more)	⊕OOO VERY LOW	CRITICAL
								55%		99 more per 1000 (from 16 more to 193 more)		
Antidepressant use at follow-up - Simple collaborative care (follow-up mean 12 months)												
6	randomised trials	very serious[1]	serious[2]	no serious indirectness	serious[4]	none	358/686 (52.2%)	338/697 (48.5%)	RR 1.14 (0.9 to 1.46)	68 more per 1000 (from 48 fewer to 223 more)	⊕OOO VERY LOW	CRITICAL
								38%		53 more per 1000 (from 38 fewer to 175 more)		
Antidepressant use at follow-up - Complex collaborative care (follow-up mean 12 months)												
4	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	798/1113 (71.7%)	634/1122 (56.5%)	RR 1.26 (1.17 to 1.35)	147 more per 1000 (from 96 more to 198 more)	⊕OOO VERY LOW	CRITICAL
								61.9%		161 more per 1000 (from 105 more to 217 more)		
Non-remission at 6 months (sinple collaborative care)												
1	randomised trials	very serious[1,5]	no serious inconsistency	no serious indirectness	serious[4]	none	64/115 (55.7%)	66/96 (68.8%)	RR 0.81 (0.66 to 1)	131 fewer per 1000 (from 234 fewer to 0 more)	⊕OOO VERY LOW	CRITICAL

Non-remission at follow-up (follow-up mean 12 months)												
2	randomised trials	serious[6]	very serious[3]	no serious indirectness	serious[4]	none	88/197 (44.7%)	156/198 (78.8%)	RR 0.58 (0.38 to 0.89)	331 fewer per 1000 (from 87 fewer to 488 fewer)	⊕OOO VERY LOW	CRITICAL
Non-remission at follow-up - simple collaborative care (follow-up mean 12 months)												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	serious[7]	none	47/110 (42.7%)	95/104 (91.3%)	RR 0.47 (0.37 to 0.59)	484 fewer per 1000 (from 375 fewer to 575 fewer)	⊕⊕OO LOW	CRITICAL
Non-remission at follow-up - complex collaborative care (follow-up mean 12 months)												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	serious[4]	none	41/87 (47.1%)	61/954 (6.4%)	RR 0.73 (0.56 to 0.95)	17 fewer per 1000 (from 3 fewer to 28 fewer)	⊕⊕OO LOW	CRITICAL

[1] ROB high or unclear across multiple domains in most studies
[2] I2 >50%
[3] I2 >80%
[4] 95% CI crosses one clinical decision threshold
[5] ROB high or unclear across multiple domains
[6] ROB high or unclear across a two to three domains
[7] OIS not met (<300 events)

Collaborative care versus active intervention

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**COLLABORATIVE CARE**	**OTHER COMPARISON**	**Relative (95% CI)**	**Absolute**		
Simple collaborative care: Standards CC vs patient centred CC- remission at follow-up (follow-up mean 12 months)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	27/65 (41.5%)	22/67 (32.8%)	RR 1.27 (0.81 to 1.98)	89 more per 1000 (from 62 fewer to 322 more)	⊕⊕OO LOW	CRITICAL

								32.8%		89 more per 1000 (from 62 fewer to 321 more)		
Telebased CC vs Practice based CC- response- 6 months (follow-up mean 6 months)													
1	randomised trials	serious	no serious inconsistency	no serious indirectness	no serious imprecision	none	70/153 (45.8%)	25/165 (15.2%)	RR 3.02 (2.02 to 4.51)	306 more per 1000 (from 155 more to 532 more)	⊕⊕⊕O MODERATE	CRITICAL
								15.2%		307 more per 1000 (from 155 more to 534 more)		
Telebased CC vs practice based CC- response at follow-up (follow-up mean 12 months)													
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	73/138 (52.9%)	31/149 (20.8%)	RR 2.54 (1.79 to 3.61)	320 more per 1000 (from 164 more to 543 more)	⊕⊕OO LOW	CRITICAL
								20.8%		320 more per 1000 (from 164 more to 543 more)		

[1] ROB high or unclear across two to three domains
[2] 95% CI crosses one clinical decision threshold

Stepped care versus control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**STEPPED CARE**	**CONTROL**	**Relative (95% CI)**	**Absolute**		
Remission at endpoint												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	40/74 (54.1%)	29/74 (39.2%)	RR 1.38 (0.97 to 1.96)	149 more per 1000 (from 12 fewer to 376 more)	⊕⊕OO LOW	CRITICAL
								39.2%		149 more per 1000 (from 12 fewer to 376 more)		
Depression symptoms at endpoint (measured with: PHQ-9; Better indicated by lower values)												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	137	64	-	MD 1.4 lower (2.87 lower to 0.07 higher)	⊕⊕OO LOW	CRITICAL
Antidepressant use (follow-up mean 6 months)												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	none	28/86 (32.6%)	23/84 (27.4%)	RR 1.19 (0.75 to 1.89)	52 more per 1000 (from 68 fewer to 244 more)	⊕OOO VERY LOW	CRITICAL

[1] ROB high or unclear in two to three domains
[2] 95% CI crosses one clinical decision threshold
[3] High or unclear ROB in most domains
[4] 95% CI crosses two clinical decision thresholds

Medication management versus control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**MEDICATION MANAGEMENT**	**CONTROL**	**Relative (95% CI)**	**Absolute**		
Mean change in depression scores (Better indicated by lower values)												
9	randomised trials	very serious[1]	serious[2]	no serious indirectness	no serious imprecision	none	0	-	-	SMD 0.13 lower (0.33 lower to 0.06 higher)	⊕OOO VERY LOW	CRITICAL
Mean change in depression scores at follow-up (follow-up mean 12 months; Better indicated by lower values)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[4,5]	none	113	106	-	MD 2 lower (4.86 lower to 0.86 higher)	⊕OOO VERY LOW	CRITICAL
Antidepressant use at endpoint												
4	randomised trials	serious[3]	serious[2]	no serious indirectness	serious[6]	none	-	-	Not estimable	-	⊕OOO VERY LOW	CRITICAL

[1] ROB high or unclear across multiple domains
[2] I2 > 50%
[3] ROB high or unclear across two to three domains
[4] OIS not met (<400 participants)
[5] 95% CI crosses two clinical decision thresholds
[6] 95% CI crosses one clinical decision threshold

Care co-ordination versus control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	CARE CO-ORDINATION	CONTROL	Relative (95% CI)	Absolute		
Mean change in depression scores at endpoint (Better indicated by lower values)												
4	randomised trials	very serious[1,2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	0	-	-	SMD 0.05 lower (0.35 lower to 0.25 higher)	⊕⊕OO LOW	CRITICAL
Antidepressant adherence at follow-up (follow-up mean 12 months)												
3	randomised trials	serious[1]	serious[3]	no serious indirectness	very serious[4]	none	-	-	RR 1.79 (0.68 to 4.72)	-	⊕OOO VERY LOW	CRITICAL
								0%		-		

[1] ROB high or unclear in two to three domains
[2] ROB high or unclear across multiple domains
[3] I2 > 50%
[4] 95% CI crosses two clinical decision thresholds

Integrated care versus control

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	INTEGRATED CARE	CONTROL	Relative (95% CI)	Absolute		
Mean change in depression scores at endpoint (Better indicated by lower values)												
3	randomised trials	very serious[1]	serious[2]	no serious indirectness	no serious imprecision	none	0	-	-	SMD 0.05 lower (0.26 lower to 0.16 higher)	⊕OOO VERY LOW	CRITICAL
Mean change in depression scores at endpoint - Integrated care vs control (Better indicated by lower values)												
2	randomised trials	very serious[1]	serious[2]	no serious indirectness	serious[3,4]	none	0	-	-	SMD 0.19 lower (0.55 lower to 0.17 higher)	⊕OOO VERY LOW	CRITICAL
Mean change in depression scores at endpoint - Integrated care vs speciality referral system (Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	0	-	-	SMD 0.08 higher (0.03 lower to 0.19 higher)	⊕⊕OO LOW	CRITICAL
Mean change in depression scores at follow-up (follow-up mean 12 months; Better indicated by higher values)												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[6]	none	189	186	-	MD 0.01 higher (0.11 lower to 0.13 higher)	⊕⊕OO LOW	CRITICAL
Antidepressant adherence												
2	randomised trials	very serious[1]	serious[2]	no serious indirectness	very serious[3]	none	-	-	Not estimable	-	⊕OOO VERY LOW	CRITICAL

[1] ROB high or unclear in multiple domains
[2] I2 > 50%
[3] 95% CI crosses two clinical decision thresholds
[4] 95% CI crosses one clinical decision threshold
[5] ROB high or unclear in two to three domains
[6] OIS not met (<400 participants)

Service delivery models for relapse prevention

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	RELAPSE PREVENTION	Control	Relative (95% CI)	Absolute		
Collaborative care (simple)- depression symptoms at endpoint (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	174	153	-	MD 0.09 lower (0.2 lower to 0.02 higher)	⊕OOO VERY LOW	CRITICAL
Collaborative care (simple)- relapse at follow-up (follow-up mean 12 months)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	67/192 (34.9%)	67/194 (34.5%)	RR 1.01 (0.77 to 1.33)	3 more per 1000 (from 79 fewer to 114 more)	⊕⊕OO LOW	CRITICAL
								34.5%		3 more per 1000 (from 79 fewer to 114 more)		
Stepped care at follow-up (follow-up mean 12 months)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	24/74 (32.4%)	16/62 (25.8%)	RR 1.26 (0.74 to 2.15)	67 more per 1000 (from 67 fewer to 297 more)	⊕OOO VERY LOW	CRITICAL
								25.8%		67 more per 1000 (from 67 fewer to 297 more)		

[1] ROB high or unclear in multiple domains
[2] OIS not met (<400 participants)
[3] 95% CI crosses one clinical decision threshold
[4] 95% CI crosses two clinical decision thresholds

Settings for care

Crisis resolution team care versus standard care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Crisis resolution team care	Standard care	Relative (95% CI)	Absolute		
Lost to follow-up (follow-up mean 12 months; assessed with: Number of participants lost to follow-up by the end of the study)												

1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	very serious[3]	none	17/135 (12.6%)	17/125 (13.6%)	RR 0.93 (0.49 to 1.73)	10 fewer per 1000 (from 69 fewer to 99 more)	⊕OOO VERY LOW	
								13.6%		10 fewer per 1000 (from 69 fewer to 99 more)		
Symptom severity (BPRS) (follow-up mean 8 weeks; measured with: Brief Psychiatric Rating Scale (BPRS) 8 weeks after crisis; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	serious[4]	none	107	104	-	SMD 0.29 lower (0.56 to 0.02 lower)	⊕OOO VERY LOW	
Admission as inpatient (follow-up mean 6 months; assessed with: Number of participants that had been admitted to a psychiatric ward within 6 months after crisis)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	serious[5]	none	39/134 (29.1%)	84/124 (67.7%)	RR 0.43 (0.32 to 0.57)	386 fewer per 1000 (from 291 fewer to 461 fewer)	⊕OOO VERY LOW	
								67.7%		386 fewer per 1000 (from 291 fewer to 460 fewer)		
Bed days in hospital (follow-up mean 6 months; measured with: Number of bed days in hospital for those admitted within 6 months after crisis; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	serious[4]	none	134	123	-	MD 18.9 lower (29.38 to 8.42 lower)	⊕OOO VERY LOW	
Satisfaction (follow-up mean 8 weeks; measured with: Client Satisfaction Questionnaire - 8 item version (CSQ-8) 8 weeks after crisis; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	serious[4]	none	118	108	-	SMD 0.23 higher (0.03 lower to 0.49 higher)	⊕OOO VERY LOW	
Quality of life (follow-up mean 8 weeks; measured with: Manchester short assessment of quality of life (MANSA) 8 weeks after crisis; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	serious[4]	none	114	103	-	SMD 0.11 lower (0.37 lower to 0.16 higher)	⊕OOO VERY LOW	
Social functioning (8 weeks after crisis) (follow-up mean 8 weeks; measured with: Life Skills Profile (LSP); Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	serious[4]	none	133	124	-	SMD 0.2 higher (0.05 lower to 0.44 higher)	⊕OOO VERY LOW	
Social functioning (at endpoint) (follow-up mean 6 months; measured with: Life Skills Profile (LSP); Better indicated by lower values)												

1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	serious[4]	none	133	122	-	SMD 0.06 higher (0.18 lower to 0.31 higher)	⊕OOO VERY LOW	

[1] High risk of bias associated with randomisation method due to significant difference between groups and baseline and non-blind participants, intervention administrator(s) and outcome assessor(s)
[2] Not depression-specific population
[3] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)
[4] N<400
[5] Events<300

Acute day hospital care versus inpatient care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Acute day hospital care**	**Inpatient care**	**Relative (95% CI)**	**Absolute**		
Lost to follow-up (follow-up 3-14 months; assessed with: Number of participants lost to follow-up by the end of the study)												
6	randomised trials	serious[1]	no serious inconsistency	serious[2]	serious[3]	none	310/907 (34.2%)	270/856 (31.5%)	RR 1.25 (0.96 to 1.63)	79 more per 1000 (from 13 fewer to 199 more)	⊕OOO VERY LOW	
								17.8%		44 more per 1000 (from 7 fewer to 112 more)		
Death (suicide) (follow-up mean 14 months; assessed with: Number of participants that committed suicide during the study period)												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	none	0/596 (0%)	3/521 (0.6%)	RR 0.12 (0.01 to 2.41)	5 fewer per 1000 (from 6 fewer to 8 more)	⊕OOO VERY LOW	
								0.6%		5 fewer per 1000 (from 6 fewer to 8 more)		
Remission of psychiatric symptoms (follow-up 3-13 months; assessed with: Present State Examination: Index of Definition≤4/<7 on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	very serious[6]	no serious inconsistency	serious[2]	very serious[7]	reporting bias[8]	33/80 (41.3%)	33/71 (46.5%)	RR 0.91 (0.65 to 1.26)	42 fewer per 1000 (from 163 fewer to 121 more)		

								36.9%		33 fewer per 1000 (from 129 fewer to 96 more)	⊕OOO VERY LOW	
Response (follow-up mean 3 months; assessed with: Number of people showing ≥47% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[9]	no serious inconsistency	no serious indirectness	very serious[7]	reporting bias[10]	6/24 (25%)	8/20 (40%)	RR 0.62 (0.26 to 1.5)	152 fewer per 1000 (from 296 fewer to 200 more)	⊕OOO VERY LOW	
								40%		152 fewer per 1000 (from 296 fewer to 200 more)		
Symptom severity (2-3 months post-admission) (follow-up 2-3 months; measured with: Comprehensive Psychopathological Rating Scale (CPRS; change score)/Brief Psychiatric Rating Scale (BPRS; change score)/Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
3	randomised trials	very serious[11]	serious[12]	serious[2]	no serious imprecision	none	682	599	-	SMD 0.05 higher (0.22 lower to 0.33 higher)	⊕OOO VERY LOW	
Symptom severity (12-14 months post-admission) (follow-up 12-14 months; measured with: Comprehensive Psychopathological Rating Scale (CPRS; change score)/Brief Psychiatric Rating Scale (BPRS; change score); Better indicated by lower values)												
2	randomised trials	very serious[11]	very serious[13]	serious[2]	serious[14]	none	663	586	-	SMD 0.19 lower (0.81 lower to 0.42 higher)	⊕OOO VERY LOW	
Duration of index admission (follow-up 12-14 months; measured with: Number of days/months in hospital; Better indicated by lower values)												
4	randomised trials	very serious[11]	no serious inconsistency	serious[2]	no serious imprecision	none	800	735	-	SMD 0.55 higher (0.44 to 0.65 higher)	⊕OOO VERY LOW	
Readmission (follow-up mean 12 months; assessed with: Number of patients readmitted to hospital)												
3	randomised trials	serious[15]	serious[12]	serious[2]	very serious[5]	reporting bias[8]	39/183 (21.3%)	47/189 (24.9%)	RR 0.79 (0.41 to 1.52)	52 fewer per 1000 (from 147 fewer to 129 more)	⊕OOO VERY LOW	
								21.5%		45 fewer per 1000 (from 127 fewer to 112 more)		
Discharge (follow-up mean 3 months; assessed with: Number of participants discharged from hospital within 3 months of admission)												

1	randomised trials	serious[15]	no serious inconsistency	serious[2]	serious[16]	reporting bias[8]	17/41 (41.5%)	33/48 (68.8%)	RR 0.6 (0.4 to 0.91)	275 fewer per 1000 (from 62 fewer to 412 fewer) 275 fewer per 1000 (from 62 fewer to 413 fewer)	⊕OOO VERY LOW	
								68.8%				
Service utilisation: Emergency contacts (follow-up mean 4 months; assessed with: Number of participants making emergency contacts within 4 months post-admission)												
1	randomised trials	serious[17]	no serious inconsistency	serious[2]	serious[3]	reporting bias[8]	12/38 (31.6%)	6/45 (13.3%)	RR 2.37 (0.98 to 5.71)	183 more per 1000 (from 3 fewer to 628 more) 182 more per 1000 (from 3 fewer to 626 more)	⊕OOO VERY LOW	
								13.3%				
Service utilisation: Outpatient contact (follow-up mean 4 months; assessed with: Number of participants making outpatient contacts within 4 months post-admission)												
1	randomised trials	serious[17]	no serious inconsistency	serious[2]	very serious[5]	reporting bias[8]	14/38 (36.8%)	12/45 (26.7%)	RR 1.38 (0.73 to 2.62)	101 more per 1000 (from 72 fewer to 432 more) 101 more per 1000 (from 72 fewer to 433 more)	⊕OOO VERY LOW	
								26.7%				
Satisfaction (follow-up mean 4 months; assessed with: Number of participants satisfied or very satisfied with their treatment)												
1	randomised trials	very serious[17]	no serious inconsistency	serious[2]	serious[16]	reporting bias[8]	31/38 (81.6%)	19/45 (42.2%)	RR 1.93 (1.33 to 2.81)	393 more per 1000 (from 139 more to 764 more) 392 more per 1000 (from 139 more to 764 more)	⊕OOO VERY LOW	
								42.2%				
Satisfaction (follow-up mean 2 months; measured with: Cliet Assessment of Treatment (CAT); Better indicated by lower values)												
1	randomised trials	very serious[11]	no serious inconsistency	serious[2]	no serious imprecision	none	596	521	-	SMD 0.03 higher (0.09 lower to 0.15 higher)	⊕OOO VERY LOW	
Quality of life (2-months post-admission) (follow-up mean 2 months; measured with: Manchester short assessment of quality of life (MANSA); Better indicated by lower values)												

1	randomised trials	very serious[11]	no serious inconsistency	serious[2]	no serious imprecision	none	596	521	-	SMD 0.01 higher (0.11 lower to 0.13 higher)	⊕OOO VERY LOW	
Quality of life (14-months post-admission) (follow-up mean 14 months; measured with: Manchester short assessment of quality of life (MANSA); Better indicated by lower values)												
1	randomised trials	very serious[11]	no serious inconsistency	serious[2]	no serious imprecision	none	596	521	-	SMD 0.01 higher (0.11 lower to 0.13 higher)	⊕OOO VERY LOW	
Social functioning response (follow-up 12-13 months; assessed with: 2 role disabilities or less on Groningen Social Disabilities Schedule (GSDS)/Number of participants living in the community and social functioning at previous level (according to the social performance and behaviour assessment schedule))												
2	randomised trials	very serious[18]	no serious inconsistency	serious[2]	serious[19]	reporting bias[8]	41/91 (45.1%)	30/90 (33.3%)	RR 1.36 (0.94 to 1.96)	120 more per 1000 (from 20 fewer to 320 more)	⊕OOO VERY LOW	
								34.2%		123 more per 1000 (from 21 fewer to 328 more)		
Social functioning impairment (2-months post-admission) (follow-up mean 2 months; measured with: Groningen Social Disabilities Schedule, Second revision (GSDS-II); Better indicated by lower values)												
1	randomised trials	very serious[11]	no serious inconsistency	serious[2]	no serious imprecision	none	596	521	-	SMD 0.3 lower (0.42 to 0.19 lower)	⊕OOO VERY LOW	
Social functioning impairment (14-months post-admission) (follow-up mean 14 months; measured with: Groningen Social Disabilities Schedule, Second revision (GSDS-II); Better indicated by lower values)												
1	randomised trials	very serious[11]	no serious inconsistency	serious[2]	no serious imprecision	none	596	521	-	SMD 0.15 lower (0.27 to 0.04 lower)	⊕OOO VERY LOW	
Carer distress (3-months post-admission) (follow-up mean 3 months; measured with: General Health Questionnaire (GHQ; change score); Better indicated by lower values)												
1	randomised trials	very serious[15]	no serious inconsistency	serious[2]	serious[14]	none	38	39	-	MD 1.1 lower (3.15 lower to 0.95 higher)	⊕OOO VERY LOW	
Carer distress (12-months post-admission) (follow-up mean 12 months; measured with: General Health Questionnaire (GHQ; change score); Better indicated by lower values)												

1	randomised trials	very serious[15]	no serious inconsistency	serious[2]	serious[14]	none	24	31	-	MD 0.4 lower (2.98 lower to 2.18 higher)	⊕OOO VERY LOW	

[1] Randomisation method was unclear (or high risk associated with it due to significant baseline differences). Non-blind participants, intervention administrator(s) and unclear blinding of, or non-blind, outcome assessor(s)
[2] Non depression-specific population
[3] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 1.25)
[4] High risk of bias associated with randomisation method due to significant difference between groups at baseline. Non-blind participants, intervention administrator(s) and outcome assessor(s). Unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)
[5] 95% CI crosses line of no effect and both threshold for clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)
[6] Unclear randomisation method and method of allocation concealment. Non-blind participants and intervention administrator(s) and unclear blinding of outcome assessment
[7] 95% CI crosses line of no effect and threshold for clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)
[8] Data cannot be extracted for all outcomes (measure of variance not reported)
[9] Unclear blinding of allocation concealment. Non-blind participants and intervention administrator(s) and unclear blinding of outcome assessment. Unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)
[10] A non-standard definition of response selected (e.g. 47% rather than 50%)
[11] High risk of bias associated with randomisation method due to significant difference between groups at baseline. Non-blind participants, intervention administrator(s) and outcome assessment. Unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)
[12] I-squared>50%
[13] I-squared>80%
[14] 95% CI crosses both line of no effect and threshold for clinically important benefit (SMD -0.5)
[15] Non-blind participants, intervention administrator(s) and outcome assessment
[16] Events<300
[17] Unclear randomisation method and allocation concealment, and non-blind participants, intervention administrator(s) and outcome assessment
[18] Non-blind participants and intervention administrator(s) and non-blind, or unclear blinding of, outcome assessment. Unclear risk of attrition bias (drop-out>20% but difference between groups<20%)
[19] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)

Non-acute day hospital care versus outpatient care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Non-acute day hospital care versus outpatient care**	**Control**	**Relative (95% CI)**	**Absolute**		
Lost to follow-up (follow-up 6-24 months; assessed with: Number of participants lost to follow-up by the end of the study)												
3	randomised trials	serious[1]	serious[2]	serious[3]	very serious[4]	reporting bias[5]	24/136 (17.6%)	30/145 (20.7%)	RR 0.81 (0.24 to 2.7)	39 fewer per 1000 (from 157 fewer to 352 more)		

								20.7%		39 fewer per 1000 (from 157 fewer to 352 more)	⊕OOO VERY LOW	
Death (all causes) (follow-up mean 24 months; assessed with: Number of participants who died due to any causes during the study period)												
1	randomised trials	serious[6]	no serious inconsistency	serious[3]	very serious[4]	none	2/48 (4.2%)	1/58 (1.7%)	RR 2.42 (0.23 to 25.85)	24 more per 1000 (from 13 fewer to 428 more)	⊕OOO VERY LOW	
								1.7%		24 more per 1000 (from 13 fewer to 422 more)		
Symptom severity (4-6 months post-admission) (follow-up 4-6 months; measured with: Psychiatric Evaluation Form (change score)/Present State Examination (change score); Better indicated by lower values)												
2	randomised trials	serious[7]	very serious[8]	serious[3]	very serious[9]	none	75	69	-	SMD 0.08 higher (0.72 lower to 0.88 higher)	⊕OOO VERY LOW	
Symptom severity (8-12 months post-admission) (follow-up 8-12 months; measured with: Psychiatric Evaluation Form (change score)/Present State Examination (change score); Better indicated by lower values)												
2	randomised trials	serious[7]	no serious inconsistency	serious[3]	serious[10]	reporting bias[11]	73	66	-	SMD 0.15 lower (0.49 lower to 0.19 higher)	⊕OOO VERY LOW	
Admission as inpatient (follow-up 6-12 months; assessed with: Number of participants admitted into inpatient care during the study period)												
3	randomised trials	serious[12]	no serious inconsistency	serious[3]	very serious[4]	none	16/136 (11.8%)	12/145 (8.3%)	RR 1.26 (0.52 to 3.06)	22 more per 1000 (from 40 fewer to 170 more)	⊕OOO VERY LOW	
								8%		21 more per 1000 (from 38 fewer to 165 more)		
Satisfaction (follow-up 4-6 months; assessed with: Number of participants satisfied or very satisfied with their treatment)												
2	randomised trials	serious[1]	very serious[8]	serious[3]	very serious[13]	none	59/92 (64.1%)	67/106 (63.2%)	RR 1 (0.47 to 2.12)	0 fewer per 1000 (from 335 fewer to 708 more)	⊕OOO VERY LOW	
								62.8%		0 fewer per 1000 (from 333 fewer to 703 more)		

Global functioning (6-months post-admission) (follow-up mean 6 months; measured with: Global Assessment Scale (GAS; change score); Better indicated by lower values)												
1	randomised trials	very serious[14]	no serious inconsistency	serious[3]	very serious[9]	none	34	18	-	SMD 0.04 higher (0.53 lower to 0.61 higher)	⊕OOO VERY LOW	
Global functioning (12-months post-admission) (follow-up mean 12 months; measured with: Global Assessment Scale (GAS; change score); Better indicated by lower values)												
1	randomised trials	very serious[14]	no serious inconsistency	serious[3]	serious[15]	none	33	18	-	SMD 0.12 lower (0.7 lower to 0.45 higher)	⊕OOO VERY LOW	
Social functioning (4-6 months post-admission) (follow-up 4-6 months; measured with: Social Adjustment Scale-Self Report (SAS-SR; change score)/Social Functioning Scale (SFS; change score); Better indicated by lower values)												
2	randomised trials	serious[7]	no serious inconsistency	serious[3]	serious[15]	reporting bias[11]	74	67	-	SMD 0.2 lower (0.54 lower to 0.14 higher)	⊕OOO VERY LOW	
Social functioning (8-12 months post-admission) (follow-up 8-12 months; measured with: Social Adjustment Scale-Self Report (SAS-SR; change score)/Social Functioning Scale (SFS; change score); Better indicated by lower values)												
2	randomised trials	serious[7]	no serious inconsistency	serious[3]	serious[15]	reporting bias[11]	73	67	-	SMD 0.31 lower (0.65 lower to 0.03 higher)	⊕OOO VERY LOW	

[1] Unclear randomisation method and non-blind participants and intervention administrator(s)
[2] I-squared>50%
[3] Non-depression specific population
[4] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)
[5] Data cannot be extracted or is not reported for all outcomes
[6] Unclear randomisation method and non-blind participants and intervention administrator(s). Unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)
[7] Unclear randomisation method and non-blind participants and intervention administrator(s). Risk of attrition bias is unclear or high (drop-out>20% and ITT analysis not used)
[8] I-squared>80%
[9] 95% CI crosses line of no effect and threshold for both clinically important benefit (SMD -0.5) and clinically important harm (SMD 0.5)
[10] N<400
[11] Data is not reported for longest follow-up
[12] Unclear randomisation method and method of allocation concealment. Non-blind participants and intervention administrator(s) and unclear blinding of outcome assessment. Unclear risk of attriiton bias (drop-out>20%)
[13] 95% CI crosses line of no effect and threshold for both clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)
[14] Unclear randomisation method and method of allocation concealment. Non-blind participants and intervention administrator(s) and unclear blinding of outcome assessment. High risk of attrition bias as drop-out>20%, difference between groups>20% and completer analysis used
[15] 95% CI crosses both line of no effect and threshold for clinically important benefit (SMD-0.5)

Specialist depression service versus usual specialist mental health care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Specialist depression service	Usual specialist mental health care	Relative (95% CI)	Absolute		
Lost to follow-up (follow-up mean 18 months; assessed with: Number of participants lost to follow-up by the end of the study)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	31/93 (33.3%)	46/94 (48.9%)	RR 0.68 (0.48 to 0.97)	157 fewer per 1000 (from 15 fewer to 254 fewer) 156 fewer per 1000 (from 15 fewer to 254 fewer)	⊕⊕OO LOW	
								48.9%				
Self-harm (follow-up mean 18 months; assessed with: Number of participants who had episodes of self-harm during the study period)												
1	randomised trials	serious[3]	no serious inconsistency	serious	very serious[4]	none	1/93 (1.1%)	2/94 (2.1%)	RR 0.51 (0.05 to 5.48)	10 fewer per 1000 (from 20 fewer to 95 more) 10 fewer per 1000 (from 20 fewer to 94 more)	⊕OOO VERY LOW	
								2.1%				
Response (follow-up mean 18 months; assessed with: Hamilton Rating Scale for Depression (HAM-D) - definition for response not reported)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[2]	none	37/93 (39.8%)	23/94 (24.5%)	RR 1.63 (1.05 to 2.51)	154 more per 1000 (from 12 more to 369 more) 154 more per 1000 (from 12 more to 370 more)	⊕⊕OO LOW	
								24.5%				
Remission (follow-up mean 18 months; assessed with: Hamilton Rating Scale for Depression (HAM-D) - cut-off for remission not reported)												

1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[2]	none	24/93 (25.8%)	12/94 (12.8%) 12.8%	RR 2.02 (1.08 to 3.8)	130 more per 1000 (from 10 more to 357 more) 131 more per 1000 (from 10 more to 358 more)	⊕⊕OO LOW	
Depression symptomatology (follow-up mean 18 months; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[5]	none	93	94	-	SMD 0.62 lower (0.92 to 0.33 lower)	⊕⊕OO LOW	
Global functioning (follow-up mean 18 months; measured with: Global Assessment of Functioning (GAF; change score); Better indicated by lower values)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[5]	none	93	94	-	SMD 0.49 higher (0.19 to 0.78 higher)	⊕⊕OO LOW	
Social functioning (follow-up mean 18 months; measured with: Social Adjustment Scale-modified (SAS-M; change score); Better indicated by lower values)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[5]	none	93	94	-	SMD 0.46 higher (0.17 to 0.75 higher)	⊕⊕OO LOW	

[1] Non-blind participants and intervention administrator(s)
[2] Events<300
[3] Non-blind participants and intervention administrator(s). Risk of attrition bias is unclear (drop-out>20% but difference between groups<20% and ITT analysis used)
[4] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)
[5] N<400

Community mental health teams (CMHTs) versus standard care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Community mental health teams (CMHTs) versus standard care**	**Control**	**Relative (95% CI)**	**Absolute**		
Lost to follow-up (follow-up mean 3 months; assessed with: Number of participants lost to follow-up by the end of the study)												
1	randomised trials	serious[1]	no serious inconsistency	serious[2]	very serious[3]	reporting bias[4]	8/48 (16.7%)	7/52 (13.5%)	RR 1.24 (0.49 to 3.16)	32 more per 1000 (from 69 fewer to 291 more)		

								13.5%		32 more per 1000 (from 69 fewer to 292 more)	⊕OOO VERY LOW	
Death (all causes) (follow-up mean 3 months; assessed with: Number of participants who died due to any causes during the study period)												
1	randomised trials	serious[1]	no serious inconsistency	serious[2]	very serious[3]	reporting bias[4]	1/48 (2.1%)	2/52 (3.8%)	RR 0.54 (0.05 to 5.78)	18 fewer per 1000 (from 37 fewer to 184 more)	⊕OOO VERY LOW	
								3.9%		18 fewer per 1000 (from 37 fewer to 186 more)		
Symptom severity (follow-up mean 3 months; measured with: Comprehensive Psychopathological Rating Scale (CPRS) at endpoint; Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	serious[2]	serious[5]	reporting bias[4]	48	52	-	SMD 0.06 lower (0.45 lower to 0.33 higher)	⊕OOO VERY LOW	
Admission as inpatient (follow-up mean 3 months; assessed with: Number of participants admitted into inpatient care during the study period)												
1	randomised trials	serious[1]	no serious inconsistency	serious[2]	serious[6]	reporting bias[4]	7/48 (14.6%)	16/52 (30.8%)	RR 0.47 (0.21 to 1.05)	163 fewer per 1000 (from 243 fewer to 15 more)	⊕OOO VERY LOW	
								30.8%		163 fewer per 1000 (from 243 fewer to 15 more)		
Admission as inpatient for >10 days (follow-up mean 3 months; assessed with: umber of participants admitted into inpatient care for more than 10 days during the study period)												
1	randomised trials	serious[1]	no serious inconsistency	serious[2]	serious[7]	reporting bias[4]	2/48 (4.2%)	11/52 (21.2%)	RR 0.2 (0.05 to 0.84)	169 fewer per 1000 (from 34 fewer to 201 fewer)	⊕OOO VERY LOW	
								21.2%		170 fewer per 1000 (from 34 fewer to 201 fewer)		
Satisfaction (follow-up mean 3 months; assessed with: Number of participants satisfied with their treatment)												
1	randomised trials	serious[1]	no serious inconsistency	serious[2]	serious[5]	reporting bias[4]	34/41 (82.9%)	25/46 (54.3%)	RR 1.53 (1.13 to 2.06)	288 more per 1000 (from 71 more to 576 more)		

								54.4%		288 more per 1000 (from 71 more to 577 more)	⊕OOO VERY LOW	
Satisfaction (follow-up mean 3 months; measured with: Service Satisfaction Score; Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	serious[2]	serious[5]	reporting bias[4]	41	46	-	SMD 0.85 higher (0.41 to 1.29 higher)	⊕OOO VERY LOW	

[1] Unclear randomisation method and non-blind participants and intervention administrator(s)
[2] Non-depression specific population
[3] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)
[4] Data cannot be extracted for all outcomes (no measure of variance reported)
[5] N<400
[6] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 0.75)
[7] Events<300

First-line treatment (chapter 7)

NMA sub-analysis

Nortriptyline for depression in older adults

Nortriptyline versus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Nortriptyline versus placebo**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression symptomatology at endpoint (measured with: HAMD; Better indicated by lower values)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	53	56	-	MD 6.24 lower (9.17 to 3.3 lower)	⊕⊕OO LOW	CRITICAL
Depression symptomatology at endpoint - milder depression (measured with: HAMD; Better indicated by lower values)												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	12	11	-	MD 8.10 lower (13.17 to 3.03 lower)	⊕⊕OO LOW	CRITICAL
Depression symptomatology at endpoint - more severe (measured with: HAMD; Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	41	45	-	MD 5.3 lower (8.89 to 1.71 lower)	⊕⊕OO LOW	CRITICAL
Remission at endpoint - milder depression (assessed with: CGI/HAMD)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	7/12 (58.3%)	1/11 (9.1%)	RR 6.42 (0.93 to 44.16)	493 more per 1000 (from 6 fewer to 1000 more)	⊕⊕OO LOW	CRITICAL
Remission at endpoint (assessed with: CGI/HAMD)												
3	randomised trials	serious[1]	serious[4]	no serious indirectness	serious[3]	none	44/72 (61.1%)	23/76 (30.3%)	RR 2.62 (1 to 6.85)	490 more per 1000 (from 0 more to 1000 more)	⊕OOO VERY LOW	CRITICAL
								15%		243 more per 1000 (from 0 more to 877 more)		
Remission at endpoint - more severe depression (assessed with: CGI/HAMD)												
2	randomised trials	serious[1]	serious[4]	no serious indirectness	serious[3]	none	37/60 (61.7%)	22/65 (33.8%)	RR 2.14 (0.81 to 5.72)	386 more per 1000 (from 64 fewer to 1000 more)	⊕OOO VERY LOW	CRITICAL
Treatment discontinuations due to side effects - milder depression												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	none	2/25 (8%)	0/28 (0%)	RR 5.58 (0.28 to 110.89)	-	⊕OOO VERY LOW	CRITICAL
Treatment discontinuation												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	39/99 (39.4%)	29/94 (30.9%)	RR 1.25 (0.85 to 1.82)	77 more per 1000 (from 46 fewer to 253 more)	⊕⊕OO LOW	CRITICAL
								33.3%		83 more per 1000 (from 50 fewer to 273 more)		

Treatment discontinuations due to side effects - more severe depression												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[6]	none	10/38 (26.3%)	1/35 (2.9%)	RR 9.21 (1.24 to 68.31)	235 more per 1000 (from 7 more to 1000 more)	⊕⊕⊕O MODERATE	CRITICAL
Treatment discontinuations due to side effects												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[6]	none	12/63 (19%)	1/63 (1.6%)	RR 7.88 (1.49 to 41.65)	109 more per 1000 (from 8 more to 645 more)	⊕⊕OO LOW	CRITICAL
								1.4%		96 more per 1000 (from 7 more to 569 more)		

[1] High ROB in one domain and unclear in several others
[2] OIS not met (<400 participants)
[3] 95% CI crosses one clinical decision threshold
[4] I2 >50% but <80%
[5] 95% CI crosses two clinical decision thresholds
[6] OIS not met (<300 events)

Pairwise comparisons: Acupuncture

Acupuncture versus sham acupuncture

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Acupuncture versus sham acupuncture**	**Control**	**Relative (95% CI)**	**Absolute**		
Discontinuation due to side effects - Mild/moderate symptom severity (follow-up 8-12 weeks)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	1/53 (1.9%)	0/54 (0%)	RR 3.1 (0.13 to 73.12)	-	⊕OOO VERY LOW	
								0%		-		
Discontinuation for any reason - Mild/moderate symptom severity (follow-up 8-12 weeks)												

2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	7/53 (13.2%)	8/51 (15.7%)	RR 0.92 (0.24 to 3.55)	13 fewer per 1000 (from 119 fewer to 400 more) 11 fewer per 1000 (from 109 fewer to 365 more)	⊕OOO VERY LOW	
								14.3%				
Remission - Mild/moderate symptom severity (follow-up mean 8 weeks; assessed with: HAMD endpoint score of 7 or below)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[4]	none	14/25 (56%)	1/22 (4.5%)	RR 12.32 (1.76 to 86.26)	515 more per 1000 (from 35 more to 1000 more) 521 more per 1000 (from 35 more to 1000 more)	⊕⊕OO LOW	
								4.6%				
Response - Mild/moderate symptom severity (follow-up mean 8 weeks; assessed with: HAMD reduction of at least 50% from the baseline score)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[4]	none	18/25 (72%)	4/22 (18.2%)	RR 3.96 (1.58 to 9.93)	538 more per 1000 (from 105 more to 1000 more) 539 more per 1000 (from 106 more to 1000 more)	⊕⊕OO LOW	
								18.2%				
Depression symptomatology - Mild/moderate symptom severity (follow-up 8-12 weeks; measured with: HAMD; endpoint/change score; completer analysis; Better indicated by lower values)												
2	randomised trials	serious[1]	very serious[5]	no serious indirectness	very serious[6]	none	45	43	-	MD 2.86 lower (9.06 lower to 3.34 higher)	⊕OOO VERY LOW	

[1] Randomisation method and method for allocation concealment are not reported
[2] 95% CI crosses line of no effect and two clinical decision thresholds (RR 0.8 and 1.25) and events<300
[3] Allocation sequence not concealed
[4] Events<300
[5] I-squared is over 80%
[6] 95% CI crosses line of no effect and two clinical decision thresholds (SMD -0.5 and 0.5)

Acupuncture versus fluoxetine

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Acupuncture versus fluoxetine	Control	Relative (95% CI)	Absolute		
Discontinuation due to side effects - Mild/moderate symptom severity (follow-up mean 6 weeks)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	0/50 (0%)	0/25 (0%)	not pooled	not pooled	⊕⊕OO LOW	
								0%		not pooled		
Discontinuation for any reason - Mild/moderate symptom severity (follow-up mean 6 weeks)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	14/50 (28%)	0/25 (0%)	See comment	-	⊕OOO VERY LOW	
								0%		-		
Response - Mild/moderate symptom severity (follow-up mean 6 weeks; assessed with: HAMD reduction of at least 50% from the baseline score)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	27/36 (75%)	15/25 (60%)	RR 1.25 (0.86 to 1.81)	150 more per 1000 (from 84 fewer to 486 more)	⊕⊕OO LOW	
								60%		150 more per 1000 (from 84 fewer to 486 more)		
Depression symptomatology - Mild/moderate symptom severity (follow-up mean 6 weeks; measured with: HAMD; endpoint score; completer analysis; Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	36	25	-	MD 2.45 lower (4.39 to 0.51 lower)	⊕⊕OO LOW	

[1] No attempt at blinding and high risk of attrition bias
[2] Events<300
[3] 95% CI crosses a clinical decision threshold (RR 1.25) and events<300
[4] 95% CI crosses clinical decision threshold (SMD -0.5) and N<400

Acupuncture + SSRI versus SSRI

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Acupuncture + SSRI (fluoxetine/paroxetine) versus SSRI (fluoxetine/paroxetine)	Control	Relative (95% CI)	Absolute		
Discontinuation due to side effects - Moderate/severe symptom severity (follow-up mean 6 weeks)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	6/160 (3.8%)	4/95 (4.2%)	RR 0.95 (0.25 to 3.71)	2 fewer per 1000 (from 32 fewer to 114 more)	⊕OOO VERY LOW	
								4.2%		2 fewer per 1000 (from 32 fewer to 114 more)		
Discontinuation for any reason - Moderate/severe symptom severity (follow-up mean 6 weeks)												
2	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	none	14/160 (8.8%)	8/95 (8.4%)	RR 0.92 (0.39 to 2.17)	7 fewer per 1000 (from 51 fewer to 99 more)	⊕OOO VERY LOW	
								8.4%		7 fewer per 1000 (from 51 fewer to 98 more)		
Remission - Moderate/severe symptom severity (follow-up mean 6 weeks; assessed with: HAMD endpoint score of 7 or below)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	none	28/109 (25.7%)	11/48 (22.9%)	RR 1.12 (0.61 to 2.06)	28 more per 1000 (from 89 fewer to 243 more)	⊕OOO VERY LOW	
								22.9%		27 more per 1000 (from 89 fewer to 243 more)		
Response - Moderate/severe symptom severity (follow-up mean 6 weeks; assessed with: HAMD reduction of at least 50% from the baseline score)												
2	randomised trials	serious[1]	serious[4]	no serious indirectness	serious[5]	none	102/157 (65%)	43/95 (45.3%)	RR 1.37 (0.91 to 2.06)	167 more per 1000 (from 41 fewer to 480 more)	⊕OOO VERY LOW	
								45.3%		168 more per 1000 (from 41		

										fewer to 480 more)		
Depression symptomatology - Moderate/severe symptom severity (follow-up mean 6 weeks; measured with: HAMD; endpoint/change score; completer analysis; Better indicated by lower values)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[6]	none	146	87	-	SMD 0.57 lower (0.84 to 0.29 lower)	⊕⊕OO LOW	

[1] Randomisation method and method for allocation concealment not reported and no attempt at blinding participants or personnel
[2] 95% CI crosses line of no effect and both clinical decision thresholds (RR 0.8 and 1.25) and events<300
[3] No attempt at blinding participants or personnel
[4] I-squared is over 50%
[5] 95% CI crosses both line of no effect and clinical decision threshold (RR 1.25) and events<300
[6] 95% CI crosses clinical decision threshold (SMD -0.5) and N<400

Acupuncture + fluoxetine versus sham acupuncture + fluoxetine

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Acupuncture + fluoxetine versus sham acupuncture + fluoxetine	Control	Relative (95% CI)	Absolute		
Discontinuation due to side effects - Mild/moderate symptom severity (follow-up mean 3 weeks)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	5/38 (13.2%)	2/35 (5.7%)	RR 2.3 (0.48 to 11.11)	74 more per 1000 (from 30 fewer to 578 more)	⊕OOO VERY LOW	
								5.7%		74 more per 1000 (from 30 fewer to 576 more)		
Discontinuation for any reason - Mild/moderate symptom severity (follow-up mean 3 weeks)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	6/38 (15.8%)	3/35 (8.6%)	RR 1.84 (0.5 to 6.81)	72 more per 1000 (from 43 fewer to 498 more)		

								8.6%		72 more per 1000 (from 43 fewer to 500 more)	⊕OOO VERY LOW	
Depression symptomatology - Mild/moderate symptom severity (follow-up mean 3 weeks; measured with: HAMD; change score; ITT analysis; Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	36	34	-	MD 4.68 lower (7.62 to 1.74 lower)	⊕⊕OO LOW	

[1] Method of randomisation not reported and significant difference between groups at baseline in proportion of females (69.4% in intervention relative to 97.1% in control). Allocation concealment method is also not reported. Personnel also non-blind and blinding of outcome assessor not reported

[2] 95% CI crosses line of no effect and both clinical decision thresholds (RR 0.8 and 1.25) and events<300

[3] 95% CI crosses clinical decision threshold (SMD -0.5) and N<400

Acupuncture + TAU versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Acupuncture + TAU versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Discontinuation due to side effects - Mild/moderate symptom severity (follow-up mean 13 weeks)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	7/302 (2.3%)	3/151 (2%)	RR 1.17 (0.31 to 4.45)	3 more per 1000 (from 14 fewer to 69 more)	⊕OOO VERY LOW	
								2%		3 more per 1000 (from 14 fewer to 69 more)		
Discontinuation for any reason - Mild/moderate symptom severity (follow-up mean 13 weeks)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	53/302 (17.5%)	21/151 (13.9%)	RR 1.26 (0.79 to 2.01)	36 more per 1000 (from 29 fewer to 140 more)	⊕OOO VERY LOW	
								13.9%		36 more per 1000 (from 29 fewer to 140 more)		
Depression symptomatology - Mild/moderate symptom severity (follow-up mean 13 weeks; measured with: PHQ-9; endpoint score; completer analysis; Better indicated by lower values)												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	249	128	-	MD 3.3 lower (4.67 to 1.93 lower)	⊕⊕OO LOW	

[1] No attempts at blinding
[2] 95% CI crosses line of no effect and both clinical decision thresholds (RR 0.8 and 1,25)
[3] 95% CI crosses clinical decision threshold (SMD -0.5) and N<400

Acupuncture + TAU versus counselling + TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Acupuncture + TAU versus Counselling + TAU	Control	Relative (95% CI)	Absolute		
Discontinuation due to side effects - Mild/moderate symptom severity (follow-up mean 13 weeks)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	7/302 (2.3%)	2/302 (0.7%)	RR 3.5 (0.73 to 16.71)	17 more per 1000 (from 2 fewer to 104 more)	⊕OOO VERY LOW	
								0.7%		18 more per 1000 (from 2 fewer to 110 more)		
Discontinuation for any reason - Mild/moderate symptom severity (follow-up mean 13 weeks)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	53/302 (17.5%)	65/302 (21.5%)	RR 0.82 (0.59 to 1.13)	39 fewer per 1000 (from 88 fewer to 28 more)	⊕OOO VERY LOW	
								21.5%		39 fewer per 1000 (from 88 fewer to 28 more)		
Depression symptomatology - Mild/moderate symptom severity (follow-up mean 13 weeks; measured with: PHQ-9; endpoint score; completer analysis; Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	249	237	-	MD 1.5 lower (2.64 to 0.36 lower)	⊕⊕⊕O MODERATE	

[1] No attempts at blinding
[2] 95% CI crosses line of no effect and both clinical decision thresholds (RR 0.8 and 1.25)
[3] 95% CI crosses both line of no effect and clinical decision threshold (RR 0.8)

Pairwise comparisons: Behavioural couples therapy

Behavioural couples therapy versus CBT

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Behavioural couples therapy versus CBT	Control	Relative (95% CI)	Absolute		
Depression symptomatology at endpoint (across severity) (follow-up 10-78 weeks; measured with: BDI/HAMD; Better indicated by lower values)												
4	randomised trials	very serious[1]	serious[2]	no serious indirectness	serious[3]	none	67	68	-	SMD 0.03 higher (0.49 lower to 0.54 higher)	⊕OOO VERY LOW	CRITICAL
Treatment discontinuation rates (more severe depression) (follow-up mean 15 weeks; assessed with: Number of participants discontinuing for any reason)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	3/12 (25%)	3/12 (25%)	RR 1 (0.25 to 4)	0 fewer per 1000 (from 188 fewer to 750 more)	⊕OOO VERY LOW	
								25%		0 fewer per 1000 (from 188 fewer to 750 more)		
Depression symptomatology at endpoint (milder depression) (follow-up 16-78 weeks; measured with: BDI/HAMD; Better indicated by lower values)												
3	randomised trials	very serious[1]	serious[2]	no serious indirectness	very serious[4]	none	52	53	-	SMD 0.14 higher (0.49 lower to 0.78 higher)	⊕OOO VERY LOW	CRITICAL
Depression symptomatology at endpoint (more severe depression) (follow-up mean 10 weeks; measured with: BDI; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	15	15	-	SMD 0.34 lower (1.07 lower to 0.38 higher)	⊕OOO VERY LOW	CRITICAL
Remission (assessed with: BDI<10)												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	13/19 (68.4%)	16/19 (84.2%)	RR 0.81 (0.57 to 1.17)	160 fewer per 1000 (from 362 fewer to 143 more)	⊕OOO VERY LOW	CRITICAL
								0%		-		
Treatment discontinuation rates (across severity) (follow-up 15-78 weeks; assessed with: Number of participants discontinuing for any reason)												
4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	20/72 (27.8%)	9/70 (12.9%)	RR 1.97 (0.98 to 3.98)	125 more per 1000 (from 3 fewer to 383 more)	⊕⊕OO LOW	
								15.5%		150 more per 1000 (from 3 fewer to 462 more)		
Treatment discontinuation rates (milder depression) (follow-up 16-78 weeks; assessed with: Number of participants discontinuing for any reason)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	17/60 (28.3%)	6/58 (10.3%)	RR 2.49 (1.11 to 5.61)	154 more per 1000 (from 11 more to 477 more)	⊕⊕OO LOW	
								14.3%		213 more per 1000 (from 16 more to 659 more)		

[1] High or unclear ROB in most domains
[2] I2 <80% but >50%
[3] 95% confidence interval crosses one clinical decision threshold
[4] 95% CI crosses two clinical decision thresholds
[5] Events<300

Behavioural couples therapy versus waitlist

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Behavioural couples therapy versus waitlist control**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression symptomatology at endpoint (more severe depression) (follow-up mean 10 weeks; measured with: BDI; Better indicated by lower values)												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	15	15	-	MD 12.07 lower (18.32 to 5.82 lower)	⊕OOO VERY LOW	CRITICAL
Treatment discontinuation rates (more severe depression) (follow-up mean 15 weeks; assessed with: Number of participants discontinuing for any reason)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	3/12 (25%)	0/12 (0%)	RR 7 (0.4 to 122.44)	-	⊕OOO VERY LOW	
								0%		-		

[1] High or unclear ROB in most domains
[2] OIS not met (<400 participants)
[3] 95% CI crosses two clinical decision thresholds

Behavioural couples therapy versus interpersonal psychotherapy

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Behavioural couples therapy versus IPT**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression symptomatology at endpoint (milder depression) (follow-up mean 78 weeks; measured with: BDI; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	20	20	-	MD 1.56 higher (5.07 lower to 8.19 higher)	⊕OOO VERY LOW	
Treatment discontinuation rates (milder depression) (follow-up mean 78 weeks; assessed with: Number of participants discontinuing for any reason)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	2/20 (10%)	2/20 (10%)	RR 1 (0.16 to 6.42)	0 fewer per 1000 (from 84 fewer to 542 more)	⊕OOO VERY LOW	
								10%		0 fewer per 1000 (from 84 fewer to 542 more)		

[1] High or unclear ROB in most domains
[2] 95% CI crosses one clinical decision threshold
[3] Data not reported for all outcomes
[4] 95% CI crosses two clinical decision thresholds

Behavioural couples therapy (BCT) versus combined BCT and CBT (individual)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Behavioural couples therapy versus combined BCT and CBT (individual CBT for the depressed wife)	Control	Relative (95% CI)	Absolute		
Depression symptomatology at endpoint (milder depression) (measured with: HAMD; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	19	21	-	MD 4.12 higher (0.66 lower to 8.9 higher)	⊕OOO VERY LOW	
Remission (milder depression) (assessed with: BDI<10)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	13/19 (68.4%)	12/21 (57.1%)	RR 1.2 (0.74 to 1.94)	114 more per 1000 (from 149 fewer to 537 more)	⊕OOO VERY LOW	
								57.1%		114 more per 1000 (from 148 fewer to 537 more)		
Treatment discontinuation rates (milder depression) (assessed with: Number of participants discontinuing for any reason)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	8/27 (29.6%)	0/21 (0%)	RR 13.36 (0.81 to 218.99)	-	⊕⊕OO LOW	
								0%		-		

[1] High or unclear ROB in most domains
[2] 95% CI crosses one clinical decision threshold
[3] 95% CI crosses two clinical decision thresholds

Pairwise comparisons: Omega-3 fatty acids

Omega-3 fatty acids versus placebo

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Omega-3 fatty acids versus placebo	Control	Relative (95% CI)	Absolute		
Remission (milder depression) (follow-up 3-8 weeks; assessed with: BDI=>10 or HAMD <=7 at endpoint)												
2	randomised trials	no serious risk of bias	serious[1]	no serious indirectness	very serious[2]	reporting bias[3]	44/143 (30.8%)	21/74 (28.4%)	RR 1.43 (0.48 to 4.29)	122 more per 1000 (from 148 fewer to 934 more)	⊕OOO VERY LOW	
								25.7%		111 more per 1000 (from 134 fewer to 846 more)		
Response (milder depression) (follow-up mean 8 weeks; assessed with: HAMD reduced by >50% at endpoint)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	52/131 (39.7%)	28/65 (43.1%)	RR 0.92 (0.65 to 1.31)	34 fewer per 1000 (from 151 fewer to 134 more)	⊕OOO VERY LOW	
								43.1%		34 fewer per 1000 (from 151 fewer to 134 more)		
Treatment discontinuation (milder depression) (follow-up 3-8 weeks; assessed with: Number of participants discontinuing for any reason)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	reporting bias[3]	16/144 (11.1%)	13/75 (17.3%)	RR 0.63 (0.32 to 1.24)	64 fewer per 1000 (from 118 fewer to 42 more)	⊕⊕OO LOW	
								14.2%		53 fewer per 1000 (from 97 fewer to 34 more)		
Discontinuation due to side effects (milder depression) (follow-up mean 8 weeks; assessed with: Number of participants discontinuing due to side effects)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	1/131 (0.8%)	0/65 (0%)	RR 1.5 (0.06 to 36.32)	-	⊕OOO VERY LOW	
								0%		-		

[1] I-squared >50%
[2] 95% CI crosses two clinical decision thresholds
[3] Data not reported for all outcomes
[4] 95% CI crosses one clinical decision threshold

Omega-3 fatty acids plus SSRI/antidepressant versus placebo plus SSRI/antidepressant

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Omega-3 fatty acids + SSRI/antidepressants versus placebo + SSRI/antidepressants	Control	Relative (95% CI)	Absolute		
Remission (more severe depression) (follow-up mean 8 weeks; assessed with: HAMD <=7 at endpoint)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	8/18 (44.4%)	4/22 (18.2%)	RR 2.44 (0.88 to 6.82)	262 more per 1000 (from 22 fewer to 1000 more)	⊕OOO VERY LOW	
								18.2%		262 more per 1000 (from 22 fewer to 1000 more)		
Response (more severe depression) (follow-up mean 8 weeks; assessed with: HAMD reduced by >50% at endpoint)												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	13/16 (81.3%)	8/16 (50%)	RR 1.62 (0.94 to 2.8)	310 more per 1000 (from 30 fewer to 900 more)	⊕OOO VERY LOW	
								50%		310 more per 1000 (from 30 fewer to 900 more)		
Treatment discontinuation (across severity) (follow-up 8-12 weeks; assessed with: Number of participants discontinuing for any reason)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	13/58 (22.4%)	16/59 (27.1%)	RR 0.85 (0.44 to 1.63)	41 fewer per 1000 (from 152 fewer to 171 more)	⊕OOO VERY LOW	
								29.4%		44 fewer per 1000 (from 165 fewer to 185 more)		
Treatment discontinuation (milder depression) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing for any reason)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	6/18 (33.3%)	5/17 (29.4%)	RR 1.13 (0.42 to 3.03)	38 more per 1000 (from 171 fewer to 597 more)		

								29.4%		38 more per 1000 (from 171 fewer to 597 more)	⊕OOO VERY LOW	
Treatment discontinuation (more severe depression) (follow-up mean 8 weeks; assessed with: Number of participants discontinuing for any reason)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	7/40 (17.5%)	11/42 (26.2%)	RR 0.68 (0.29 to 1.62)	84 fewer per 1000 (from 186 fewer to 162 more)	⊕OOO VERY LOW	
								25.9%		83 fewer per 1000 (from 184 fewer to 161 more)		
Discontinuation due to side effects (more severe depression) (follow-up mean 8 weeks; assessed with: Number of participants discontinuing due to side effects)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	2/40 (5%)	1/42 (2.4%)	RR 2 (0.2 to 20.33)	24 more per 1000 (from 19 fewer to 460 more)	⊕OOO VERY LOW	
								2.5%		25 more per 1000 (from 20 fewer to 483 more)		

[1] High or unclear risk in multiple ROB domains
[2] 95% CI crosses one clinical decision threshold
[3] Data not reported for all outcomes
[4] Unclear risk across multiple ROB domains
[5] 95% CI crosses two clinical decision thresholds

Pairwise comparisons: Psychosocial interventions (peer support)

Peer support versus waitlist

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Peer support group versus waitlist**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression symptoms at endpoint (milder depression) (follow-up mean 4 weeks; measured with: BDI; Better indicated by lower values)												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	19	67	-	MD 7.09 lower (9.77 to 4.41 lower)	⊕OOO VERY LOW	

[1] Unclear allocation concealment and non-blind participants, intervention administrators and outcome assessment
[2] N<400
[3] Data is not reported or cannot be extracted for all outcomes

Peer support (online support group) versus attention-placebo control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Peer support (online support group) versus attention control**	**Control**	**Relative (95% CI)**	**Absolute**		
Treatment discontinuation (milder depression) (follow-up mean 12 weeks; assessed with: Number of participants who discontinued for any reason)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias[2]	36/89 (40.4%)	11/82 (13.4%)	RR 3.02 (1.65 to 5.52)	271 more per 1000 (from 87 more to 606 more)	⊕⊕OO LOW	
								13.4%		271 more per 1000 (from 87 more to 606 more)		

[1] Events<300
[2] Data is not reported or cannot be extracted for all outcomes

Peer support group versus CBT group

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Peer support group versus CBT group**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression symptoms at endpoint (milder depression) (follow-up mean 4 weeks; measured with: BDI; Better indicated by lower values)												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	19	50	-	MD 1.72 lower (4.8 lower to 1.36 higher)	⊕OOO VERY LOW	

[1] Unclear allocation concealment and non-blind participants, intervention administrators and outcome assessment
[2] 95% CI crosses one clinical decision threshold
[3] Data is not reported or cannot be extracted for all outcomes

Peer support group versus self-help (without support)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Peer support group versus self-help (without support)**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression symptoms at endpoint (milder depression) (follow-up mean 4 weeks; measured with: BDI; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	19	28	-	MD 2.87 lower (6.53 lower to 0.79 higher)	⊕OOO VERY LOW	

[1] Unclear allocation concealment and non-blind participants, intervention administrators and outcome assessment
[2] 95% CI crosses one clinical decision threshold
[3] Data is not reported or cannot be extracted for all outcomes

Light therapy

Is bright light effective for depression with a seasonal pattern/SAD compared with waitlist control?

Quality assessment							Summary of findings					
							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Limitations**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Bright light**	**Waitlist**	**Relative (95% CI)**	**Absolute**		
Leaving study early for any reason (overall) (total number not completing study)												

2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[1]	none	3/42 (7.1%)	3/40 (7.5%) 8.7%	RR 0.95 (0.21 to 4.32)	0 fewer per 100 (from 6 fewer to 25 more) 0 fewer per 100 (from 7 fewer to 29 more)	⊕⊕OO LOW	
Leaving study early due to side effects - Light box vs waitlist control												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	0/16 (0%)	0/15 (0%) 0%	not pooled	not pooled not pooled	⊕⊕⊕O MODERATE	
Leaving study early - Light room vs waitlist control												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	1/26 (3.8%)	1/25 (4%) 0%	RR 0.96 (0.06 to 14.55)	0 fewer per 100 (from 4 fewer to 54 more) 0 fewer per 100 (from 0 fewer to 0 more)	⊕⊕⊕O MODERATE	
Mean self rated SAD depression scores at endpoint - Light room vs waitlist control (measured with: SIGH-SAD-SR; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	24	24	-	MD 12.8 lower (18.52 to 7.08 lower)	⊕⊕⊕O MODERATE	
Mean clinician rated SAD depression scores at endpoint - Light box vs waitlist control (measured with: SIGH-SAD; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	16	15	-	MD 10.4 lower (15.99 to 4.81 lower)	⊕⊕⊕O MODERATE	
Mean clinician rated typical depression scores at endpoint - Light box vs waitlist control (measured with: HRSD-21; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	16	15	-	MD 6.3 lower (10.34 to 2.26 lower)	⊕⊕⊕⊕ HIGH	
Mean self-rated depression score - overall (Better indicated by lower values)												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	40	39	-	MD 1.15 lower (1.63 to 0.67 lower)	⊕⊕⊕⊕ HIGH	
Mean self rated depression scores at endpoint - Light room vs waitlist control (measured with: HRSD-21-SR; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	24	24	-	MD 7.7 lower (11.58 to 3.82 lower)	⊕⊕⊕O MODERATE	

Mean self rated depression scores at endpoint - Light box vs waitlist control (measured with: BDI; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	16	15	-	MD 10.9 lower (16.99 to 4.81 lower)	⊕⊕⊕O MODERATE	
Mean clinician rated atypical depression scores at endpoint - Light box vs waitlist control (measured with: SAD subscale; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	16	15	-	MD 4 lower (6.73 to 1.27 lower)	⊕⊕⊕O MODERATE	
Mean self rated atypical depression scores at endpoint - Light room vs waitlist control (measured with: SAD-SR subscale of SIGH-SAD); Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	24	24	-	MD 5.2 lower (7.39 to 3.01 lower)	⊕⊕⊕O MODERATE	
Non remission (SIGH-SAD-SR) (overall)												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	20/42 (47.6%)	36/40 (90%) 88%	RR 0.53 (0.38 to 0.74)	42 fewer per 100 (from 23 fewer to 56 fewer) 41 fewer per 100 (from 23 fewer to 55 fewer)	⊕⊕⊕⊕ HIGH	
Non remission (SIGH-SAD-SR) - Light room vs waitlist control												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	12/26 (46.2%)	24/25 (96%) 96%	RR 0.48 (0.31 to 0.73)	50 fewer per 100 (from 26 fewer to 66 fewer) 50 fewer per 100 (from 26 fewer to 66 fewer)	⊕⊕⊕O MODERATE	
Non remission (SIGH-SAD-SR) - Light box vs waitlist control												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	8/16 (50%)	12/15 (80%) 80%	RR 0.62 (0.36 to 1.08)	30 fewer per 100 (from 51 fewer to 6 more) 30 fewer per 100 (from 51 fewer to 6 more)	⊕⊕⊕O MODERATE	
Non response (SIGH-SAD) - Light room vs waitlist control												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	13/26 (50%)	25/25 (100%) 100%	RR 0.50 (0.34 to 0.73)	50 fewer per 100 (from 27 fewer to 66 fewer) 50 fewer per 100 (from 27 fewer to 66 fewer)	⊕⊕⊕O MODERATE	

[1] Inconclusive effect size
[2] Single study

Is bright light effective for depression with a seasonal pattern/SAD compared with attentional control?

Quality assessment							Summary of findings					
							No of patients		Effect		Quality	Importance
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Bright light	Attentional control	Relative (95% CI)	Absolute		
Leaving study early for any reason (overall)												
5	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[1]	none	18/134 (13.4%)	18/124 (14.5%) 13.1%	RR 0.92 (0.51 to 1.64)	1 fewer per 100 (from 7 fewer to 9 more) 1 fewer per 100 (from 6 fewer to 8 more)	⊕⊕OO LOW	
Leaving study early for any reason - Light box vs deactivated negative ion generator												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	8/41 (19.5%)	9/40 (22.5%) 22.5%	RR 0.87 (0.37 to 2.02)	3 fewer per 100 (from 14 fewer to 23 more) 3 fewer per 100 (from 14 fewer to 23 more)	⊕⊕OO LOW	
Leaving study early for any reason - Low dose (<5000lux hours/day) LED light vs negative ion generator												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	1/15 (6.7%)	2/11 (18.2%) 18.2%	RR 0.37 (0.04 to 3.55)	11 fewer per 100 (from 17 fewer to 46 more) 11 fewer per 100 (from 17 fewer to 46 more)	⊕⊕OO LOW	
Leaving study early for any reason - Light box vs high dose (>300lux) dim red light box												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	6/33 (18.2%)	5/26 (19.2%) 19.2%	RR 0.95 (0.32 to 2.76)	1 fewer per 100 (from 13 fewer to 34 more) 1 fewer per 100 (from 13 fewer to 34 more)	⊕⊕OO LOW	
Leaving study early for any reason - Light box vs low-density ionisation												

1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	2/23 (8.7%)	2/25 (8%) 8%	RR 1.09 (0.17 to 7.1)	1 more per 100 (from 7 fewer to 49 more) 1 more per 100 (from 7 fewer to 49 more)	⊕⊕OO LOW	
Leaving study early for any reason - Low dose (<5000lux hours/day) light box vs no light box												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	1/10 (10%)	0/12 (0%) 0%	RR 3.55 (0.16 to 78.56)	0 more per 100 (from 0 fewer to 0 more) 0 more per 100 (from 0 fewer to 0 more)	⊕⊕OO LOW	
Leaving study early for any reason - Low dose (<5000lux hours/day) light visor vs no light visor												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	0/12 (0%)	0/10 (0%) 0%	not pooled	not pooled not pooled	⊕⊕⊕O MODERATE	
Leaving study early due to lack of efficacy - Low dose (<5000lux hours/day) LED light vs negative ion generator												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	0/15 (0%)	1/11 (9.1%) 9.1%	RR 0.25 (0.01 to 5.62)	7 fewer per 100 (from 9 fewer to 42 more) 7 fewer per 100 (from 9 fewer to 42 more)	⊕⊕OO LOW	
Reported side effects (overall)												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	25/45 (55.6%)	21/36 (58.3%) 44.6%	RR 0.98 (0.73 to 1.32)	1 fewer per 100 (from 16 fewer to 19 more) 1 fewer per 100 (from 12 fewer to 14 more)	⊕⊕OO LOW	
Reported side effects - Low dose (<5000lux hours/day) LED light vs negative ion generator												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	2/15 (13.3%)	1/11 (9.1%) 9.1%	RR 1.47 (0.15 to 14.21)	4 more per 100 (from 8 fewer to 120 more) 4 more per 100 (from 8 fewer to 120 more)	⊕⊕⊕O MODERATE	
Reported side effects - Light visor vs dim light visor												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious	none	23/30 (76.7%)	20/25 (80%)	RR 0.96 (0.73 to 1.27)	3 fewer per 100 (from 22 fewer to 22 more)		

								80%		3 fewer per 100 (from 22 fewer to 22 more)	⊕⊕OO LOW	
Mean clinician rated SAD depression scores at endpoint (overall) (measured with: SIGH-SAD; Better indicated by lower values)												
6	randomised trials	no serious limitations	serious[3]	no serious indirectness	serious[1]	none	139	131	-	MD 2.78 lower (6.81 lower to 1.26 higher)	⊕⊕OO LOW	
Mean clinician rated SAD depression scores at endpoint - Low dose (<5000lux hours/day) LED light vs negative ion generator (measured with: SIGH-SAD; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[3]	none	14	9	-	MD 4.7 lower (10.34 lower to 0.94 higher)	⊕⊕⊕O MODERATE	
Mean clinician rated SAD depression scores at endpoint - Light visor vs dim light visor (measured with: SIGH-SAD; Better indicated by lower values)												
2	randomised trials	no serious limitations	serious[1]	no serious indirectness	serious[3]	none	64	58	-	MD 0.86 higher (7.56 lower to 9.29 higher)	⊕⊕OO LOW	
Mean clinician rated SAD depression scores at endpoint - Light box vs low-density ionisation (measured with: SIGH-SAD; Better indicated by lower values)												
2	randomised trials	no serious limitations	serious[2]	no serious indirectness	no serious imprecision	none	40	42	-	MD 8.56 lower (14.73 to 2.39 lower)	⊕⊕⊕O MODERATE	
Mean clinician rated SAD depression scores at endpoint - Low dose (<5000lux hours/day) light box vs no light box (measured with: SIGH-SAD; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	9	12	-	MD 1.4 higher (4.93 lower to 7.73 higher)	⊕⊕OO LOW	
Mean clinician rated SAD depression scores at endpoint - Low dose (<5000lux hours/day) light visor vs no light visor (measured with: SIGH-SAD; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[3]	none	12	10	-	MD 0.2 lower (6.22 lower to 5.82 higher)	⊕⊕OO LOW	
Mean clinician rated typical depression scores at endpoint (measured with: HAMD-17/HRSD-21; Better indicated by lower values)												
5	randomised trials	no serious limitations	serious[1]	no serious indirectness	serious[1]	none	106	103	-	SMD 0.07 lower (0.51 lower to 0.37 higher)	⊕⊕OO LOW	
Mean clinician rated typical depression scores at endpoint - Light visor vs dim light visor (measured with: HAMD-17/HRSD-21; Better indicated by lower values)												
2	randomised trials	no serious limitations	serious[3]	no serious indirectness	serious[4]	none	64	58	-	SMD 0.05 higher (0.52 lower to 0.63 higher)	⊕⊕OO LOW	

Mean clinician rated typical depression scores at endpoint - Light box vs low-density ionisation (measured with: HAMD-17/HRSD-21; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[4]	none	21	23	-	SMD 0.81 lower (1.43 to 0.19 lower)	⊕⊕⊕O MODERATE	
Mean clinician rated typical depression scores at endpoint - Low dose (<5000lux hours/day) light box vs no light box (measured with: HAMD-17/HRSD-21; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[4]	none	9	12	-	SMD 0.26 higher (0.61 lower to 1.13 higher)	⊕⊕⊕O MODERATE	
Mean clinician rated typical depression scores at endpoint - Low dose (<5000lux hours/day) light visor vs no light visor (measured with: HAMD-17/HRSD-21; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[4]	none	12	10	-	SMD 0.2 higher (0.64 lower to 1.04 higher)	⊕⊕⊕O MODERATE	
Mean clinician rated atypical depression scores at endpoint (measured with: SAD subscale; Better indicated by lower values)												
3	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	55	55	-	MD 1.25 lower (2.77 lower to 0.27 higher)	⊕⊕⊕⊕ HIGH	
Mean clinician rated atypical depression scores at endpoint - Light visor vs dim light visor (measured with: SAD subscale; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[4]	none	34	33	-	MD 2.1 lower (4.31 lower to 0.11 higher)	⊕⊕⊕O MODERATE	
Mean clinician rated atypical depression scores at endpoint - Low dose (<5000lux hours/day) light box vs no light box (measured with: SAD subscale; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[4]	none	9	12	-	MD 1.2 higher (2.48 lower to 4.88 higher)	⊕⊕⊕O MODERATE	
Mean clinician rated atypical depression scores at endpoint - Low dose (<5000lux hours/day) light visor vs no light visor (measured with: SAD subscale; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[4]	none	12	10	-	MD 1.3 lower (3.84 lower to 1.24 higher)	⊕⊕⊕O MODERATE	
Mean self rated depression scores at endpoint - Light box vs deactivated negative ion generator (measured with: BDI; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[3]	none	33	31	-	MD 2.6 lower (6.72 lower to 1.52 higher)	⊕⊕⊕O MODERATE	
Non remission (SIGH-SAD or SIGH-SAD-SR or HDRS) (overall)												

6	randomised trials	no serious limitations	serious[1]	no serious indirectness	serious[3]	none	99/176 (56.3%)	98/160 (61.3%) 70.5%	RR 0.89 (0.66 to 1.2)	7 fewer per 100 (from 21 fewer to 12 more) 8 fewer per 100 (from 24 fewer to 14 more)	⊕⊕OO LOW	
Non remission (SIGH-SAD or SIGH-SAD-SR or HDRS) - Low dose (<5000lux hours/day) LED light vs negative ion generator												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[3]	none	7/15 (46.7%)	10/11 (90.9%) 90.9%	RR 0.51 (0.29 to 0.91)	45 fewer per 100 (from 8 fewer to 65 fewer) 45 fewer per 100 (from 8 fewer to 65 fewer)	⊕⊕⊕O MODERATE	
Non remission (SIGH-SAD or SIGH-SAD-SR or HDRS) - Light box vs deactivated negative ion generator												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[3]	none	21/41 (51.2%)	30/40 (75%) 75%	RR 0.68 (0.48 to 0.97)	24 fewer per 100 (from 2 fewer to 39 fewer) 24 fewer per 100 (from 2 fewer to 39 fewer)	⊕⊕⊕O MODERATE	
Non remission (SIGH-SAD or SIGH-SAD-SR or HDRS) - Light visor vs dim light visor												
2	randomised trials	no serious limitations	serious[1]	no serious indirectness	serious[4]	none	33/64 (51.6%)	22/58 (37.9%) 38.7%	RR 1.34 (0.79 to 2.27)	13 more per 100 (from 8 fewer to 48 more) 13 more per 100 (from 8 fewer to 49 more)	⊕⊕OO LOW	
Non remission (SIGH-SAD or SIGH-SAD-SR or HDRS) - Light box vs high dose (>300lux) dim red light box												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	25/33 (75.8%)	19/26 (73.1%) 73.1%	RR 1.04 (0.77 to 1.4)	3 more per 100 (from 17 fewer to 29 more) 3 more per 100 (from 17 fewer to 29 more)	⊕⊕OO LOW	
Non remission (SIGH-SAD or SIGH-SAD-SR or HDRS) - Light box vs low-density ionisation												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	13/23 (56.5%)	17/25 (68%) 68%	RR 0.83 (0.53 to 1.3)	12 fewer per 100 (from 32 fewer to 20 more) 12 fewer per 100 (from 32 fewer to 20 more)	⊕⊕OO LOW	
Non response (SIGH-SAD) (overall)												

7	randomised trials	no serious limitations	serious[3]	no serious indirectness	serious[1]	none	83/183 (45.4%)	92/171 (53.8%) 58.3%	RR 0.86 (0.64 to 1.15)	8 fewer per 100 (from 19 fewer to 8 more) 8 fewer per 100 (from 21 fewer to 9 more)	⊕⊕OO LOW	
Non response (SIGH-SAD) - Light box vs deactivated negative ion generator												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[4]	none	19/41 (46.3%)	25/40 (62.5%) 62.5%	RR 0.74 (0.49 to 1.11)	16 fewer per 100 (from 32 fewer to 7 more) 16 fewer per 100 (from 32 fewer to 7 more)	⊕⊕⊕O MODERATE	
Non response (SIGH-SAD) - Light visor vs dim light visor												
2	randomised trials	no serious limitations	serious[3]	no serious indirectness	serious[4]	none	30/64 (46.9%)	22/58 (37.9%) 37.2%	RR 1.24 (0.56 to 2.75)	9 more per 100 (from 17 fewer to 66 more) 9 more per 100 (from 16 fewer to 65 more)	⊕⊕OO LOW	
Non response (SIGH-SAD) - Light box vs high dose (>300lux) dim red light box												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[4]	none	13/33 (39.4%)	14/26 (53.8%) 53.9%	RR 0.73 (0.42 to 1.27)	15 fewer per 100 (from 31 fewer to 15 more) 15 fewer per 100 (from 31 fewer to 15 more)	⊕⊕⊕O MODERATE	
Non response (SIGH-SAD) - Light box vs low-density ionisation												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[4]	none	9/23 (39.1%)	18/25 (72%) 72%	RR 0.54 (0.31 to 0.96)	33 fewer per 100 (from 3 fewer to 50 fewer) 33 fewer per 100 (from 3 fewer to 50 fewer)	⊕⊕⊕O MODERATE	
Non response (SIGH-SAD) - Low dose (<5000lux hours/day) light box vs no light box												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[4]	none	7/10 (70%)	7/12 (58.3%) 58.3%	RR 1.2 (0.64 to 2.25)	12 more per 100 (from 21 fewer to 73 more) 12 more per 100 (from 21 fewer to 73 more)	⊕⊕⊕O MODERATE	
Non response (SIGH-SAD) - Low dose (<5000lux hours/day) light visor vs no light visor												

1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[4]	none	5/12 (41.7%)	6/10 (60%) 60%	RR 0.69 (0.3 to 1.61)	19 fewer per 100 (from 42 fewer to 37 more) 19 fewer per 100 (from 42 fewer to 37 more)	⊕⊕⊕O MODERATE	

[1] Inconclusive effect size
[2] Single study; inconclusive effect size
[3] Significant heterogeneity; random effects model used
[4] Single study

Is bright light effective for depression with a seasonal pattern/SAD compared with active treatments?

Quality assessment							Summary of findings					Importance
							No of patients		Effect		Quality	
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Bright light	Active treatment control	Relative (95% CI)	Absolute		
Leaving study early for any reason - Light box vs group CBT												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[1]	none	2/25 (8%)	4/24 (16.7%) 17.8%	RR 0.53 (0.12 to 2.31)	8 fewer per 100 (from 15 fewer to 22 more) 8 fewer per 100 (from 16 fewer to 23 more)	⊕⊕⊕O MODERATE	
Leaving study early for any reason - Light box + placebo pill vs dim light box + fluoxetine												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[1]	none	12/68 (17.6%)	8/68 (11.8%) 9.8%	RR 1.5 (0.65 to 3.44)	6 more per 100 (from 4 fewer to 29 more) 5 more per 100 (from 3 fewer to 24 more)	⊕⊕⊕O MODERATE	
Leaving study early for any reason - Light box + hypericum vs dim light + hypericum												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	0/10 (0%)	0/10 (0%) 0%	not pooled	not pooled not pooled	⊕⊕⊕⊕ HIGH	
Leaving study early due to side effects - Light box + placebo pill vs dim light box + fluoxetine												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	1/48 (2.1%)	2/48 (4.2%)	RR 0.5 (0.05 to 5.33)	2 fewer per 100 (from 4 fewer to 18 more)		

								4.2%		2 fewer per 100 (from 4 fewer to 18 more)	⊕⊕OO LOW	
Leaving study early due to side effects - Light box vs group CBT												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[3]	none	0/16 (0%)	0/15 (0%) 0%	not pooled	not pooled not pooled	⊕⊕⊕O MODERATE	
Leaving study early due to lack of efficacy - Light box + placebo pill vs dim light box + fluoxetine												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	2/43 (4.7%)	0/48 (0%) 0%	RR 5.57 (0.27 to 112.85)	0 more per 100 (from 0 fewer to 0 more) 0 more per 100 (from 0 fewer to 0 more)		
Reported side effects - Light box + placebo pill vs dim light box + fluoxetine												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[3]	none	37/48 (77.1%)	75%	RR 1.03 (0.82 to 1.29)	22 more per 1000 (from 135 fewer to 217 more)		
Mean clinician rated SAD depression scores at endpoint - Light box vs group CBT (measured with: SIGH-SAD; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	16	15	-	MD 0.2 lower (6.5 lower to 6.1 higher)	⊕⊕OO LOW	
Mean clinician rated SAD depression scores at endpoint - Light box + placebo pill vs dim light box + fluoxetine (measured with: SIGH-SAD; Better indicated by lower values)												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	68	68	-	MD 0.49 lower (3.72 lower to 2.74 higher)	⊕⊕⊕⊕ HIGH	
Mean clinician rated typical depression scores at endpoint - Light box vs group CBT (measured with: HAMD-17/HRSD-21; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	16	15	-	SMD 0.13 lower (0.83 lower to 0.58 higher)	⊕⊕OO LOW	
Mean clinician rated typical depression scores at endpoint - Light box + placebo pill vs dim light box + fluoxetine (measured with: HAMD-17/HRSD-21; Better indicated by lower values)												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	68	68	-	SMD 0.04 lower (0.38 lower to 0.29 higher)	⊕⊕⊕⊕ HIGH	
Mean clinician rated typical depression scores at endpoint - Light box + hypericum vs dim light + hypericum (measured with: HAMD-17/HRSD-21; Better indicated by lower values)												

1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	10	10	-	SMD 0.32 lower (1.2 lower to 0.57 higher)	⊕⊕OO LOW	
Mean clinician rated atypical depression scores at endpoint - Light box vs group CBT (measured with: SAD subscale; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[3]	none	16	15	-	MD 0.4 higher (2.68 lower to 3.48 higher)	⊕⊕⊕O MODERATE	
Mean clinician rated atypical depression scores at endpoint - Light box + placebo pill vs dim light box + fluoxetine (measured with: SAD subscale; Better indicated by lower values)												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[1]	none	68	68	-	MD 0.3 lower (1.75 lower to 1.15 higher)	⊕⊕OO LOW	
Mean self rated depression scores at endpoint - Light box vs group CBT (measured with: BDI; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	16	15	-	MD 0.7 lower (7.16 lower to 5.76 higher)		
Mean self rated depression scores at endpoint - Light box + placebo pill vs dim light box + fluoxetine (measured with: BDI; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	48	48	-	MD 1.6 lower (5.68 lower to 2.48 higher)	⊕⊕OO LOW	
Non remission - Light box + placebo pill vs dim light box + fluoxetine												
2	randomised trials	no serious limitations	serious[4]	no serious indirectness	serious[1]	none	34/68 (50%)	37/68 (54.4%) 60.4%	RR 0.92 (0.67 to 1.27)	4 fewer per 100 (from 18 fewer to 15 more) 5 fewer per 100 (from 20 fewer to 16 more)	⊕⊕OO LOW	
Non remission - Light box vs group CBT												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	12/25 (48%)	15/24 (62.5%) 63.3%	RR 0.77 (0.46 to 1.28)	14 fewer per 100 (from 34 fewer to 17 more) 15 fewer per 100 (from 34 fewer to 18 more)	⊕⊕⊕⊕ HIGH	
Non response - Light box + placebo pill vs dim light box + fluoxetine												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[1]	none	22/68 (32.4%)	23/68 (33.8%)	RR 0.96 (0.59 to 1.54)	1 fewer per 100 (from 14 fewer to 18 more)		

								34.2%		1 fewer per 100 (from 14 fewer to 18 more)	⊕⊕OO LOW	

[1] Inconclusive effect size
[2] Inconclusive effect size/single study
[3] Single study
[4] Significant heterogeneity; random effects model used

Is bright light effective for depression with a seasonal pattern/SAD compared with a combination of bright light and CBT?

Quality assessment							Summary of findings					Importance
							No of patients		Effect		Quality	
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Bright light	Light + CBT combo	Relative (95% CI)	Absolute		
Leaving study early for any reason												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[1]	none	2/25 (8%)	2/23 (8.7%) 9.6%	RR 0.92 (0.17 to 4.91)	1 fewer per 100 (from 7 fewer to 34 more) 1 fewer per 100 (from 8 fewer to 38 more)	⊕⊕⊕O MODERATE	
Leaving study early due to side effects												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	0/16 (0%)	1/15 (6.7%) 6.7%	RR 0.31 (0.01 to 7.15)	5 fewer per 100 (from 7 fewer to 41 more) 5 fewer per 100 (from 7 fewer to 41 more)	⊕⊕OO LOW	
Mean clinician rated SAD depression scores at endpoint (measured with: SIGH-SAD; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[3]	none	16	15	-	MD 4.2 higher (0.52 lower to 8.92 higher)	⊕⊕⊕O MODERATE	
Mean clinician rated typical depression scores at endpoint (measured with: HAMD-17/HRSD-21; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	16	15	-	SMD 0.46 higher (0.26 lower to 1.17 higher)	⊕⊕⊕O MODERATE	
Mean clinician rated atypical depression scores at endpoint (measured with: SAD subscale; Better indicated by lower values)												

1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[3]	none	16	15	-	MD 2 higher (0.12 lower to 4.12 higher)	⊕⊕⊕O MODERATE	
Mean self rated depression scores at endpoint (measured with: BDI; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	16	15	-	MD 2.3 higher (2.47 lower to 7.07 higher)	⊕⊕OO LOW	
Non remission (SIGH-SAD)												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	12/25 (48%)	5/23 (21.7%) 19.6%	RR 2.22 (0.92 to 5.32)	27 more per 100 (from 2 fewer to 94 more) 24 more per 100 (from 2 fewer to 85 more)	⊕⊕⊕⊕ HIGH	

[1] Inconclusive effect size
[2] Inconclusive effect size; single study
[3] Single study

Does the time of day increase the effectiveness of bright light box therapy?

Quality assessment							Summary of findings					Importance
							No of patients		Effect		Quality	
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Morning	Afternoon/evening bright light box	Relative (95% CI)	Absolute		
Leaving study early for any reason (overall)												
3	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[1]	none	8/66 (12.1%)	8/64 (12.5%) 0%	RR 0.98 (0.41 to 2.35)	0 fewer per 100 (from 7 fewer to 17 more) 0 fewer per 100 (from 0 fewer to 0 more)	⊕⊕⊕O MODERATE	
Leaving study early for any reason - SAD												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[1]	none	8/50 (16%)	8/49 (16.3%) 10%	RR 0.98 (0.41 to 2.35)	0 fewer per 100 (from 10 fewer to 22 more) 0 fewer per 100 (from 6 fewer to 13 more)	⊕⊕⊕O MODERATE	

Leaving study early for any reason - Subsyndromal SAD												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	0/16 (0%)	0/15 (0%) 0%	not pooled	not pooled not pooled	⊕⊕⊕O MODERATE	
Leaving study early due to side effects - Subsyndromal SAD												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	0/16 (0%)	0/15 (0%) 0%	not pooled	not pooled not pooled	⊕⊕⊕O MODERATE	
Reported side effects - Subsyndromal SAD												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[3]	none	1/16 (6.3%)	2/15 (13.3%) 13.3%	RR 0.47 (0.05 to 4.65)	7 fewer per 100 (from 13 fewer to 49 more) 7 fewer per 100 (from 13 fewer to 49 more)	⊕⊕OO LOW	
Mean clinician rated SAD depression scores at endpoint (overall) (measured with: SIGH-SAD; Better indicated by lower values)												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[1]	none	35	33	-	MD 1.38 lower (5.49 lower to 2.73 higher)	⊕⊕OO LOW	
Mean clinician rated SAD depression scores at endpoint - Subsyndromal SAD (measured with: SIGH-SAD; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[3]	none	16	14	-	MD 0.6 higher (3.89 lower to 5.09 higher)	⊕⊕OO LOW	
Mean clinician rated SAD depression scores at endpoint - SAD (measured with: SIGH-SAD; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[3]	none	19	19	-	MD 3.6 lower (8.5 lower to 1.3 higher)	⊕⊕OO LOW	
Mean clinician rated typical depression scores at endpoint (overall) (measured with: HAMD-17/HRSD-31; Better indicated by lower values)												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[1]	none	25	22	-	SMD 0.05 lower (0.63 lower to 0.52 higher)	⊕⊕⊕O MODERATE	
Mean clinician rated typical depression scores at endpoint - Subsyndromal SAD (measured with: HAMD-17/HRSD-21; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[3]	none	16	14	-	SMD 0.15 lower (0.87 lower to 0.57 higher)	⊕⊕OO LOW	

Mean clinician rated typical depression scores at endpoint - SAD (HRSD-31) (measured with: HAMD-17/HRSD-21; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[3]	none	9	8	-	SMD 0.12 higher (0.83 lower to 1.07 higher)	⊕⊕OO LOW	
Mean clinician rated atypical depression scores at endpoint - Subsyndromal SAD (measured with: SAD subscale; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[3]	none	16	14	-	MD 1 higher (1.72 lower to 3.72 higher)	⊕⊕OO LOW	
Mean self rated depression scores at endpoint - SAD (measured with: BDI; Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[3]	none	33	32	-	MD 0.9 lower (4.66 lower to 2.86 higher)	⊕⊕OO LOW	
Non remission - SAD												
2	randomised trials	no serious limitations	serious[4]	no serious indirectness	serious[1]	none	27/50 (54%)	26/48 (54.2%) 42.5%	RR 1.00 (0.69 to 1.45)	0 fewer per 100 (from 17 fewer to 24 more) 0 fewer per 100 (from 13 fewer to 19 more)	⊕⊕OO LOW	
Non response (overall)												
3	randomised trials	no serious limitations	serious[1]	no serious indirectness	serious[1]	none	29/66 (43.9%)	27/63 (42.9%) 40%	RR 1 (0.51 to 1.98)	0 fewer per 100 (from 21 fewer to 42 more) 0 fewer per 100 (from 20 fewer to 39 more)	⊕⊕OO LOW	
Non response - SAD												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[1]	none	24/50 (48%)	18/48 (37.5%) 32.5%	RR 1.26 (0.78 to 2.01)	10 more per 100 (from 8 fewer to 38 more) 8 more per 100 (from 7 fewer to 33 more)	⊕⊕⊕O MODERATE	
Non response - Subsyndromal SAD												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[3]	none	5/16 (31.3%)	9/15 (60%) 60%	RR 0.52 (0.23 to 1.2)	29 fewer per 100 (from 46 fewer to 12 more) 29 fewer per 100 (from 46 fewer to 12 more)	⊕⊕⊕O MODERATE	

[1] Inconclusive effect size
[2] Single study
[3] Inconclusive effect size; single study
[4] Significant heterogeneity; random effects model used

Is dawn simulation effective for depression with a seasonal pattern/SAD?

Quality assessment							Summary of findings					Importance
							No of patients		Effect		Quality	
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Dawn simulation	Attentional control	Relative (95% CI)	Absolute		
Leaving study early for any reason												
3	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[1]	none	2/70 (2.9%)	10/71 (14.1%) 19.4%	RR 0.33 (0.05 to 2.22)	9 fewer per 100 (from 13 fewer to 17 more) 13 fewer per 100 (from 18 fewer to 24 more)	⊕⊕OO LOW	
Leaving study early due to side effects												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	0/31 (0%)	1/31 (3.2%) 3.2%	RR 0.33 (0.01 to 7.88)	2 fewer per 100 (from 3 fewer to 22 more) 2 fewer per 100 (from 3 fewer to 22 more)	⊕⊕OO LOW	
Leaving study early due to lack of efficacy												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[1]	none	0/45 (0%)	6/44 (13.6%) 11.9%	RR 0.14 (0.02 to 1.1)	12 fewer per 100 (from 13 fewer to 1 more) 10 fewer per 100 (from 12 fewer to 1 more)	⊕⊕⊕O MODERATE	
Reported side effects												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	6/14 (42.9%)	1/13 (7.7%) 7.7%	RR 5.57 (0.77 to 40.26)	35 more per 100 (from 2 fewer to 302 more) 35 more per 100 (from 2 fewer to 302 more)	⊕⊕OO LOW	

Mean clinician rated typical depression scores at endpoint (measured with: HAMD-17/HRSD-21; Better indicated by lower values)												
2	randomised trials	no serious limitations	serious[3]	no serious indirectness	no serious imprecision	none	37	36	-	SMD 0.53 lower (1.62 lower to 0.15 higher)	⊕⊕⊕O MODERATE	
Mean clinician rated atypical depression scores at endpoint (measured with: SAD subscale; Better indicated by lower values)												
2	randomised trials	no serious limitations	serious[3]	no serious indirectness	very serious[2]	none	37	36	-	MD 2.20 lower (7.52 lower to 3.11 higher)	⊕OOO VERY LOW	
Non remission (SIGH-SAD)												
2	randomised trials	no serious limitations	serious[3]	no serious indirectness	serious[1]	none	25/56 (44.6%)	29/58 (50%) 49.9%	RR 0.9 (0.46 to 1.78)	5 fewer per 100 (from 27 fewer to 39 more) 5 fewer per 100 (from 27 fewer to 39 more)	⊕⊕OO LOW	
Non response (SIGH-SAD)												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[1]	none	14/56 (25%)	21/58 (36.2%) 36.3%	RR 0.71 (34 to 1.48)	11 fewer per 100 (from 17 more to 1195 more) 11 fewer per 100 (from 17 more to 1198 more)	⊕⊕⊕O MODERATE	

[1] Inconclusive effect size
[2] Inconclusive effect size; single study
[3] Significant heterogeneity; random effects model used

Is dawn simulation more effective than bright light box therapy for depression with a seasonal pattern/SAD?

Quality assessment							Summary of findings					
							No of patients		Effect		Quality	Importance
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Bright light box	Dawn simulation	Relative (95% CI)	Absolute		
Leaving study early for any reason												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[1]	none	5/56 (8.9%)	1/56 (1.8%)	RR 3.72 (0.62 to 22.22)	5 more per 100 (from 1 fewer to 38 more)		

								2%		5 more per 100 (from 1 fewer to 42 more)	⊕⊕⊕O MODERATE	
Leaving study early due to side effects												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	2/33 (6.1%)	0%	RR 4.71 (0.23 to 94.31)	0 more per 1000 (from 0 fewer to 0 more)		
Leaving study early due to lack of efficacy												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	0/31 (0%)	0/31 (0%) 0%	not pooled	not pooled not pooled	⊕⊕⊕⊕ HIGH	
Non remission (SIGH-SAD)												
2	randomised trials	no serious limitations	serious[3]	no serious indirectness	very serious[1]	none	30/56 (53.6%)	25/56 (44.6%) 46.1%	RR 1.19 (0.7 to 2)	8 more per 100 (from 13 fewer to 45 more) 9 more per 100 (from 14 fewer to 46 more)	⊕OOO VERY LOW	
Non response (SIGH-SAD)												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[1]	none	20/56 (35.7%)	14/56 (25%) 26.1%	RR 1.45 (0.82 to 2.58)	11 more per 100 (from 5 fewer to 39 more) 12 more per 100 (from 5 fewer to 41 more)	⊕⊕⊕O MODERATE	
Depression: mean endpoint scores (Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	21	24	-	MD 0.9 lower (4 lower to 2.2 higher)	⊕⊕OO LOW	
SAD: mean endpoint scores (Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	21	24	-	MD 1.8 lower (6.98 lower to 3.38 higher)	⊕⊕OO LOW	

[1] Inconclusive effect size
[2] Inconclusive effect size; single study
[3] Significant effect size - random effects model used

Non-light therapies for depression with a seasonal pattern/SAD

Are antidepressants effective in depression with a seasonal pattern/SAD? (Acute phase efficacy data)

Quality assessment							Summary of findings					Importance
							No of patients		Effect		Quality	
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Acute phase treatment :antidepressants	Control	Relative (95% CI)	Absolute		
Number not achieving =/> 50% reduction in SIGH-SAD score at endpoint (overall)												
2	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	57/129 (44.2%)	68/126 (54%) 57.8%	RR 0.82 (0.63 to 1.05)	10 fewer per 100 (from 20 fewer to 3 more) 10 fewer per 100 (from 21 fewer to 3 more)	⊕⊕⊕⊕ HIGH	
Number not achieving =/> 50% reduction SIGH-SAD score												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[1]	none	41/93 (44.1%)	47/94 (50%) 50%	RR 0.88 (0.65 to 1.2)	6 fewer per 100 (from 18 fewer to 10 more) 6 fewer per 100 (from 18 fewer to 10 more)	⊕⊕OO LOW	
Number not achieving =/> 50% reduction in outcome score at endpoint - Fluoxetine vs Placebo												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[1]	none	16/36 (44.4%)	21/32 (65.6%) 65.6%	RR 0.68 (0.43 to 1.05)	21 fewer per 100 (from 37 fewer to 3 more) 21 fewer per 100 (from 37 fewer to 3 more)	⊕⊕OO LOW	
Mean endpoint SIGH-SAD (clinician rated) (antidepressants) (Better indicated by lower values)												
2	randomised trials	no serious limitations	serious[2]	no serious indirectness	serious	none	52	47	-	SMD 0.11 lower (0.65 lower to 0.42 higher)	⊕⊕OO LOW	
Mean endpoint (clinician rated) (antidepressants) - Moclobemide vs Placebo (Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[1]	none	16	15	-	SMD 0.23 higher (0.48 lower to 0.94 higher)	⊕⊕OO LOW	

Mean endpoint (clinician rated) (antidepressants) - Fluoxetine vs Placebo (Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[1]	none	36	32	-	SMD 0.33 lower (0.81 lower to 0.15 higher)	⊕⊕OO LOW	
Mean endpoint BDI (self rated) - Fluoxetine vs Placebo (Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[1]	none	36	32	-	MD 1.7 lower (6.53 lower to 3.13 higher)	⊕⊕OO LOW	
Mean change (clinician rated) - Sertraline vs Placebo (Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[3]	none	93	93	-	MD 4.51 lower (8.23 to 0.79 lower)	⊕⊕⊕O MODERATE	
Relapse Prevention - Number of patients expriencing a recurrence												
3	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	92/542 (17%)	153/519 (29.5%) 31.9%	RR 0.58 (0.46 to 0.72)	12 fewer per 100 (from 8 fewer to 16 fewer) 13 fewer per 100 (from 9 fewer to 17 fewer)	⊕⊕⊕⊕ HIGH	

[1] Single study; inconclusive effect size
[2] Significant heterogeneity - random effects model used
[3] Single study

Are antidepressants effective in depression with a seasonal pattern/SAD? (Acute phase acceptability/tolerability data)

Quality assessment							Summary of findings					Importance
							No of patients		Effect		Quality	
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Acute phase acceptability and tolerability (antidepressants)	Placebo	Relative (95% CI)	Absolute		
Number leaving the study early for any reason (overall)												
2	randomised trials	no serious limitations	serious[1]	no serious indirectness	very serious[2]	none	20/109 (18.3%)	23/112 (20.5%) 19%	RR 0.7 (0.16 to 3.05)	6 fewer per 100 (from 17 fewer to 42 more) 6 fewer per 100 (from 16 fewer to 39 more)	⊕OOO VERY LOW	

Number leaving the study early for any reason - Sertraline vs Placebo												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[3]	none	20/93 (21.5%)	20/94 (21.3%) 21.3%	RR 1.01 (0.58 to 1.75)	0 more per 100 (from 9 fewer to 16 more) 0 more per 100 (from 9 fewer to 16 more)	⊕⊕OO LOW	
Number leaving the study early for any reason - Moclobemide vs Placebo												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[3]	none	0/16 (0%)	3/18 (16.7%) 16.7%	RR 0.16 (0.01 to 2.87)	14 fewer per 100 (from 17 fewer to 31 more) 14 fewer per 100 (from 17 fewer to 31 more)	⊕⊕OO LOW	
Number leaving the study early due to side effects												
3	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	12/145 (8.3%)	8/144 (5.6%) 5.3%	RR 1.48 (0.63 to 3.47)	3 more per 100 (from 2 fewer to 14 more) 3 more per 100 (from 2 fewer to 13 more)	⊕⊕OO LOW	
Number leaving the study early due to side effects - Sertraline vs Placebo												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	10/93 (10.8%)	5/94 (5.3%) 5.3%	RR 2.02 (0.72 to 5.69)	5 more per 100 (from 1 fewer to 25 more) 5 more per 100 (from 1 fewer to 25 more)	⊕⊕OO LOW	
Number leaving the study early due to side effects - Moclobemide vs Placebo												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[3]	none	0/16 (0%)	2/18 (11.1%) 11.1%	RR 0.22 (0.01 to 4.34)	9 fewer per 100 (from 11 fewer to 37 more) 9 fewer per 100 (from 11 fewer to 37 more)	⊕⊕OO LOW	
Number leaving the study early due to side effects - Fluoxetine vs Placebo												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[3]	none	2/36 (5.6%)	1/32 (3.1%)	RR 1.78 (0.17 to 18.69)	2 more per 100 (from 3 fewer to 55 more)		

								3.1%		2 more per 100 (from 3 fewer to 55 more)	⊕⊕OO LOW	
Number reporting side effects - Sertraline vs Placebo												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[4]	none	76/93 (81.7%)	47/94 (50%)	RR 1.63 (1.31 to 2.04)	31 more per 100 (from 15 more to 52 more)	⊕⊕⊕O MODERATE	
								50%		31 more per 100 (from 15 more to 52 more)		
Number reporting side effects - Fluoxetine vs Placebo												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[4]	none	35/36 (97.2%)	29/32 (90.6%)	RR 1.07 (0.95 to 1.21)	6 more per 100 (from 5 fewer to 19 more)	⊕⊕⊕O MODERATE	
								90.6%		6 more per 100 (from 5 fewer to 19 more)		

[1] Significant heterogeneity - random effects model used
[2] Inconclusive effect size
[3] Single study; inconclusive effect size
[4] Single study

Which antidepressant is more effective in depression with a seasonal pattern/SAD?

Quality assessment							Summary of findings					Importance
							No of patients		Effect		Quality	
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Acute phase treatment: antidepressants	Active control	Relative (95% CI)	Absolute		
Number not achieving =/> 50% reduction in SIGH-SAD score at endpoint - High ion density v Low ion density												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[1]	none	5/12 (41.7%)	11/13 (84.6%)	RR 0.49 (0.24 to 1)	43 fewer per 100 (from 64 fewer to 0 more)	⊕⊕⊕O MODERATE	
								84.6%		43 fewer per 100 (from 64 fewer to 0 more)		
Mean endpoint SIGH-SAD (clinician rated) - Moclobemide vs Fluoxetine (Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[1]	none	11	18	-	MD 1.6 lower (7.01 lower to 3.81 higher)	⊕⊕OO LOW	

[1] Single study; inconclusive effect size

Is continuation treatment effective for depression with a seasonal pattern/SAD?

Quality assessment							Summary of findings					Importance
							No of patients		Effect		Quality	
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Continuation treatment	Control	Relative (95% CI)	Absolute		
Mean endpoint HAMD-21 (clinician-rated) - Propanolol vs Placebo (Better indicated by lower values)												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[1]	none	12	11	-	MD 7 lower (11.24 to 2.76 lower)	⊕⊕⊕O MODERATE	
Number leaving the study early for any reason - Propanolol vs Placebo												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	none	1/13 (7.7%)	0/11 (0%) 0%	RR 2.57 (0.12 to 57.44)	0 more per 100 (from 0 fewer to 0 more) 0 more per 100 (from 0 fewer to 0 more)	⊕⊕OO LOW	

[1] Single study

[2] Single study; inconclusive effect size

Further-line treatment (chapter 8)

Increasing the dose of antidepressant versus continuing with the antidepressant at the same dose

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Increasing the dose of antidepressant	Continuing with the antidepressant at the same dose	Relative (95% CI)	Absolute		

Remission (follow-up 5-6 weeks; assessed with: Number of people scoring ≤7/8 on Hamilton Rating Scale for Depression (HAM-D))												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	50/165 (30.3%)	51/162 (31.5%)	RR 1.07 (0.63 to 1.83)	22 more per 1000 (from 116 fewer to 261 more) 23 more per 1000 (from 120 fewer to 269 more)	⊕OOO VERY LOW	
								32.4%				
Response (follow-up 5-6 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias[3]	64/127 (50.4%)	79/125 (63.2%)	RR 0.8 (0.65 to 0.99)	126 fewer per 1000 (from 6 fewer to 221 fewer) 107 fewer per 1000 (from 5 fewer to 188 fewer)	⊕OOO VERY LOW	
								53.7%				
Response (follow-up 5-6 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
2	randomised trials	serious[1]	very serious[5]	no serious indirectness	very serious[2]	reporting bias[3]	96/135 (71.1%)	105/135 (77.8%)	RR 1.03 (0.59 to 1.8)	23 more per 1000 (from 319 fewer to 622 more) 21 more per 1000 (from 292 fewer to 570 more)	⊕OOO VERY LOW	
								71.2%				
Depression symptomatology (follow-up mean 6 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	serious[7]	reporting bias[3]	30	27	-	MD 1.7 higher (1.09 lower to 4.49 higher)	⊕OOO VERY LOW	
Discontinuation for any reason (follow-up 5-6 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
3	randomised trials	serious[1]	serious[8]	no serious indirectness	very serious[9]	reporting bias[3]	20/166 (12%)	23/166 (13.9%)	RR 0.7 (0.21 to 2.38)	42 fewer per 1000 (from 109 fewer to 191 more) 41 fewer per 1000 (from 107	⊕OOO VERY LOW	
								13.5%				

										fewer to 186 more)		
Discontinuation due to adverse events (follow-up mean 6 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[9]	reporting bias[3]	0/30 (0%)	4/30 (13.3%)	RR 0.11 (0.01 to 1.98)	119 fewer per 1000 (from 132 fewer to 131 more)	⊕OOO VERY LOW	
							13.3%			118 fewer per 1000 (from 132 fewer to 130 more)		

[1] Unclear blinding of intervention administrator and outcome assessor
[2] 95% CI crosses line of no effect, and threshold for both clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)
[3] Data cannot be extracted/is not reported for all outcomes and study funded by pharmaceutical company
[4] Events<300
[5] I-squared>80%
[6] Unclear blinding of intervention administration and possible risk of attrition bias difference in drop-out between groups>20%) although ITT analysis used
[7] 95% CI crosses both line of no effect and threshold for clinically important harm (SMD 0.5)
[8] I-squared>50%
[9] 95% CI crosses line of no effect and both threshold for clinically important benefit (RR 0.75) and threshold for clinically important harm (RR 1.25)

1 Increasing the dose of antidepressant versus switching to another antidepressant

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Increasing the dose of antidepressant	Switching to another antidepressant	Relative (95% CI)	Absolute		
Remission (follow-up mean 8 weeks; assessed with: Number of people scoring ≤10 on Montgomery Asberg Depression Rating Scale (MADRS))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	124/229 (54.1%)	102/243 (42%)	RR 1.29 (1.07 to 1.56)	122 more per 1000 (from 29 more to 235 more)	⊕OOO VERY LOW	
								42%		122 more per 1000 (from 29 more to 235 more)		
Response (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Montgomery Asberg Depression Rating Scale (MADRS))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[3]	167/229 (72.9%)	170/243 (70%)	RR 1.04 (0.93 to 1.17)	28 more per 1000 (from 49 fewer to 119 more)	⊕⊕OO LOW	
								70%		28 more per 1000 (from 49 fewer to 119 more)		
Response (follow-up mean 8 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[3]	176/229 (76.9%)	182/243 (74.9%)	RR 1.03 (0.93 to 1.14)	22 more per 1000 (from 52 fewer to 105 more)	⊕⊕OO LOW	
								74.9%		22 more per 1000 (from 52 fewer to 105 more)		
Depression symptomatology (follow-up mean 8 weeks; measured with: Quick Inventory of Depressive Symptomatology (QIDS; change score); Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[3]	229	243	-	MD 0.9 lower (1.88 lower to 0.08 higher)	⊕⊕OO LOW	

Discontinuation for any reason (follow-up mean 8 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	reporting bias[4]	56/238 (23.5%)	53/246 (21.5%)	RR 1.09 (0.78 to 1.52)	19 more per 1000 (from 47 fewer to 112 more) 19 more per 1000 (from 47 fewer to 112 more)	⊕⊕OO LOW	
								21.5%				
Discontinuation due to adverse events (follow-up mean 8 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	13/238 (5.5%)	13/246 (5.3%)	RR 1.03 (0.49 to 2.18)	2 more per 1000 (from 27 fewer to 62 more) 2 more per 1000 (from 27 fewer to 63 more)	⊕OOO VERY LOW	
								5.3%				

[1] Unclear blinding of outcome assessment and risk of attrition bias (drop-out>20% [23%]) although difference between groups<20% and ITT analysis
[2] Events<300
[3] Data cannot be extracted/is not reported for all outcomes and study funded by pharmaceutical company
[4] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 1.25)
[5] 95% CI crosses line of no effect and both threshold for clinically important benefit (RR 0.75) and threshold for clinically important harm (RR 1.25)

Increasing the dose of antidepressant versus augmenting with another antidepressant/non-antidepressant agent

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Increasing the dose of antidepressant**	**Augmenting with another antidepressant/non-antidepressant agent**	**Relative (95% CI)**	**Absolute**		
Remission (increasing dose of SSRI versus TCA augmentation) (follow-up mean 4 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[1]	reporting bias[2]	22/48 (45.8%)	13/46 (28.3%)	RR 1.6 (0.91 to 2.81)	170 more per 1000 (from 25 fewer to 512 more) 163 more per 1000 (from 24	⊕OOO VERY LOW	
								27.2%				

										fewer to 492 more)		
Remission (increasing dose of SSRI versus lithium augmentation) (follow-up mean 4 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias[2]	22/48 (45.8%)	12/48 (25%)	RR 1.83 (1.03 to 3.25)	208 more per 1000 (from 7 more to 562 more)	⊕OOO VERY LOW	
							26.1%			217 more per 1000 (from 8 more to 587 more)		
Remission (increasing dose of SSRI versus TeCA [mianserin] augmentation) (follow-up mean 4 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias[2]	28/97 (28.9%)	43/98 (43.9%)	RR 0.66 (0.45 to 0.97)	149 fewer per 1000 (from 13 fewer to 241 fewer)	⊕OOO VERY LOW	
							43.9%			149 fewer per 1000 (from 13 fewer to 241 fewer)		
Response (increasing dose of SSRI versus TeCA [mianserin] augmentation) (follow-up mean 5 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[6]	reporting bias[2]	54/97 (55.7%)	66/98 (67.3%)	RR 0.83 (0.66 to 1.03)	114 fewer per 1000 (from 229 fewer to 20 more)	⊕OOO VERY LOW	
							67.4%			115 fewer per 1000 (from 229 fewer to 20 more)		
Response (increasing dose of SSRI versus TeCA [mianserin] augmentation) (follow-up mean 5 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[6]	reporting bias[2]	66/97 (68%)	76/98 (77.6%)		93 fewer per 1000 (from 202		

									RR 0.88 (0.74 to 1.04)	fewer to 31 more) 93 fewer per 1000 (from 202 fewer to 31 more)	⊕OOO VERY LOW	
								77.6%				
Depression symptomatology (iIncreasing dose of SSRI versus TCA augmentation) (follow-up mean 4 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
2	randomised trials	serious[3]	serious[7]	no serious indirectness	serious[8]	reporting bias[2]	48	46	-	MD 2.97 lower (6.08 lower to 0.13 higher)	⊕OOO VERY LOW	
Depression symptomatology (iIncreasing dose of SSRI versus lithium augmentation) (follow-up mean 4 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
2	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[8]	reporting bias[2]	48	48	-	MD 2 lower (4.32 lower to 0.33 higher)	⊕OOO VERY LOW	
Discontinuation for any reason (increasing dose of SSRI versus TCA augmentation) (follow-up mean 4 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[9]	reporting bias[2]	5/48 (10.4%)	8/46 (17.4%)	RR 0.58 (0.21 to 1.64)	73 fewer per 1000 (from 137 fewer to 111 more) 84 fewer per 1000 (from 157 fewer to 127 more)	⊕OOO VERY LOW	
								19.9%				
Discontinuation for any reason (increasing dose of SSRI versus lithium augmentation) (follow-up mean 4 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[9]	reporting bias[2]	5/48 (10.4%)	7/48 (14.6%)	RR 0.72 (0.24 to 2.11)	41 fewer per 1000 (from 111 fewer to 162 more) 41 fewer per 1000 (from 110 fewer to 161 more)	⊕OOO VERY LOW	
								14.5%				

Discontinuation for any reason (increasing dose of SSRI versus TeCA [mianserin] augmentation) (follow-up mean 5 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))											
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[9]	reporting bias[2]	15/98 (15.3%)	17/98 (17.3%)	RR 0.88 (0.47 to 1.67)	21 fewer per 1000 (from 92 fewer to 116 more)	⊕OOO VERY LOW	
								17.4%		21 fewer per 1000 (from 92 fewer to 117 more)		
Discontinuation due to adverse events (increasing dose of SSRI versus TCA augmentation) (follow-up mean 4 weeks; assessed with: Number of participants discontinuing due to adverse events)											
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[9]	reporting bias[2]	0/15 (0%)	2/12 (16.7%)	RR 0.16 (0.01 to 3.09)	140 fewer per 1000 (from 165 fewer to 348 more)	⊕OOO VERY LOW	
								16.7%		140 fewer per 1000 (from 165 fewer to 349 more)		
Discontinuation due to adverse events (increasing dose of SSRI versus lithium augmentation) (follow-up mean 4 weeks; assessed with: Number of participants discontinuing due to adverse events)											
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[9]	reporting bias[2]	0/15 (0%)	1/14 (7.1%)	RR 0.31 (0.01 to 7.09)	49 fewer per 1000 (from 71 fewer to 435 more)	⊕OOO VERY LOW	
								7.1%		49 fewer per 1000 (from 70 fewer to 432 more)		

[1] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[2] Data cannot be extracted/is not reported for all outcomes and/or funding from pharmaceutical company
[3] Unclear randomization method and allocation concealment and unclear blinding of intervention administration and outcome assessment
[4] Events<300
[5] Unclear blinding of intervention administration and outcome assessment
[6] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 0.75)
[7] I-squared>50%
[8] 95% CI crosses both line of effect and threshold for clinically important benefit (SMD -0.5)
[9] 95% CI crosses line of no effect and both threshold for clinically important benefit (RR 0.75) and threshold for clinically important harm (RR 1.25)

1 Augmenting the antidepressant with another antidepressant or a non-antidepressant agent versus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Augmenting the antidepressant with another antidepressant or a non-antidepressant agent	Placebo	Relative (95% CI)	Absolute		
Remission (atypical antidepressant) (follow-up mean 4 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	23/41 (56.1%)	9/45 (20%)	RR 2.72 (1.44 to 5.14)	344 more per 1000 (from 88 more to 828 more)	⊕OOO VERY LOW	
								18.3%		315 more per 1000 (from 81 more to 758 more)		
Remission (antipsychotic) (follow-up 4-8 weeks; assessed with: Number of people scoring <10/11 on Montgomery Asberg Depression Rating Scale (MADRS)/≤7 on Hamilton Rating Scale for Depression (HAM-D))												
9	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[3]	500/1408 (35.5%)	226/1173 (19.3%)	RR 1.56 (1.36 to 1.78)	108 more per 1000 (from 69 more to 150 more)	⊕⊕OO LOW	
								17.2%		96 more per 1000 (from 62 more to 134 more)		
Remission (lithium) (follow-up 2-6 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	24/54 (44.4%)	12/56 (21.4%)	RR 2.07 (1.16 to 3.69)	229 more per 1000 (from 34 more to 576 more)	⊕OOO VERY LOW	
								25%		267 more per 1000 (from 40 more to 673 more)		

Remission (thyroid hormone [T3]) (follow-up mean 2 weeks; assessed with: Number of people scoring <7 on Hamilton Rating Scale for Depression (HAM-D) AND responding (≥50% improvement on HAM-D))													
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	7/17 (41.2%)	2/16 (12.5%)	RR 3.29 (0.8 to 13.57)	286 more per 1000 (from 25 fewer to 1000 more)	⊕⊕OO LOW	
								12.5%		286 more per 1000 (from 25 fewer to 1000 more)		
Remission (stimulant [methylphenidate]) (follow-up mean 4 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))													
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias[3]	4/30 (13.3%)	1/30 (3.3%)	RR 4 (0.47 to 33.73)	100 more per 1000 (from 18 fewer to 1000 more)	⊕OOO VERY LOW	
								3.3%		99 more per 1000 (from 17 fewer to 1000 more)		
Response (any augmentation agent) (follow-up 0.3-12 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D)/Montgomery Asberg Depression Rating Scale (MADRS))													
20	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[3]	759/1689 (44.9%)	416/1421 (29.3%)	RR 1.35 (1.23 to 1.49)	102 more per 1000 (from 67 more to 143 more)	⊕⊕OO LOW	
								25.3%		89 more per 1000 (from 58 more to 124 more)		
Response (atypical antidepressant) (follow-up mean 4 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))													
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	7/11 (63.6%)	3/15 (20%)	RR 3.18 (1.05 to 9.62)	436 more per 1000 (from 10 more to 1000 more)	⊕OOO VERY LOW	
								20%		436 more per 1000 (from 10 more to 1000 more)		

Response (antipsychotic) (follow-up 4-8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D)/Montgomery Asberg Depression Rating Scale (MADRS))												
10	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[3]	660/1420 (46.5%)	344/1184 (29.1%)	RR 1.4 (1.25 to 1.57)	116 more per 1000 (from 73 more to 166 more) 114 more per 1000 (from 71 more to 162 more)	⊕⊕OO LOW	
								28.5%				
Response (lithium) (follow-up 0.3-6 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
4	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias[3]	9/38 (23.7%)	6/38 (15.8%)	RR 1.55 (0.61 to 3.91)	87 more per 1000 (from 62 fewer to 459 more) 83 more per 1000 (from 59 fewer to 439 more)	⊕OOO VERY LOW	
								15.1%				
Response (anticonvulsant [lamotrigine]) (follow-up 8-10 weeks; assessed with: Number of people showing ≥50% improvement on Montgomery Asberg Depression Rating Scale (MADRS))												
2	randomised trials	very serious[7]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias[3]	21/65 (32.3%)	22/65 (33.8%)	RR 0.96 (0.59 to 1.56)	14 fewer per 1000 (from 139 fewer to 190 more) 14 fewer per 1000 (from 141 fewer to 192 more)	⊕OOO VERY LOW	
								34.3%				
Response (omega-3 fatty acid) (follow-up mean 12 weeks; assessed with: Number of people showing ≥50% improvement on Montgomery Asberg Depression Rating Scale (MADRS))												
1	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias[3]	16/52 (30.8%)	4/17 (23.5%)	RR 1.31 (0.51 to 3.38)	73 more per 1000 (from 115 fewer to 560 more) 73 more per 1000 (from 115 fewer to 559 more)	⊕OOO VERY LOW	
								23.5%				
Response (stimulant [methylphenidate]) (follow-up 4-5 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D)/Montgomery Asberg Depression Rating Scale (MADRS))												

2	randomised trials	very serious[7]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	46/103 (44.7%)	37/102 (36.3%)	RR 1.21 (0.87 to 1.68)	76 more per 1000 (from 47 fewer to 247 more)	⊕OOO VERY LOW	
								32.5%		68 more per 1000 (from 42 fewer to 221 more)		
Response (any augmentation agent) (follow-up 4-8 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
5	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	46/127 (36.2%)	37/130 (28.5%)	RR 1.29 (0.85 to 1.97)	83 more per 1000 (from 43 fewer to 276 more)	⊕OOO VERY LOW	
								26.7%		77 more per 1000 (from 40 fewer to 259 more)		
Response (atypical antidepressant) (follow-up mean 4 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	7/11 (63.6%)	3/15 (20%)	RR 3.18 (1.05 to 9.62)	436 more per 1000 (from 10 more to 1000 more)	⊕OOO VERY LOW	
								20%		436 more per 1000 (from 10 more to 1000 more)		
Response (lithium) (follow-up mean 6 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias[3]	5/18 (27.8%)	4/17 (23.5%)	RR 1.18 (0.38 to 3.67)	42 more per 1000 (from 146 fewer to 628 more)	⊕OOO VERY LOW	
								23.5%		42 more per 1000 (from 146 fewer to 627 more)		
Response (anticonvulsant [lamotrigine]) (follow-up mean 8 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[9]	no serious inconsistency	no serious indirectness	very serious[6]	none	4/17 (23.5%)	6/17 (35.3%)	RR 0.67 (0.23 to 1.95)	116 fewer per 1000 (from 272 fewer to 335 more)	⊕OOO VERY LOW	

								35.3%		116 fewer per 1000 (from 272 fewer to 335 more)		
Response (anxiolytic) (follow-up mean 6 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[10]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias[3]	17/51 (33.3%)	16/51 (31.4%)	RR 1.06 (0.61 to 1.86)	19 more per 1000 (from 122 fewer to 270 more)	⊕OOO VERY LOW	
								31.4%		19 more per 1000 (from 122 fewer to 270 more)		
Response (stimulant [methylphenidate]) (follow-up mean 4 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	13/30 (43.3%)	8/30 (26.7%)	RR 1.62 (0.79 to 3.34)	165 more per 1000 (from 56 fewer to 624 more)	⊕OOO VERY LOW	
								26.7%		166 more per 1000 (from 56 fewer to 625 more)		
Depression symptomatology (atypical antidepressant) (follow-up mean 4 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[11]	reporting bias[3]	11	15	-	SMD 1.12 lower (1.96 to 0.27 lower)	⊕OOO VERY LOW	
Depression symptomatology (antipsychotic) (follow-up 4-8 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
3	randomised trials	serious[4]	serious[12]	no serious indirectness	serious[13]	reporting bias[3]	172	290	-	SMD 0.4 lower (0.86 lower to 0.06 higher)	⊕OOO VERY LOW	
Depression symptomatology (lithium) (follow-up 2-3 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												

3	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[13]	reporting bias[3]	41	42	-	SMD 0.23 lower (0.86 lower to 0.39 higher)	⊕OOO VERY LOW	
Depression symptomatology (thyroid hormone [T3]) (follow-up mean 2 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[14]	no serious inconsistency	no serious indirectness	serious[11]	none	17	16	-	SMD 0.78 lower (1.5 to 0.07 lower)	⊕⊕OO LOW	
Depression symptomatology (anticonvulsant [lamotrigine]) (follow-up 8-10 weeks; measured with: Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
2	randomised trials	very serious[15]	no serious inconsistency	no serious indirectness	serious[13]	reporting bias[3]	65	65	-	SMD 0.13 lower (0.54 lower to 0.27 higher)	⊕OOO VERY LOW	
Depression symptomatology (stimulant [methylphenidate]) (follow-up mean 5 weeks; measured with: Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
1	randomised trials	very serious[16]	no serious inconsistency	no serious indirectness	serious[11]	reporting bias[3]	72	72	-	SMD 0.06 higher (0.27 lower to 0.38 higher)	⊕OOO VERY LOW	
Discontinuation for any reason (atypical antidepressant) (follow-up mean 4 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[17]	reporting bias[3]	1/41 (2.4%) 6.7%	2/45 (4.4%)	RR 0.68 (0.07 to 6.61)	14 fewer per 1000 (from 41 fewer to 249 more) 21 fewer per 1000 (from 62 fewer to 376 more)	⊕OOO VERY LOW	
Discontinuation for any reason (antipsychotic) (follow-up 4-8 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
10	randomised trials	serious[18]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[3]	265/1480 (17.9%)	165/1226 (13.5%)	RR 1.24 (1.02 to 1.52)	32 more per 1000 (from 3 more to 70 more)	⊕⊕OO LOW	

								14.1%		34 more per 1000 (from 3 more to 73 more)		
Discontinuation for any reason (lithium) (follow-up 2-6 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
6	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[17]	reporting bias[3]	10/99 (10.1%)	12/101 (11.9%)	RR 0.87 (0.41 to 1.84)	15 fewer per 1000 (from 70 fewer to 100 more)	⊕OOO VERY LOW	
								5.6%		7 fewer per 1000 (from 33 fewer to 47 more)		
Discontinuation for any reason (thyroid hormone [T3]) (follow-up mean 2 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[14]	no serious inconsistency	no serious indirectness	serious[2]	none	0/27 (0%)	0/24 (0%)	not pooled	not pooled	⊕⊕OO LOW	
								0%		not pooled		
Discontinuation for any reason (anticonvulsant [lamotrigine]) (follow-up 8-10 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	very serious[19]	no serious inconsistency	no serious indirectness	very serious[17]	reporting bias[3]	17/65 (26.2%)	21/65 (32.3%)	RR 0.81 (0.48 to 1.38)	61 fewer per 1000 (from 168 fewer to 123 more)	⊕OOO VERY LOW	
								29.5%		56 fewer per 1000 (from 153 fewer to 112 more)		
Discontinuation for any reason (anxiolytic) (follow-up mean 6 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
1	randomised trials	serious[20]	no serious inconsistency	no serious indirectness	very serious[17]	reporting bias[3]	6/51 (11.8%)	10/51 (19.6%)	RR 0.6 (0.24 to 1.53)	78 fewer per 1000 (from 149 fewer to 104 more)	⊕OOO VERY LOW	
								19.6%		78 fewer per 1000 (from 149 fewer to 104 more)		
Discontinuation for any reason (omega-3 fatty acid) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[17]	reporting bias[3]	6/52 (11.5%)	4/18 (22.2%)	RR 0.52 (0.17 to 1.63)	107 fewer per 1000 (from 184 fewer to 140 more)	⊕OOO VERY LOW	
										107 fewer per 1000 (from 184 fewer to 140 more)		
								22.2%				
Discontinuation for any reason (stimulant [methylphenidate]) (follow-up mean 5 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
1	randomised trials	very serious[16]	no serious inconsistency	no serious indirectness	serious[21]	reporting bias[3]	11/73 (15.1%)	4/72 (5.6%)	RR 2.71 (0.91 to 8.12)	95 more per 1000 (from 5 fewer to 396 more)	⊕OOO VERY LOW	
										96 more per 1000 (from 5 fewer to 399 more)		
								5.6%				
Discontinuation due to adverse events (atypical antidepressant) (follow-up mean 4 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[22]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	0/30 (0%)	0/30 (0%)	not pooled	not pooled	⊕OOO VERY LOW	
								0%		not pooled		
Discontinuation due to adverse events (antipsychotic) (follow-up 4-8 weeks; assessed with: Number of participants discontinuing due to adverse events)												
10	randomised trials	serious[18]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	120/1480 (8.1%)	21/1226 (1.7%)	RR 3.16 (1.97 to 5.06)	37 more per 1000 (from 17 more to 70 more)	⊕OOO VERY LOW	
										43 more per 1000 (from 19 more to 81 more)		
								2%				
Discontinuation due to adverse events (lithium) (follow-up 2-6 weeks; assessed with: Number of participants discontinuing due to adverse events)												
5	randomised trials	serious[14]	no serious inconsistency	no serious indirectness	very serious[17]	reporting bias[3]	4/81 (4.9%)	3/84 (3.6%)	RR 1.3 (0.33 to 5.14)	11 more per 1000 (from 24 fewer to 148 more)	⊕OOO VERY LOW	
								0%		-		
Discontinuation due to adverse events (thyroid hormone [T3]) (follow-up mean 2 weeks; assessed with: Number of participants discontinuing due to adverse events)												

2	randomised trials	serious[14]	no serious inconsistency	no serious indirectness	serious[2]	none	0/27 (0%)	0/24 (0%)	not pooled	not pooled	⊕⊕OO LOW	
								0%		not pooled		
Discontinuation due to adverse events (anticonvulsant [lamotrigine]) (follow-up 8-10 weeks; assessed with: Number of participants discontinuing due to adverse events)												
2	randomised trials	very serious[19]	no serious inconsistency	no serious indirectness	very serious[17]	reporting bias[3]	9/65 (13.8%)	10/65 (15.4%)	RR 1.12 (0.21 to 5.94)	18 more per 1000 (from 122 fewer to 760 more)	⊕OOO VERY LOW	
								10.4%		12 more per 1000 (from 82 fewer to 514 more)		
Discontinuation due to adverse events (anxiolytic) (follow-up mean 6 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[20]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	0/51 (0%)	0/51 (0%)	not pooled	not pooled	⊕OOO VERY LOW	
								0%		not pooled		
Discontinuation due to adverse events (omega-3 fatty acid) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[17]	reporting bias[3]	1/52 (1.9%)	1/18 (5.6%)	RR 0.35 (0.02 to 5.25)	36 fewer per 1000 (from 54 fewer to 236 more)	⊕OOO VERY LOW	
								5.6%		36 fewer per 1000 (from 55 fewer to 238 more)		
Discontinuation due to adverse events (stimulant [methylphenidate]) (follow-up 4-5 weeks; assessed with: Number of participants discontinuing due to adverse events)												
2	randomised trials	serious[18]	serious[12]	no serious indirectness	very serious[17]	reporting bias[3]	8/103 (7.8%)	2/102 (2%)	RR 2.92 (0.21 to 40.65)	38 more per 1000 (from 15 fewer to 777 more)	⊕OOO VERY LOW	
								3.3%		63 more per 1000 (from 26 fewer to 1000 more)		

[1] Unclear randomisation method and method for allocation concealment. Blinding of intervention administration and outcome assessment is also unclear

[2] Events<300

[3] Data cannot be extracted/is not reported for all outcomes and/or funding from pharmaceutical company

[4] Unclear randomisation method and method for allocation concealment. Blinding of intervention administration and outcome assessment is also unclear for studies that make up >50% weighting in the analysis

[5] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[6] 95% CI crosses line of no effect and threshold for both clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)
[7] Significant group differences in baseline demographics at baseline in studies contributing to>50% weighting in analysis and unclear blinding of intervention administration and outcome assessment
[8] Unclear blinding of outcome assessment
[9] Unclear blinding of outcome assessment and unclear risk of attrition bias (drop-out>20% [21%] but difference between groups<20% and ITT analysis)
[10] Unclear randomisation method and method of allocation concealment. Blinding of outcome assessment is also unclear
[11] N<400
[12] I-squared>50%
[13] 95% CI crosses both line of no effect and threshold for clinically important benefit (SMD -0.5)
[14] Unclear randomisation method and method of allocation concealment, and blinding of intervention administration unclear
[15] High risk of bias associated with randomisation method as significant differences between groups at baseline in studies contributing >50% to weighting in analysis. Unclear blinding of outcome assessment and unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)
[16] High risk of bias associated with randomisation method as significant difference between groups at baseline and method of allocation concealment is unclear. Blinding of intervention administration is also unclear
[17] 95% CI crosses line of no effect and both threshold for clinically important benefit (RR 0.75) and threshold for clinically important harm (RR 1.25)
[18] Unclear or high risk of bias associated with randomisation method, unclear method of allocation concealment, and unclear blinding of intervention administration in studies contributing to >50% of weighting in analysis
[19] High risk of bias associated with randomisation method as significant difference between groups at baseline
[20] Unclear randomisation method and method of allocation concealment
[21] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 1.25)
[22] Unclear randomisation method and method of allocation concealment, and blinding of intervention administration is unclear

Augmenting the antidepressant with another antidepressant/non-antidepressant agent versus continuing with the antidepressant-only

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Augmenting the antidepressant with another antidepressant/non-antidepressant agent versus continuing with the antidepressant-only	Control	Relative (95% CI)	Absolute		
Remission (TeCA [mianserin] + SSRI versus SSRI-only) (follow-up 5-6 weeks; assessed with: Number of people scoring ≤7/8 on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	serious[1]	serious[2]	no serious indirectness	serious[3]	reporting bias[4]	57/130 (43.8%)	44/136 (32.4%)	RR 1.52 (0.77 to 3.01)	168 more per 1000 (from 74 fewer to 650 more)	⊕OOO VERY LOW	
								28.1%		146 more per 1000 (from 65 fewer to 565 more)		

Remission (antipsychotic + SSRI versus SSRI-only) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D)/≤10 on Montgomery Asberg Depression Rating Scale (MADRS))												
3	randomised trials	serious[5]	serious[2]	no serious indirectness	very serious[6]	reporting bias[4]	71/283 (25.1%)	56/268 (20.9%)	RR 1.12 (0.46 to 2.75)	25 more per 1000 (from 113 fewer to 366 more)	⊕OOO VERY LOW	
								16.8%		20 more per 1000 (from 91 fewer to 294 more)		
Remission (anticonvulsant + SSRI versus SSRI-only) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[7]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias[4]	19/39 (48.7%)	21/45 (46.7%)	RR 1.04 (0.67 to 1.63)	19 more per 1000 (from 154 fewer to 294 more)	⊕OOO VERY LOW	
								46.7%		19 more per 1000 (from 154 fewer to 294 more)		
Remission (anxiolytic + SSRI versus SSRI-only) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[7]	no serious inconsistency	no serious indirectness	serious[8]	reporting bias[4]	15/46 (32.6%)	21/45 (46.7%)	RR 0.7 (0.42 to 1.18)	140 fewer per 1000 (from 271 fewer to 84 more)	⊕OOO VERY LOW	
								46.7%		140 fewer per 1000 (from 271 fewer to 84 more)		
Remission (SARI + SSRI versus SSRI-only) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[7]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias[4]	20/47 (42.6%)	21/45 (46.7%)	RR 0.91 (0.58 to 1.44)	42 fewer per 1000 (from 196 fewer to 205 more)	⊕OOO VERY LOW	
								46.7%		42 fewer per 1000 (from 196 fewer to 205 more)		

Remission (thyroid hormone + SSRI versus SSRI-only) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[7]	no serious inconsistency	no serious indirectness	serious[3]	reporting bias[4]	18/48 (37.5%)	12/45 (26.7%)	RR 1.41 (0.77 to 2.58)	109 more per 1000 (from 61 fewer to 421 more)	⊕OOO VERY LOW	
								26.7%		109 more per 1000 (from 61 fewer to 422 more)		
Response (TeCA [mianserin] + SSRI versus SSRI-only) (follow-up 5-6 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	serious[1]	serious[2]	no serious indirectness	very serious[6]	reporting bias[4]	86/130 (66.2%)	83/136 (61%)	RR 1.22 (0.69 to 2.15)	134 more per 1000 (from 189 fewer to 702 more)	⊕OOO VERY LOW	
								53.6%		118 more per 1000 (from 166 fewer to 616 more)		
Response (lithium + SSRI versus SSRI-only) (follow-up mean 1 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[9]	no serious inconsistency	no serious indirectness	serious[10]	reporting bias[4]	6/10 (60%)	2/14 (14.3%)	RR 4.2 (1.06 to 16.68)	457 more per 1000 (from 9 more to 1000 more)	⊕OOO VERY LOW	
								14.3%		458 more per 1000 (from 9 more to 1000 more)		
Response (antipsychotic + SSRI versus SSRI-only) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D)/Montgomery Asberg Depression Rating Scale (MADRS))												
3	randomised trials	serious[5]	serious[2]	no serious indirectness	very serious[6]	reporting bias[4]	111/283 (39.2%)	92/268 (34.3%)	RR 1.12 (0.61 to 2.07)	41 more per 1000 (from 134 fewer to 367 more)	⊕OOO VERY LOW	
								29.6%		36 more per 1000 (from 115		

										fewer to 317 more)		
Response (anticonvulsant + SSRI versus SSRI-only) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[7]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias[4]	24/39 (61.5%)	30/45 (66.7%)	RR 0.92 (0.67 to 1.27)	53 fewer per 1000 (from 220 fewer to 180 more)	⊕OOO VERY LOW	
								66.7%		53 fewer per 1000 (from 220 fewer to 180 more)		
Response (anxiolytic + SSRI versus SSRI-only) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[7]	no serious inconsistency	no serious indirectness	serious[8]	reporting bias[4]	26/46 (56.5%)	30/45 (66.7%)	RR 0.85 (0.61 to 1.18)	100 fewer per 1000 (from 260 fewer to 120 more)	⊕OOO VERY LOW	
								66.7%		100 fewer per 1000 (from 260 fewer to 120 more)		
Response (SARI + SSRI versus SSRI-only) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[7]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias[4]	29/47 (61.7%)	30/45 (66.7%)	RR 0.93 (0.68 to 1.26)	47 fewer per 1000 (from 213 fewer to 173 more)	⊕OOO VERY LOW	
								66.7%		47 fewer per 1000 (from 213 fewer to 173 more)		
Response (thyroid hormone + SSRI versus SSRI-only) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[7]	no serious inconsistency	no serious indirectness	serious[3]	reporting bias[4]	28/48 (58.3%)	21/45 (46.7%)	RR 1.25 (0.84 to 1.85)	117 more per 1000 (from 75 fewer to 397 more)	⊕OOO VERY LOW	

								46.7%		117 more per 1000 (from 75 fewer to 397 more)		
Response (TeCA [mianserin] + SSRI versus SSRI-only) (follow-up 5-6 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
2	randomised trials	serious[1]	very serious[11]	no serious indirectness	very serious[6]	reporting bias[4]	99/130 (76.2%)	101/136 (74.3%)	RR 1.17 (0.65 to 2.12)	126 more per 1000 (from 260 fewer to 832 more)	⊕OOO VERY LOW	
								65.2%		111 more per 1000 (from 228 fewer to 730 more)		
Depression symptomatology (any augmentation agent) (follow-up 6-8 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
3	randomised trials	serious[12]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[4]	270	261	-	SMD 0.37 lower (0.55 to 0.2 lower)	⊕⊕OO LOW	
Depression symptomatology (TeCA [mianserin]) + SSRI versus SSRI-only) (follow-up mean 6 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	very serious[13]	no serious inconsistency	no serious indirectness	serious[14]	reporting bias[4]	32	38	-	SMD 0.66 lower (1.14 to 0.17 lower)	⊕OOO VERY LOW	
Depression symptomatology (antipsychotic + SSRI versus SSRI-only) (follow-up mean 8 weeks; measured with: Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
2	randomised trials	serious[15]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[4]	238	223	-	SMD 0.33 lower (0.52 to 0.15 lower)	⊕⊕OO LOW	
Discontinuation for any reason (any augmentation agent) (follow-up 5-8 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
4	randomised trials	serious[16]	no serious inconsistency	no serious indirectness	serious[10]	reporting bias[4]	96/371 (25.9%)	62/363 (17.1%)	RR 1.43 (1.07 to 1.91)	73 more per 1000 (from 12 more to 155 more)	⊕OOO VERY LOW	

								18.9%		81 more per 1000 (from 13 more to 172 more)		
Discontinuation for any reason (TeCA [mianserin] + SSRI versus SSRI-only) (follow-up 5-6 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[17]	no serious inconsistency	no serious indirectness	serious[18]	reporting bias[4]	23/130 (17.7%)	17/137 (12.4%)	RR 1.43 (0.79 to 2.56)	53 more per 1000 (from 26 fewer to 194 more)	⊕OOO VERY LOW	
								14.3%		61 more per 1000 (from 30 fewer to 223 more)		
Discontinuation for any reason (antipsychotic + SSRI versus SSRI-only) (follow-up mean 8 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[19]	no serious inconsistency	no serious indirectness	serious[10]	reporting bias[4]	73/241 (30.3%)	45/226 (19.9%)	RR 1.44 (1.03 to 2)	88 more per 1000 (from 6 more to 199 more)	⊕OOO VERY LOW	
								22.2%		98 more per 1000 (from 7 more to 222 more)		
Discontinuation due to adverse events (any augmentation agent) (follow-up 6-8 weeks; assessed with: Number of participants discontinuing due to adverse events)												
3	randomised trials	serious[12]	no serious inconsistency	no serious indirectness	serious[10]	reporting bias[4]	45/273 (16.5%)	5/264 (1.9%)	RR 6.19 (2.65 to 14.47)	98 more per 1000 (from 31 more to 255 more)	⊕OOO VERY LOW	
								0%		-		
Discontinuation due to adverse events (TeCA [mianserin] + SSRI versus SSRI-only) (follow-up mean 6 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[20]	no serious inconsistency	no serious indirectness	very serious[21]	reporting bias[4]	2/32 (6.3%)	0/38 (0%)	RR 5.91 (0.29 to 118.78)	-	⊕OOO VERY LOW	
								0%		-		

Discontinuation due to adverse events (antipsychotic + SSRI versus SSRI-only) (follow-up mean 8 weeks; assessed with: Number of participants discontinuing due to adverse events)												
2	randomised trials	serious[19]	no serious inconsistency	no serious indirectness	serious[10]	reporting bias[4]	43/241 (17.8%)	5/226 (2.2%)	RR 6.22 (2.57 to 15.07)	115 more per 1000 (from 35 more to 311 more)	⊕OOO VERY LOW	
								1.2%		63 more per 1000 (from 19 more to 169 more)		

[1] Unclear blinding of intervention administration, and unclear blinding or non-blind outcome assessment
[2] I-squared>50%
[3] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[4] Data cannot be extracted or is not reported for all outcomes and/or funding from pharmaceutical company
[5] Unclear or high risk of bias associated with randomisation method and unclear method of allocation concealment, unclear blinding of intervention administration and outcome assessment, and unclear risk of attrition bias (drop-out>20% and some differences between groups but ITT analysis used) in studies contributing>50% to weighting in analysis
[6] 95% CI crosses line of no effect and both threshold for clinically important harm (RR 0.75) and threshold for clinically important benefit (RR 1.25)
[7] High risk of bias associated with randomisation method due to significant difference between groups at baseline, and unclear blinding of intervention administration and outcome assessment
[8] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 0.75)
[9] Unclear randomisation method and method of allocation concealment, and unclear blinding of intervention administration and outcome assessment
[10] Events<300
[11] I-squared>80%
[12] Unclear randomisation method and method of allocation concealment, unclear or non-blind intervention administration and outcome assessment, and unclear risk of attrition bias (drop-out>20% and some differences between groups but ITT analysis used), in studies contributing >50% to weighting in analysis
[13] Unclear randomisation method and method of allocation concealment, and unclear blinding of intervention administrator. Outcome assessment was non-blind. There was also an unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)
[14] N<400
[15] Unclear randomisation method and method of allocation concealment, and unclear or non-blind intervention administration. There was also an unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used) in studies contributing >50% to weighing in analysis
[16] Unclear randomisation method and method of allocation concealment, and unclear or non-blind intervention administration, in studies contributing >50% to weighting in analysis
[17] Unclear blinding of intervention administration
[18] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 1.25)
[19] Unclear randomisation method and method of allocation concealment, and unclear or non-blind intervention administration
[20] Unclear randomisation method and method of allocation concealment, and unclear blinding of intervention administrator
[21] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

Augmenting the antidepressant with lithium compared to 'other' augmentation agents (head-to-head comparisons)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Augmenting the antidepressant with lithium	'other' augmentation agent	Relative (95% CI)	Absolute		
Remission (lithium versus any 'other' augmentation agent) (follow-up 2-14 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D)/≤8/10 on Montgomery Asberg Depression Rating Scale (MADRS)/<7 on HAM-D AND responding (≥50% improvement on HAM-D))												
7	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	95/382 (24.9%)	123/392 (31.4%)	RR 0.79 (0.63 to 0.99)	66 fewer per 1000 (from 3 fewer to 116 fewer) 62 fewer per 1000 (from 3 fewer to 109 fewer)	⊕OOO VERY LOW	
								29.4%				
Remission (lithium versus TCA) (follow-up mean 4 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	12/48 (25%)	13/46 (28.3%)	RR 0.88 (0.45 to 1.74)	34 fewer per 1000 (from 155 fewer to 209 more) 33 fewer per 1000 (from 150 fewer to 201 more)	⊕OOO VERY LOW	
								27.2%				
Remission (lithium versus antipsychotic) (follow-up 6-8 weeks; assessed with: Number of people scoring ≤8/<10 on Montgomery Asberg Depression Rating Scale (MADRS))												
2	randomised trials	serious[1]	serious[6]	no serious indirectness	very serious[5]	reporting bias[3]	63/231 (27.3%)	81/239 (33.9%)	RR 0.65 (0.31 to 1.39)	119 fewer per 1000 (from 234 fewer to 132 more) 196 fewer per 1000 (from 386 fewer to 218 more)	⊕OOO VERY LOW	
								55.9%				
Remission (lithium versus thyroid hormone [T3]) (follow-up 2-14 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D)/<7 AND responding (≥50% improvement on HAM-D))												

2	randomised trials	very serious[7]	no serious inconsistency	no serious indirectness	serious[8]	reporting bias[3]	17/86 (19.8%)	25/90 (27.8%)	RR 0.72 (0.42 to 1.22)	78 fewer per 1000 (from 161 fewer to 61 more) 92 fewer per 1000 (from 191 fewer to 72 more)	⊕OOO VERY LOW	
								32.9%				
Remission (lithium versus anticonvulsant [lamotrigine]) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	none	3/17 (17.6%)	4/17 (23.5%)	RR 0.75 (0.2 to 2.86)	59 fewer per 1000 (from 188 fewer to 438 more) 59 fewer per 1000 (from 188 fewer to 437 more)	⊕OOO VERY LOW	
								23.5%				
Response (lithium versus any 'other' augmentation agent) (follow-up 6-14 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D)/Montgomery Asberg Depression Rating Scale (MADRS)/Quick Inventory of Depressive Symptomatology (QIDS))												
4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	135/317 (42.6%)	154/329 (46.8%)	RR 0.91 (0.78 to 1.08)	42 fewer per 1000 (from 103 fewer to 37 more) 47 fewer per 1000 (from 116 fewer to 42 more)	⊕OOO VERY LOW	
								52.7%				
Response (lithium versus antipsychotic) (follow-up 6-8 weeks; assessed with: Number of people showing ≥50% improvement on Montgomery Asberg Depression Rating Scale (MADRS))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[8]	reporting bias[3]	117/231 (50.6%)	128/239 (53.6%)	RR 0.89 (0.63 to 1.25)	59 fewer per 1000 (from 198 fewer to 134 more) 73 fewer per 1000 (from 245 fewer to 165 more)	⊕OOO VERY LOW	
								66.2%				
Response (lithium versus thyroid hormone [T3]) (follow-up mean 14 weeks; assessed with: Number of people showing ≥50% improvement on Quick Inventory of Depressive Symptomatology (QIDS))												
1	randomised trials	very serious[9]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	11/69 (15.9%)	17/73 (23.3%)	RR 0.68 (0.35 to 1.36)	75 fewer per 1000 (from 151 fewer to 84 more)		

								23.3%		75 fewer per 1000 (from 151 fewer to 84 more)	⊕OOO VERY LOW	
Response (lithium versus anticonvulsant [lamotrigine]) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	none	7/17 (41.2%)	9/17 (52.9%)	RR 0.78 (0.38 to 1.6)	116 fewer per 1000 (from 328 fewer to 318 more)	⊕OOO VERY LOW	
								52.9%		116 fewer per 1000 (from 328 fewer to 317 more)		
Response (lithium versus antipsychotic) (follow-up mean 6 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[10]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	133/221 (60.2%)	153/229 (66.8%)	RR 0.9 (0.78 to 1.04)	67 fewer per 1000 (from 147 fewer to 27 more)	⊕OOO VERY LOW	
								66.8%		67 fewer per 1000 (from 147 fewer to 27 more)		
Depression symptomatology (lithium versus any 'other' augmentation agent) (follow-up 2-14 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Quick Inventory of Depressive Symptomatology (QIDS; change score); Better indicated by lower values)												
5	randomised trials	serious[11]	no serious inconsistency	no serious indirectness	serious[12]	reporting bias[3]	151	153	-	SMD 0.14 higher (0.14 lower to 0.42 higher)	⊕OOO VERY LOW	
Depression symptomatology (lithium versus TCA) (follow-up mean 4 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
2	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[12]	reporting bias[3]	48	46	-	SMD 0.09 lower (0.49 lower to 0.32 higher)	⊕OOO VERY LOW	
Depression symptomatology (lithium versus thyroid hormone [T3]) (follow-up 2-14 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Quick Inventory of Depressive Symptomatology (QIDS; change score); Better indicated by lower values)												

2	randomised trials	serious[11]	no serious inconsistency	no serious indirectness	serious[12]	reporting bias[3]	86	90	-	SMD 0.15 higher (0.14 lower to 0.45 higher)	⊕OOO VERY LOW	
Depression symptomatology (lithium versus anticonvulsant [lamotrigine]) (follow-up mean 8 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[13]	none	17	17	-	SMD 0.81 higher (0.11 to 1.51 higher)	⊕⊕OO LOW	
Discontinuation for any reason (lithium versus any 'other' augmentation agent) (follow-up 2-8 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
7	randomised trials	serious[10]	no serious inconsistency	no serious indirectness	serious[14]	reporting bias[3]	60/331 (18.1%) 11.8%	45/331 (13.6%) 11.8%	RR 1.29 (0.91 to 1.84)	39 more per 1000 (from 12 fewer to 114 more) 34 more per 1000 (from 11 fewer to 99 more)	⊕OOO VERY LOW	
Discontinuation for any reason (lithium versus TCA) (follow-up mean 4 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[15]	no serious inconsistency	no serious indirectness	very serious[16]	reporting bias[3]	7/48 (14.6%) 19.9%	8/46 (17.4%) 19.9%	RR 0.83 (0.33 to 2.11)	30 fewer per 1000 (from 117 fewer to 193 more) 34 fewer per 1000 (from 133 fewer to 221 more)	⊕OOO VERY LOW	
Discontinuation for any reason (lithium versus antipsychotic) (follow-up 6-8 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[16]	reporting bias[3]	50/239 (20.9%) 7.6%	35/241 (14.5%) 7.6%	RR 1.66 (0.57 to 4.79)	96 more per 1000 (from 62 fewer to 550 more) 50 more per 1000 (from 33 fewer to 288 more)	⊕OOO VERY LOW	
Discontinuation for any reason (lithium versus thyroid hormone [T3]) (follow-up mean 2 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												

2	randomised trials	serious[15]	no serious inconsistency	no serious indirectness	very serious[16]	none	1/27 (3.7%)	0/27 (0%)	RR 2.84 (0.12 to 65.34)	-	⊕OOO VERY LOW	
								0%		-		
Discontinuation for any reason (lithium versus anticonvulsant [lamotrigine]) (follow-up mean 8 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
1	randomised trials	serious[15]	no serious inconsistency	no serious indirectness	very serious[16]	none	2/17 (11.8%)	2/17 (11.8%)	RR 1 (0.16 to 6.3)	0 fewer per 1000 (from 99 fewer to 624 more)	⊕OOO VERY LOW	
								11.8%		0 fewer per 1000 (from 99 fewer to 625 more)		
Discontinuation due to adverse events (lithium versus any 'other' augmentation agent) (follow-up 2-14 weeks; assessed with: Number of participants discontinuing due to adverse events)												
7	randomised trials	serious[17]	no serious inconsistency	no serious indirectness	very serious[16]	reporting bias[3]	37/366 (10.1%)	32/370 (8.6%)	RR 1.27 (0.61 to 2.64)	23 more per 1000 (from 34 fewer to 142 more)	⊕OOO VERY LOW	
								0%		-		
Discontinuation due to adverse events (lithium versus TCA) (follow-up mean 4 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[15]	no serious inconsistency	no serious indirectness	very serious[16]	reporting bias[3]	1/14 (7.1%)	2/12 (16.7%)	RR 0.43 (0.04 to 4.16)	95 fewer per 1000 (from 160 fewer to 527 more)	⊕OOO VERY LOW	
								16.7%		95 fewer per 1000 (from 160 fewer to 528 more)		
Discontinuation due to adverse events (lithium versus antipsychotic) (follow-up 6-8 weeks; assessed with: Number of participants discontinuing due to adverse events)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[16]	reporting bias[3]	19/239 (7.9%)	23/241 (9.5%)	RR 0.83 (0.46 to 1.48)	16 fewer per 1000 (from 52 fewer to 46 more)	⊕OOO VERY LOW	
								5%		9 fewer per 1000 (from 27 fewer to 24 more)		
Discontinuation due to adverse events (lithium versus thyroid hormone [T3]) (follow-up 2-14 weeks; assessed with: Number of participants discontinuing due to adverse events)												

3	randomised trials	serious[17]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	17/96 (17.7%)	7/100 (7%)	RR 2.44 (1.1 to 5.43)	101 more per 1000 (from 7 more to 310 more)	⊕OOO VERY LOW	
								0%		-		
Discontinuation due to adverse events (lithium versus anticonvulsant [lamotrigine]) (follow-up mean 8 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[15]	no serious inconsistency	no serious indirectness	serious[2]	none	0/17 (0%)	0/17 (0%)	not pooled	not pooled	⊕⊕OO LOW	
								0%		not pooled		

[1] Unclear method of allocation concealment and unclear or non-blind intervention administration in studies contributing >50% to weighting in analysis
[2] Events<300
[3] Data cannot be extracted or is not reported for all outcomes and/or funding from pharmaceutical company
[4] Unclear randomisation method and method of allocation concealment, and unclear blinding of intervention administration and outcome assessment
[5] 95% CI crosses line of no effect and threshold for clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)
[6] I-squared>50%
[7] High risk of bias associated with randomisation method due to significant difference between groups at baseline (in studies contributing >50% to weighting in analysis) and unclear method of allocation concealment and unclear blinding of intervention administration
[8] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 0.75)
[9] High risk of bias associated with randomisation method due to significant difference between groups at baseline and unclear method of allocation concealment and unclear blinding of intervention administration
[10] Unclear method of allocation concealment and non-blind intervention administration
[11] Risk associated with randomisation method is high or unclear, the method of allocation concealment is unclear, and blinding of intervention administration and outcome assessment is unclear, in studies contributing to >50% of weighting in analysis
[12] N<400
[13] 95% CI crosses both line of no effect and threshold for clinically important harm (SMD 0.5)
[14] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 1.25)
[15] Unclear randomisation method and method of allocation concealment, and unclear blinding of intervention administration
[16] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)
[17] Risk associated with randomisation method is high or unclear, the method of allocation concealment is unclear, and blinding of intervention administration is unclear, in studies contributing to >50% of weighting in analysis

1 Augmenting the antidepressant with an antipsychotic compared to 'other' augmentation agents (head-to-head comparisons)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Augmenting the antidepressant with an antipsychotic	Any 'other' augmentation agent	Relative (95% CI)	Absolute		
Remission (antipsychotic versus anticonvulsant) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	12/45 (26.7%)	19/39 (48.7%)	RR 0.55 (0.31 to 0.98)	219 fewer per 1000 (from 10 fewer to 336 fewer)	⊕OOO VERY LOW	
								48.7%		219 fewer per 1000 (from 10 fewer to 336 fewer)		
Remission (antipsychotic versus anxiolytic) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	reporting bias[3]	12/45 (26.7%)	15/46 (32.6%)	RR 0.82 (0.43 to 1.55)	59 fewer per 1000 (from 186 fewer to 179 more)	⊕OOO VERY LOW	
								32.6%		59 fewer per 1000 (from 186 fewer to 179 more)		
Remission (antipsychotic versus thyroid hormone) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	reporting bias[3]	12/45 (26.7%)	18/48 (37.5%)	RR 0.71 (0.39 to 1.3)	109 fewer per 1000 (from 229 fewer to 112 more)	⊕OOO VERY LOW	
								37.5%		109 fewer per 1000 (from 229 fewer to 112 more)		
Remission (antipsychotic versus SARI) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	12/45 (26.7%)	20/47 (42.6%)	RR 0.63 (0.35 to 1.13)	157 fewer per 1000 (from 277 fewer to 55 more)	⊕OOO VERY LOW	
								26.7%		99 fewer per 1000 (from 174 fewer to 35 more)		
Response (antipsychotic versus anticonvulsant) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	21/45 (46.7%)	24/39 (61.5%)	RR 0.76 (0.51 to 1.13)	148 fewer per 1000 (from 302 fewer to 80 more)	⊕OOO VERY LOW	
								61.5%		148 fewer per 1000 (from 301 fewer to 80 more)		
Response (antipsychotic versus anxiolytic) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	21/45 (46.7%)	26/46 (56.5%)	RR 0.83 (0.55 to 1.23)	96 fewer per 1000 (from 254 fewer to 130 more)	⊕OOO VERY LOW	
								56.5%		96 fewer per 1000 (from 254 fewer to 130 more)		
Response (antipsychotic versus thyroid hormone) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	21/45 (46.7%)	28/48 (58.3%)	RR 0.8 (0.54 to 1.19)	117 fewer per 1000 (from 268 fewer to 111 more)	⊕OOO VERY LOW	
								58.3%		117 fewer per 1000 (from 268 fewer to 111 more)		
Response (antipsychotic versus SARI) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	21/45 (46.7%)	29/47 (61.7%)	RR 0.76 (0.51 to 1.11)	148 fewer per 1000 (from 302 fewer to 68 more)	⊕OOO VERY LOW	
										112 fewer per 1000 (from 229 fewer to 51 more)		
								46.7%				

[1] High risk of bias associated with randomisation method due to significant difference between groups at baseline, and unclear blinding of intervention administration and outcome assessment
[2] Events<300
[3] Data cannot be extracted or is not reported for all outcomes and/or funding from pharmaceutical company
[4] 95% CI crosses line of no effect and threshold for both clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)
[5] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 0.75)

Augmenting the antidepressant with an anticonvulsant compared to 'other' augmentation agents (head-to-head comparisons)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Augmenting the antidepressant with an anticonvulsant**	**Any 'other' augmentation agent**	**Relative (95% CI)**	**Absolute**		
Remission (anticonvulsant versus anxiolytic) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	19/39 (48.7%)	15/46 (32.6%)	RR 1.49 (0.88 to 2.53)	160 more per 1000 (from 39 fewer to 499 more)	⊕OOO VERY LOW	
										160 more per 1000 (from 39 fewer to 499 more)		
								32.6%				
Remission (anticonvulsant versus SARI) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	reporting bias[3]	19/39 (48.7%)	20/47 (42.6%)	RR 1.14 (0.72 to 1.82)	60 more per 1000 (from 119 fewer to 349 more)	⊕OOO VERY LOW	
										60 more per 1000 (from 119 fewer to 349 more)		
								42.6%				

Remission (anticonvulsant versus thyroid hormone) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))													
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	19/39 (48.7%)	18/48 (37.5%)	RR 1.3 (0.8 to 2.11)	112 more per 1000 (from 75 fewer to 416 more)	⊕OOO VERY LOW		
								37.5%		112 more per 1000 (from 75 fewer to 416 more)			
Response (anticonvulsant versus anxiolytic) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))													
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	24/39 (61.5%)	26/46 (56.5%)	RR 1.09 (0.76 to 1.55)	51 more per 1000 (from 136 fewer to 311 more)	⊕OOO VERY LOW		
								56.5%		51 more per 1000 (from 136 fewer to 311 more)			
Response (anticonvulsant versus SARI) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))													
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	reporting bias[3]	24/39 (61.5%)	29/47 (61.7%)	RR 1 (0.71 to 1.39)	0 fewer per 1000 (from 179 fewer to 241 more)	⊕OOO VERY LOW		
								61.7%		0 fewer per 1000 (from 179 fewer to 241 more)			
Response (anticonvulsant versus thyroid hormone) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))													
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	24/39 (61.5%)	28/48 (58.3%)	RR 1.05 (0.75 to 1.49)	29 more per 1000 (from 146 fewer to 286 more)	⊕OOO VERY LOW		
								58.3%		29 more per 1000 (from 146 fewer to 286 more)			

[1] High risk of bias associated with randomisation method due to significant difference between groups at baseline, and unclear blinding of intervention administration and outcome assessment
[2] 95% crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[3] Data cannot be extracted or is not reported for all outcomes and/or funding from pharmaceutical company
[4] 95% CI crosses line of no effect and both threshold for clinically important harm (RR 0.75) and for clinically important benefit (RR 1.25)

1 Augmenting the antidepressant with an anxiolytic compared to 'other' augmentation agents (head-to-head comparisons)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Augmenting the antidepressant with an anxiolytic	Any 'other' augmentation agent	Relative (95% CI)	Absolute		
Remission (anxiolytic versus atypical antidepressant) (follow-up mean 6 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	86/286 (30.1%)	83/279 (29.7%)	RR 1.01 (0.79 to 1.3)	3 more per 1000 (from 62 fewer to 89 more)	⊕OOO VERY LOW	
								29.8%		3 more per 1000 (from 63 fewer to 89 more)		
Remission (anxiolytic versus SARI) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	15/46 (32.6%)	20/47 (42.6%)	RR 0.77 (0.45 to 1.3)	98 fewer per 1000 (from 234 fewer to 128 more)	⊕OOO VERY LOW	
								42.6%		98 fewer per 1000 (from 234 fewer to 128 more)		
Remission (anxiolytic versus thyroid hormone) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	15/46 (32.6%)	18/48 (37.5%)	RR 0.87 (0.5 to 1.51)	49 fewer per 1000 (from 188 fewer to 191 more)	⊕OOO VERY LOW	
								37.5%		49 fewer per 1000 (from 188 fewer to 191 more)		
Response (anxiolytic versus atypical antidepressant) (follow-up mean 6 weeks; assessed with: Number of people showing ≥50% improvement on Quick Inventory of Depressive Symptomatology (QIDS))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[6]	reporting bias[3]	77/286 (26.9%)	88/279 (31.5%)	RR 0.85 (0.66 to 1.1)	47 fewer per 1000 (from 107 fewer to 32 more)		

								31.5%		47 fewer per 1000 (from 107 fewer to 32 more)	⊕OOO VERY LOW	
Response (anxiolytic versus SARI) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	26/46 (56.5%)	29/47 (61.7%)	RR 0.92 (0.65 to 1.29)	49 fewer per 1000 (from 216 fewer to 179 more)	⊕OOO VERY LOW	
								61.7%		49 fewer per 1000 (from 216 fewer to 179 more)		
Response (anxiolytic versus thyroid hormone) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	26/46 (56.5%)	28/48 (58.3%)	RR 0.97 (0.68 to 1.37)	17 fewer per 1000 (from 187 fewer to 216 more)	⊕OOO VERY LOW	
								58.3%		17 fewer per 1000 (from 187 fewer to 216 more)		
Depression symptomatology (anxiolytic versus atypical antidepressant) (follow-up mean 6 weeks; measured with: Quick Inventory of Depressive Symptomatology (QIDS; change score); Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[3]	286	279	-	SMD 0.17 higher (0.01 to 0.34 higher)	⊕OOO VERY LOW	
Discontinuation due to adverse events (anxiolytic versus atypical antidepressant) (follow-up mean 6 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[7]	reporting bias[3]	59/286 (20.6%)	35/279 (12.5%)	RR 1.64 (1.12 to 2.41)	80 more per 1000 (from 15 more to 177 more)	⊕OOO VERY LOW	
								12.5%		80 more per 1000 (from 15 more to 176 more)		

[1] High risk of bias associated with randomisation method due to significant difference between groups at baseline and unclear method of allocation concealment, and unclear blinding of intervention administration

[2] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)

[3] Data cannot be extracted or is not reported for all outcomes and/or funding from pharmaceutical company

[4] High risk of bias associated with randomisation method due to significant difference between groups at baseline, and unclear blinding of intervention administration and outcome assessment
[5] 95% CI crosses line of no effect and both threshold for clinically important harm (RR 0.75) and for clinically important benefit (RR 1.25)
[6] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 0.75)
[7] Events<300

Augmenting the antidepressant with a thyroid hormone compared to 'other' augmentation agents (head-to-head comparisons)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Augmenting the antidepressant with a thyroid hormone	Any 'other' augmentation agent	Relative (95% CI)	Absolute		
Remission (thyroid hormone versus SARI) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	18/48 (37.5%)	20/47 (42.6%)	RR 0.88 (0.54 to 1.44)	51 fewer per 1000 (from 196 fewer to 187 more)	⊕OOO VERY LOW	
								42.6%		51 fewer per 1000 (from 196 fewer to 187 more)		
Response (thyroid hormone versus SARI) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	28/48 (58.3%)	29/47 (61.7%)	RR 0.95 (0.68 to 1.31)	31 fewer per 1000 (from 197 fewer to 191 more)	⊕OOO VERY LOW	
								61.7%		31 fewer per 1000 (from 197 fewer to 191 more)		

[1] High risk of bias associated with randomisation method due to significant difference between groups at baseline, and unclear blinding of intervention administration and outcome assessment
[2] 95% CI crosses line of no effect and both threshold for clinically important harm (RR 0.75) and for clinically important benefit (RR 1.25)
[3] Data cannot be extracted or is not reported for all outcomes and/or funding from pharmaceutical company

Augmenting the antidepressant with a psychological intervention compared to attention-placebo

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Augmenting the antidepressant with a psych intervention	Attention-placebo	Relative (95% CI)	Absolute		
Remission (Mindfulness-based cognitive therapy [MBCT] versus attention-placebo) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	19/87 (21.8%)	12/86 (14%)	RR 1.57 (0.81 to 3.02)	80 more per 1000 (from 27 fewer to 282 more)	⊕OOO VERY LOW	
								14%		80 more per 1000 (from 27 fewer to 283 more)		
Response (Mindfulness-based cognitive therapy [MBCT] versus attention-placebo) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias[3]	27/87 (31%)	13/86 (15.1%)	RR 2.05 (1.14 to 3.71)	159 more per 1000 (from 21 more to 410 more)	⊕OOO VERY LOW	
								15.1%		159 more per 1000 (from 21 more to 409 more)		
Depression symptomatology (Mindfulness-based cognitive therapy [MBCT] versus attention-placebo) (follow-up mean 8 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[6]	none	23	20	-	MD 5.06 lower (7.78 to 2.34 lower)	⊕⊕OO LOW	
Discontinuation for any reason (Mindfulness-based cognitive therapy [MBCT] versus attention-placebo) (follow-up mean 8 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[7]	reporting bias[3]	15/113 (13.3%)	20/110 (18.2%)	RR 0.73 (0.39 to 1.34)	49 fewer per 1000 (from 111 fewer to 62 more)	⊕OOO VERY LOW	
								20.6%		56 fewer per 1000 (from 126 fewer to 70 more)		

[1] Unclear method of allocation concealment and non-blind intervention administration
[2] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[3] Data cannot be extracted/is not reported for all outcomes
[4] Events<300

[5] Non-blind intervention administration
[6] N<400
[7] 95% CI crosses line of no effect and both threshold for clinically important benefit (RR 0.75) and for clinically important harm (RR 1.25)

Augmenting the antidepressant with a psychological intervention compared to continuing with the antidepressant-only

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Augmenting the antidepressant with a psych intervention	Continuing with the antidepressant-only	Relative (95% CI)	Absolute		
Remission (CBASP + any AD versus any AD) (follow-up mean 12 weeks; assessed with: Number of people scoring <8 on Hamilton Rating Scale for Depression (HAM-D) AND responding (≥50% improvement on HAM-D))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	67/174 (38.5%)	30/76 (39.5%)	RR 0.98 (0.7 to 1.36)	8 fewer per 1000 (from 118 fewer to 142 more)	⊕OOO VERY LOW	
								39.5%		8 fewer per 1000 (from 119 fewer to 142 more)		
Remission (CBT individual [over 15 sessions] + TAU versus TAU) (follow-up 20-27 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D)/<10 on Beck Depression Inventory (BDI))												
2	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	serious[5]	none	76/286 (26.6%)	41/291 (14.1%)	RR 1.89 (1.34 to 2.66)	125 more per 1000 (from 48 more to 234 more)	⊕OOO VERY LOW	
								13.3%		118 more per 1000 (from 45 more to 221 more)		
Remission (CBT individual [under 15 sessions] + TAU versus TAU) (follow-up mean - weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	serious[5]	none	13/21 (61.9%)	4/21 (19%)		429 more per 1000 (from 51		

									RR 3.25 (1.27 to 8.35)	more to 1000 more) 430 more per 1000 (from 52 more to 1000 more)	⊕⊕OO LOW	
								19.1%				
Remission (cognitive and cognitive behavioural therapies [combined] + any AD/TAU versus any AD/TAU-only) (follow-up 12-27 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D)/<8 on HAM-D AND responding (≥50% improvement on HAM-D)/<10 on Beck Depression Inventory (BDI))												
4	randomised trials	serious[1]	serious[7]	no serious indirectness	serious[5]	none	156/481 (32.4%)	75/388 (19.3%)	RR 1.68 (1.02 to 2.78)	131 more per 1000 (from 4 more to 344 more) 116 more per 1000 (from 3 more to 303 more)	⊕OOO VERY LOW	
								17%				
Remission (IPT + TAU versus TAU) (follow-up mean 19 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	very serious[2]	none	5/16 (31.3%)	3/18 (16.7%)	RR 1.88 (0.53 to 6.63)	147 more per 1000 (from 78 fewer to 938 more) 147 more per 1000 (from 78 fewer to 940 more)	⊕OOO VERY LOW	
								16.7%				
Remission (short-term psychodynamic psychotherapy individual + any AD/TAU versus any AD/TAU) (follow-up 12-20 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D)/<8 on HAMD-D AND responding (≥50% improvement on HAM-D))												
2	randomised trials	serious[1]	very serious[9]	no serious indirectness	very serious[2]	reporting bias[10]	63/198 (31.8%)	31/106 (29.2%)	RR 2.5 (0.16 to 39.74)	439 more per 1000 (from 246 fewer to 1000 more) 321 more per 1000 (from 180 fewer to 1000 more)	⊕OOO VERY LOW	
								21.4%				

Remission (long-term psychodynamic psychotherapy + TAU versus TAU) (follow-up mean 78 weeks; assessed with: Number of people scoring ≤8 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[11]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[12]	6/67 (9%)	4/62 (6.5%)	RR 1.39 (0.41 to 4.69)	25 more per 1000 (from 38 fewer to 238 more)	⊕OOO VERY LOW	
								6.5%		25 more per 1000 (from 38 fewer to 240 more)		
Response (any psych + TAU versus TAU-only) (follow-up 19-27 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D)/Beck Depression Inventory (BDI))												
3	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	serious[5]	none	118/243 (48.6%)	55/252 (21.8%)	RR 2.22 (1.7 to 2.9)	266 more per 1000 (from 153 more to 415 more)	⊕OOO VERY LOW	
								22.2%		271 more per 1000 (from 155 more to 422 more)		
Response (CBT individual [over 15 sessions] + TAU versus TAU) (follow-up mean 27 weeks; assessed with: Number of people showing ≥50% improvement on Beck Depression Inventory (BDI))												
1	randomised trials	very serious[13]	no serious inconsistency	no serious indirectness	serious[5]	none	95/206 (46.1%)	46/213 (21.6%)	RR 2.14 (1.59 to 2.87)	246 more per 1000 (from 127 more to 404 more)	⊕OOO VERY LOW	
								21.6%		246 more per 1000 (from 127 more to 404 more)		
Response (CBT individual [under 15 sessions] + TAU versus TAU) (follow-up mean - weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	serious[5]	none	17/21 (81%)	5/21 (23.8%)	RR 3.4 (1.54 to 7.51)	571 more per 1000 (from 129 more to 1000 more)	⊕⊕OO LOW	

								23.8%		571 more per 1000 (from 129 more to 1000 more)		
Response (IPT + TAU versus TAU) (follow-up mean 19 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	very serious[2]	none	6/16 (37.5%)	4/18 (22.2%)	RR 1.69 (0.58 to 4.92)	153 more per 1000 (from 93 fewer to 871 more)	⊕OOO VERY LOW	
								22.2%		153 more per 1000 (from 93 fewer to 870 more)		
Response (cognitive and cognitive behavioural therapies [combined] + TAU versus TAU-only) (follow-up mean 27 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D)/Beck Depression Inventory (BDI))												
2	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	serious[5]	none	112/227 (49.3%)	51/234 (21.8%)	RR 2.32 (1.64 to 3.27)	288 more per 1000 (from 139 more to 495 more)	⊕OOO VERY LOW	
								22.7%		300 more per 1000 (from 145 more to 515 more)		
Depression symptomatology (CBASP + any AD versus any AD) (follow-up mean 12 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[14]	reporting bias[3]	174	76	-	SMD 0.36 lower (0.64 to 0.09 lower)	⊕OOO VERY LOW	
Depression symptomatology (CBT individual [over 15 sessions] + clinical management/TAU versus clinical management/TAU) (follow-up 20-27 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Beck Depression Inventory (BDI; change score); Better indicated by lower values)												
2	randomised trials	very serious[4]	very serious[9]	no serious indirectness	serious[15]	none	286	291	-	SMD 0.41 lower (0.85 lower to 0.04 higher)	⊕OOO VERY LOW	
Depression symptomatology (CBT individual [under 15 sessions] + TAU versus TAU) (follow-up -; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												

1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	serious[14]	none	21	21	-	SMD 1.29 lower (1.96 to 0.62 lower)	⊕⊕OO LOW	
Depression symptomatology (IPT + TAU versus TAU) (follow-up mean 19 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	serious[15]	none	16	18	-	SMD 0.66 lower (1.35 lower to 0.04 higher)	⊕⊕OO LOW	
Depression symptomatology (short-term psychodynamic psychotherapy individual + any AD versus any AD) (follow-up mean 12 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[14]	reporting bias[3]	168	76	-	SMD 0.1 lower (0.37 lower to 0.17 higher)	⊕OOO VERY LOW	
Depression symptomatology (long-term psychodynamic psychotherapy + TAU versus TAU-only) (follow-up mean 78 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	very serious[11]	no serious inconsistency	no serious indirectness	serious[15]	reporting bias[12]	67	62	-	SMD 0.26 lower (0.61 lower to 0.09 higher)	⊕OOO VERY LOW	
Depression symptomatology (cognitive bibliotherapy + any AD versus any AD) (follow-up mean 6 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	serious[15]	none	49	41	-	SMD 0.37 lower (0.79 lower to 0.05 higher)	⊕⊕OO LOW	
Depression symptomatology (mutual peer support + TAU versus TAU) (follow-up mean 24 weeks; measured with: Beck Depression Inventory (BDI-II; change score); Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[14]	reporting bias[16]	127	217	-	SMD 0.03 lower (0.25 lower to 0.19 higher)	⊕OOO VERY LOW	
Depression symptomatology (cognitive and cognitive behavioural therapies [combined] + any AD/TAU versus any AD/TAU-only) (follow-up 12-27 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Beck Depression Inventory (BDI; change score); Better indicated by lower values)												

4	randomised trials	serious[17]	serious[7]	no serious indirectness	no serious imprecision	none	481	388	-	SMD 0.52 lower (0.83 to 0.2 lower)	⊕⊕OO LOW	
Discontinuation for any reason (CBASP + any AD versus any AD) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[18]	reporting bias[3]	25/200 (12.5%)	16/96 (16.7%)	RR 0.75 (0.42 to 1.34)	42 fewer per 1000 (from 97 fewer to 57 more)	⊕OOO VERY LOW	
								16.7%		42 fewer per 1000 (from 97 fewer to 57 more)		
Discontinuation for any reason (CBT individual [over 15 sessions] + clinical management/TAU versus clinical management/TAU) (follow-up 20-27 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[19]	none	44/314 (14%)	34/313 (10.9%)	RR 1.29 (0.85 to 1.96)	32 more per 1000 (from 16 fewer to 104 more)	⊕⊕OO LOW	
								12.4%		36 more per 1000 (from 19 fewer to 119 more)		
Discontinuation for any reason (CBT individual [under 15 sessions] + TAU versus TAU) (assessed with: Number of participants discontinuing for any reason (including adverse events))												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[18]	none	1/21 (4.8%)	2/21 (9.5%)	RR 0.5 (0.05 to 5.1)	48 fewer per 1000 (from 90 fewer to 390 more)	⊕OOO VERY LOW	
								9.5%		47 fewer per 1000 (from 90 fewer to 389 more)		
Discontinuation for any reason (IPT + TAU versus TAU) (follow-up mean 19 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												

1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[18]	none	5/17 (29.4%)	2/23 (8.7%)	RR 3.38 (0.74 to 15.39)	207 more per 1000 (from 23 fewer to 1000 more)	⊕OOO VERY LOW	
								8.7%		207 more per 1000 (from 23 fewer to 1000 more)		
Discontinuation for any reason (short-term psychodynamic psychotherapy individual + any AD/TAU versus any AD/TAU) (follow-up 12-20 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[1]	serious[7]	no serious indirectness	very serious[18]	reporting bias[10]	34/225 (15.1%)	19/126 (15.1%)	RR 1.19 (0.45 to 3.13)	29 more per 1000 (from 83 fewer to 321 more)	⊕OOO VERY LOW	
								13.3%		25 more per 1000 (from 73 fewer to 283 more)		
Discontinuation for any reason (long-term psychodynamic psychotherapy + TAU versus TAU-only) (follow-up mean 78 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
1	randomised trials	very serious[11]	no serious inconsistency	no serious indirectness	very serious[18]	reporting bias[12]	10/67 (14.9%)	8/62 (12.9%)	RR 1.16 (0.49 to 2.74)	21 more per 1000 (from 66 fewer to 225 more)	⊕OOO VERY LOW	
								12.9%		21 more per 1000 (from 66 fewer to 224 more)		
Discontinuation for any reason (cognitive bibliotherapy + any AD versus any AD) (follow-up mean 6 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[18]	none	11/49 (22.4%)	6/41 (14.6%)	RR 1.53 (0.62 to 3.79)	78 more per 1000 (from 56 fewer to 408 more)	⊕OOO VERY LOW	
								14.6%		77 more per 1000 (from 55		

										fewer to 407 more)		
Discontinuation for any reason (mutual peer support + TAU versus TAU) (follow-up mean 24 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[18]	reporting bias[16]	15/144 (10.4%)	26/243 (10.7%)	RR 0.97 (0.53 to 1.78)	3 fewer per 1000 (from 50 fewer to 83 more)	⊕OOO VERY LOW	
								10.7%		3 fewer per 1000 (from 50 fewer to 83 more)		
Discontinuation for any reason (cognitive and cognitive behavioural therapies [combined] + any AD/TAU versus any AD/TAU-only) (follow-up 12-27 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
4	randomised trials	serious[17]	no serious inconsistency	no serious indirectness	very serious[18]	none	70/535 (13.1%)	52/430 (12.1%)	RR 1.06 (0.75 to 1.49)	7 more per 1000 (from 30 fewer to 59 more)	⊕OOO VERY LOW	
								12.5%		7 more per 1000 (from 31 fewer to 61 more)		
Discontinuation due to adverse events (CBASP + any AD versus any AD) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[18]	reporting bias[3]	2/200 (1%)	2/96 (2.1%)	RR 0.48 (0.07 to 3.36)	11 fewer per 1000 (from 19 fewer to 49 more)	⊕OOO VERY LOW	
								2.1%		11 fewer per 1000 (from 20 fewer to 50 more)		
Discontinuation due to adverse events (short-term psychodynamic psychotherapy individual + any AD versus any AD) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[18]	reporting bias[3]	1/195 (0.5%)	2/96 (2.1%)	RR 0.25 (0.02 to 2.68)	16 fewer per 1000 (from 20 fewer to 35 more)	⊕OOO VERY LOW	
								2.1%		16 fewer per 1000 (from 21		

										fewer to 35 more)		

[1] Method of randomisation was unclear, and non-blind participants and intervention administrator(s)
[2] 95% CI crosses line of no effect and threshold for both clinically important harm (RR 0.75) and for clinically important benefit (RR 1.25)
[3] Drugs were supplied at no cost by pharmaceutical company and authors have financial interests with pharmaceutical companies
[4] High risk of bias associated with randomisation method due to significant difference between groups at baseline, non-blind participants and intervention administrator(s), and unclear blinding of outcome assessment, in studies contributing >50% of weighting in analysis
[5] Events<300
[6] Non-blind participants and intervention administrator(s)
[7] I-squared>50%
[8] Non-blind participants and intervention administrator(s) and potential risk of attrition bias (difference in drop-out between groups>20% but ITT analysis used)
[9] I-squared>80%
[10] Data cannot be extracted or is not reported for all outcomes and/or drugs were supplied at no cost by pharmaceutical company and authors have financial interests with pharmaceutical companies
[11] High risk of bias associated with randomisation method due to significant difference between groups at baseline, non-blind participants and intervention administrator(s)
[12] Study partially funded by the International Psychoanalytic Association
[13] High risk of bias associated with randomisation method due to significant difference between groups at baseline, non-blind participants and intervention administrator(s), and unclear blinding of outcome assessment
[14] N<400
[15] 95% CI crosses both line of no effect and threshold for clinically important benefit (SMD -0.5)
[16] Data cannot be extracted/is not reported for all outcomes
[17] High or unclear risk of randomisation method and participants and intervention administrator(s) were non-blind
[18] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)
[19] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 1.25)

Augmenting the antidepressant with a psychological intervention compared to augmenting with a non-antidepressant agent

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Augmenting the antidepressant with a psych intervention**	**Augmenting with a non-AD agent**	**Relative (95% CI)**	**Absolute**		
Remission (CBT individual [under 15 sessions] + AD versus lithium + AD) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	6/23 (26.1%)	8/21 (38.1%)	RR 0.68 (0.28 to 1.65)	122 fewer per 1000 (from 274 fewer to 248 more)	⊕OOO VERY LOW	

								38.1%		122 fewer per 1000 (from 274 fewer to 248 more)		
Depression symptomatology (CBT individual [under 15 sessions] + AD versus lithium + AD) (follow-up mean 8 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	23	21	-	SMD 0.7 higher (0.09 to 1.31 higher)	⊕⊕OO LOW	
Discontinuation for any reason (CBT individual [under 15 sessions] + AD versus lithium + AD) (follow-up mean 8 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	6/23 (26.1%)	6/21 (28.6%)	RR 0.91 (0.35 to 2.4)	26 fewer per 1000 (from 186 fewer to 400 more)	⊕OOO VERY LOW	
								28.6%		26 fewer per 1000 (from 186 fewer to 400 more)		
Discontinuation due to adverse events (CBT individual [under 15 sessions] + AD versus lithium + AD) (follow-up mean 8 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	0/23 (0%)	1/21 (4.8%)	RR 0.31 (0.01 to 7.12)	33 fewer per 1000 (from 47 fewer to 291 more)	⊕OOO VERY LOW	
								4.8%		33 fewer per 1000 (from 48 fewer to 294 more)		

[1] Unclear method of randomisation and allocation concealment, non-blind participants and intervention administrator(s), and unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)
[2] 95% CI crosses line of no effect and threshold for both clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)
[3] N<400
[4] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

Augmenting the antidepressant with a psychological intervention compared to 'other' psychological intervention (head-to-head comparisons)

Quality assessment							No of patients		Effect			Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Augmenting the antidepressant with a psych intervention [head-to-head]	Control	Relative (95% CI)	Absolute		
Remission (CBASP + any AD versus short-term psychodynamic psychotherapy individual + any AD) (follow-up mean 12 weeks; assessed with: Number of people scoring <8 on Hamilton Rating Scale for Depression (HAM-D) AND responding (≥50% improvement on HAM-D))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	67/174 (38.5%)	52/168 (31%)	RR 1.24 (0.93 to 1.67)	74 more per 1000 (from 22 fewer to 207 more)	⊕⊕OO LOW	
								31%		74 more per 1000 (from 22 fewer to 208 more)		
Depression symptomatology (CBASP + any AD versus short-term psychodynamic psychotherapy individual + any AD) (follow-up mean 12 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	174	168	-	SMD 0.26 lower (0.48 to 0.05 lower)	⊕⊕OO LOW	
Discontinuation for any reason (CBASP + any AD versus short-term psychodynamic psychotherapy individual + any AD) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	25/200 (12.5%)	27/195 (13.8%)	RR 0.9 (0.54 to 1.5)	14 fewer per 1000 (from 64 fewer to 69 more)	⊕OOO VERY LOW	
								13.9%		14 fewer per 1000 (from 64 fewer to 69 more)		
Discontinuation due to adverse events (CBASP + any AD versus short-term psychodynamic psychotherapy individual + any AD) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	2/200 (1%)	1/195 (0.5%)	RR 1.95 (0.18 to 21.33)	5 more per 1000 (from 4 fewer to 104 more)	⊕OOO VERY LOW	
								0.5%		5 more per 1000 (from 4 fewer to 102 more)		

[1] Method of randomisation was unclear, and non-blind participants and intervention administrator(s)
[2] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)

[3] N<400

[4] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

Augmenting the antidepressant with a physical intervention compared to attention-placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Augmenting the antidepressant with a physical intervention	Attention-placebo	Relative (95% CI)	Absolute		
Remission (exercise + SSRI/any AD versus attention-placebo + SSRI/any AD) (follow-up 10-12 weeks; assessed with: Number of people scoring ≤7/10 on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	39/55 (70.9%)	28/47 (59.6%)	RR 1.77 (0.37 to 8.41)	459 more per 1000 (from 375 fewer to 1000 more)	⊕OOO VERY LOW	
								37.8%		291 more per 1000 (from 238 fewer to 1000 more)		
Response (exercise + any AD versus attention-placebo + any AD) (follow-up mean 12 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[4]	4/19 (21.1%)	0/10 (0%)	RR 4.95 (0.29 to 83.68)	-	⊕OOO VERY LOW	
								0%		-		
Depression symptomatology (exercise + SSRI/any AD versus attention-placebo + SSRI/any AD) (follow-up 10-12 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
2	randomised trials	serious[1]	very serious[5]	no serious indirectness	serious[6]	none	52	45	-	SMD 0.4 lower (0.86 lower to 0.06 higher)	⊕OOO VERY LOW	
Discontinuation for any reason (exercise + SSRI/any AD versus attention-placebo + SSRI/any AD) (follow-up 10-12 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[7]	none	6/58 (10.3%)	3/48 (6.3%)	RR 1.53 (0.4 to 5.86)	33 more per 1000 (from 38 fewer to 304 more)		

								7.3%		39 more per 1000 (from 44 fewer to 355 more)	⊕OOO VERY LOW	

[1] Non-blind intervention administration
[2] 95% CI crosses line of no effect and threshold for both clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)
[3] High risk of bias associated with randomisation method due to significant difference between groups at baseline and unclear method of allocation concealment. Intervention administration was non-blind
[4] Study partially funded by pharmaceutical company
[5] I-squared>80%
[6] 95% CI crosses both line of no effect and threshold for clinically important benefit (SMD -0.5)
[7] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

Switching to another antidepressant of a different class compared to placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Switch to another antidepressant of different class**	**Placebo**	**Relative (95% CI)**	**Absolute**		
Remission (SSRI to atypical antidepressant or placebo) (follow-up mean 12 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	40/165 (24.2%)	39/157 (24.8%)	RR 0.98 (0.67 to 1.43)	5 fewer per 1000 (from 82 fewer to 107 more)	⊕OOO VERY LOW	
								24.8%		5 fewer per 1000 (from 82 fewer to 107 more)		
Response (SSRI to atypical antidepressant or placebo) (follow-up mean 12 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias[3]	63/165 (38.2%)	58/157 (36.9%)	RR 1.03 (0.78 to 1.37)	11 more per 1000 (from 81 fewer to 137 more)	⊕OOO VERY LOW	
								36.9%		11 more per 1000 (from 81 fewer to 137 more)		

Response (SSRI to atypical antidepressant or placebo) (follow-up mean 12 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias[3]	79/165 (47.9%)	69/157 (43.9%)	RR 1.09 (0.86 to 1.38)	40 more per 1000 (from 62 fewer to 167 more)	⊕OOO VERY LOW	
								44%		40 more per 1000 (from 62 fewer to 167 more)		
Depression symptomatology (SSRI to atypical antidepressant or placebo) (follow-up mean 12 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	165	157	-	SMD 0.02 higher (0.19 lower to 0.24 higher)	⊕OOO VERY LOW	
Discontinuation for any reason (SSRI to atypical antidepressant or placebo) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[6]	reporting bias[3]	67/166 (40.4%)	47/159 (29.6%)	RR 1.37 (1.01 to 1.85)	109 more per 1000 (from 3 more to 251 more)	⊕OOO VERY LOW	
								29.6%		110 more per 1000 (from 3 more to 252 more)		
Discontinuation due to adverse events (SSRI to atypical antidepressant or placebo) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[7]	reporting bias[3]	39/166 (23.5%)	31/159 (19.5%)	RR 1.21 (0.79 to 1.83)	41 more per 1000 (from 41 fewer to 162 more)	⊕OOO VERY LOW	
								19.5%		41 more per 1000 (from 41 fewer to 162 more)		

[1] Unclear randomisation method and method of allocation concealment, and unclear risk of attrition bias (drop-out>20% but difference between groups <20% and ITT analysis used)
[2] 95% CI crosses line of no effect and threshold for both clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)
[3] Study run and funded by pharmaceutical company
[4] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[5] N<400
[6] Events<300
[7] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 1.25)

1 Switching to another antidepressant of a different class compared to continuing with the same antidepressant

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Switch to another antidepressant of a different class	Continuing with the antidepressant	Relative (95% CI)	Absolute		
Remission (any switch versus continuing with the antidepressant) (follow-up 6-12 weeks; assessed with: Number of people scoring ≤7/8 on Hamilton Rating Scale for Depression (HAM-D)/≤8 on Montgomery Asberg Depression Rating Scale (MADRS))												
4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	82/336 (24.4%)	53/209 (25.4%)	RR 0.93 (0.65 to 1.34)	18 fewer per 1000 (from 89 fewer to 86 more)	⊕OOO VERY LOW	
								20.4%		14 fewer per 1000 (from 71 fewer to 69 more)		
Remission (switch to SSRI versus continuing TCA/SNRI) (follow-up 8-12 weeks; assessed with: Number of people scoring ≤8 on Montgomery Asberg Depression Rating Scale (MADRS))												
2	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	29/198 (14.6%)	25/126 (19.8%)	RR 0.78 (0.47 to 1.27)	44 fewer per 1000 (from 105 fewer to 54 more)	⊕OOO VERY LOW	
								20%		44 fewer per 1000 (from 106 fewer to 54 more)		
Remission (switch to atypical AD/SNRI/TeCA [mianserin] versus continuing SSRI) (follow-up 6-8 weeks; assessed with: Number of people scoring ≤7/8 on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	very serious[5]	serious[6]	no serious indirectness	very serious[2]	reporting bias[3]	53/138 (38.4%)	28/83 (33.7%)	RR 1.19 (0.52 to 2.77)	64 more per 1000 (from 162 fewer to 597 more)	⊕OOO VERY LOW	
								32.5%		62 more per 1000 (from 156		

										fewer to 575 more)		
Response (any switch versus continuing with the antidepressant) (follow-up 6-12 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D)/Montgomery Asberg Depression Rating Scale (MADRS))												
4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[7]	reporting bias[3]	140/336 (41.7%)	94/209 (45%)	RR 0.91 (0.74 to 1.12)	40 fewer per 1000 (from 117 fewer to 54 more)	⊕OOO VERY LOW	
								43.4%		39 fewer per 1000 (from 113 fewer to 52 more)		
Response (switch to SSRI versus continuing TCA/SNRI) (follow-up 8-12 weeks; assessed with: Number of people showing ≥50% improvement on Montgomery Asberg Depression Rating Scale (MADRS))												
2	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[7]	reporting bias[3]	60/198 (30.3%)	50/126 (39.7%)	RR 0.8 (0.58 to 1.09)	79 fewer per 1000 (from 167 fewer to 36 more)	⊕OOO VERY LOW	
								40.4%		81 fewer per 1000 (from 170 fewer to 36 more)		
Response (switch to atypical AD/SNRI/TeCA [mianserin] versus continuing SSRI) (follow-up 6-8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	80/138 (58%)	44/83 (53%)	RR 1.01 (0.73 to 1.41)	5 more per 1000 (from 143 fewer to 217 more)	⊕OOO VERY LOW	
								51.8%		5 more per 1000 (from 140 fewer to 212 more)		
Response (switch to TeCA [mianserin] versus continuing SSRI) (follow-up mean 6 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	serious[9]	reporting bias[3]	21/33 (63.6%)	17/38 (44.7%)	RR 1.42 (0.92 to 2.2)	188 more per 1000 (from 36 fewer to 537 more)	⊕OOO VERY LOW	

								44.7%		188 more per 1000 (from 36 fewer to 536 more)		
Depression symptomatology (any switch versus continuing with the antidepressant) (follow-up 6-12 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
3	randomised trials	serious[10]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[3]	235	165	-	SMD 0.04 lower (0.3 lower to 0.23 higher)	⊕⊕OO LOW	
Depression symptomatology (switch to SSRI versus continuing TCA/SNRI) (follow-up 8-12 weeks; measured with: Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
2	randomised trials	serious[4]	serious[6]	no serious indirectness	serious[11]	reporting bias[3]	202	127	-	SMD 0.03 higher (0.31 lower to 0.38 higher)	⊕OOO VERY LOW	
Depression symptomatology (switch to TeCA [mianserin] versus continuing SSRI) (follow-up mean 6 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	serious[12]	reporting bias[3]	33	38	-	SMD 0.24 lower (0.71 lower to 0.23 higher)	⊕OOO VERY LOW	
Discontinuation for any reason (any switch versus continuing with the antidepressant) (follow-up 6-12 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
4	randomised trials	serious[13]	no serious inconsistency	no serious indirectness	serious[14]	reporting bias[3]	71/341 (20.8%)	38/210 (18.1%)	RR 1.23 (0.81 to 1.86)	42 more per 1000 (from 34 fewer to 156 more)	⊕OOO VERY LOW	
								18.1%		42 more per 1000 (from 34 fewer to 156 more)		
Discontinuation for any reason (switch to SSRI versus continuing TCA/SNRI) (follow-up 8-12 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[15]	serious[6]	no serious indirectness	very serious[16]	reporting bias[3]	40/202 (19.8%)	23/127 (18.1%)		24 more per 1000 (from 83		

									RR 1.13 (0.54 to 2.38)	fewer to 250 more) 24 more per 1000 (from 86 fewer to 257 more)	⊕OOO VERY LOW	
								18.6%				
Discontinuation for any reason (switch to atypical AD/SNRI/TeCA [mianserin] versus continuing SSRI) (follow-up 6-8 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[17]	no serious inconsistency	no serious indirectness	very serious[16]	reporting bias[3]	31/139 (22.3%)	15/83 (18.1%)	RR 1.37 (0.74 to 2.54)	67 more per 1000 (from 47 fewer to 278 more) 67 more per 1000 (from 47 fewer to 279 more)	⊕OOO VERY LOW	
								18.1%				
Discontinuation due to adverse events (any switch versus continuing with the antidepressant) (follow-up 6-12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
4	randomised trials	serious[13]	serious[6]	no serious indirectness	very serious[16]	reporting bias[3]	15/336 (4.5%)	4/210 (1.9%)	RR 1.74 (0.32 to 9.6)	14 more per 1000 (from 13 fewer to 164 more) 15 more per 1000 (from 14 fewer to 172 more)	⊕OOO VERY LOW	
								2%				
Discontinuation due to adverse events (switch to SSRI versus continuing TCA/SNRI) (follow-up 8-12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
2	randomised trials	serious[15]	no serious inconsistency	no serious indirectness	very serious[16]	reporting bias[3]	7/202 (3.5%)	3/127 (2.4%)	RR 1.43 (0.38 to 5.47)	10 more per 1000 (from 15 fewer to 106 more) 10 more per 1000 (from 14 fewer to 103 more)	⊕OOO VERY LOW	
								2.3%				

Discontinuation due to adverse events (switch to atypical AD/SNRI/TeCA [mianserin] versus continuing SSRI) (follow-up 6-8 weeks; assessed with: Number of participants discontinuing due to adverse events)												
2	randomised trials	serious[17]	very serious[18]	no serious indirectness	very serious[16]	reporting bias[3]	8/134 (6%)	1/83 (1.2%)	RR 1.8 (0.01 to 222.73)	10 more per 1000 (from 12 fewer to 1000 more)	⊕OOO VERY LOW	
								1.1%		9 more per 1000 (from 11 fewer to 1000 more)		

[1] Risk of randomisation method is high risk or unclear, method of allocation concealment is unclear, intervention administration is non-blind, risk of detection bias is high or unclear, in studies contributing>50% to weighting in analysis
[2] 95% CI crosses line of no effect and threshold for both clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)
[3] Funded by pharmaceutical company
[4] Unclear randomisation method and method of allocation concealment, and unclear blinding of intervention administration and outcome assessment
[5] High risk of bias associated with randomisation method due to significant difference between groups at baseline, and unclear blinding of intervention administration and unclear blinding or non-blind outcome assessment, in studies contributing >50% to weighting in analysis
[6] I-squared>50%
[7] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 0.75)
[8] Unclear randomisation method and method of allocation concealment, unclear blinding of intervention administration, non-blind outcome assessment and unclear risk of attrition bias (drop-out>20% but difference between groups <20% and ITT analysis used)
[9] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[10] Unclear randomisation method and method of allocation concealment, unclear blinding of intervention administration, unclear blinding or non-blind outcome assessment, and unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)
[11] N<400
[12] 95% CI crosses both line of no effect and threshold for clinically important benefit (SMD -0.5)
[13] Unclear or high risk of bias associated with randomisation method, method of allocation concealment is unclear and unclear blinding of intervention administration
[14] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 1.25)
[15] Unclear randomisation method and method of allocation concealment, and unclear blinding of intervention administration
[16] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)
[17] Risk of randomisation method is high or unclear and unclear blinding of intervention administration
[18] I-squared>80%

Switching to a non-antidepressant agent compared to continuing with the antidepressant

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Switch to non-antidepressant agent	Continuing with the antidepressant	Relative (95% CI)	Absolute		
Remission (switch to antipsychotic monotherapy versus continuing SSRI/TCA/SNRI) (follow-up 8-12 weeks; assessed with: Number of people scoring ≤8/10 on Montgomery Asberg Depression Rating Scale (MADRS))												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	56/400 (14%)	59/329 (17.9%)	RR 0.79 (0.56 to 1.11)	38 fewer per 1000 (from 79 fewer to 20 more)	⊕OOO VERY LOW	
								17.7%		37 fewer per 1000 (from 78 fewer to 19 more)		
Remission (switch to combined antipsychotic + SSRI versus continuing TCA/SNRI) (follow-up 8-12 weeks; assessed with: Number of people scoring ≤8 on Montgomery Asberg Depression Rating Scale (MADRS))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias[3]	94/376 (25%)	25/126 (19.8%)	RR 1.17 (0.79 to 1.75)	34 more per 1000 (from 42 fewer to 149 more)	⊕OOO VERY LOW	
								20%		34 more per 1000 (from 42 fewer to 150 more)		
Response (switch to antipsychotic monotherapy versus continuing SSRI/TCA/SNRI) (follow-up 8-12 weeks; assessed with: Number of people showing ≥50% improvement on Montgomery Asberg Depression Rating Scale (MADRS))												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	94/400 (23.5%)	110/329 (33.4%)	RR 0.69 (0.49 to 0.96)	104 fewer per 1000 (from 13 fewer to 171 fewer)	⊕OOO VERY LOW	
								30.9%		96 fewer per 1000 (from 12 fewer to 158 fewer)		
Response (switch to combined antipsychotic + SSRI versus continuing TCA/SNRI) (follow-up 8-12 weeks; assessed with: Number of people showing ≥50% improvement on Montgomery Asberg Depression Rating Scale (MADRS))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	140/376 (37.2%)	50/126 (39.7%)	RR 0.87 (0.68 to 1.12)	52 fewer per 1000 (from 127 fewer to 48 more)		

								40.4%		53 fewer per 1000 (from 129 fewer to 48 more)	⊕OOO VERY LOW	
Depression symptomatology (switch to antipsychotic monotherapy versus continuing SSRI/TCA/SNRI) (follow-up 8-12 weeks; measured with: Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
3	randomised trials	serious[1]	serious[6]	no serious indirectness	serious[7]	reporting bias[3]	403	330	-	SMD 0.22 higher (0.12 lower to 0.57 higher)	⊕OOO VERY LOW	
Depression symptomatology (switch to combined antipsychotic + SSRI versus continuing TCA/SNRI) (follow-up 8-12 weeks; measured with: Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[3]	389	127	-	SMD 0.09 lower (0.29 lower to 0.11 higher)	⊕⊕OO LOW	
Discontinuation for any reason (switch to antipsychotic monotherapy versus continuing SSRI/TCA/SNRI) (follow-up 8-12 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
3	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	122/405 (30.1%)	63/333 (18.9%)	RR 1.67 (1.26 to 2.23)	127 more per 1000 (from 49 more to 233 more)	⊕OOO VERY LOW	
								19.4%		130 more per 1000 (from 50 more to 239 more)		
Discontinuation for any reason (switch to combined antipsychotic + SSRI versus continuing TCA/SNRI) (follow-up 8-12 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	very serious[9]	reporting bias[3]	90/389 (23.1%)	23/127 (18.1%)	RR 1.22 (0.69 to 2.16)	40 more per 1000 (from 56 fewer to 210 more)	⊕OOO VERY LOW	
								18.6%		41 more per 1000 (from 58 fewer to 216 more)		
Discontinuation due to adverse events (switch to antipsychotic monotherapy versus continuing SSRI/TCA/SNRI) (follow-up 8-12 weeks; assessed with: Number of participants discontinuing due to adverse events)												

3	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	51/405 (12.6%)	8/333 (2.4%)	RR 5.34 (2.57 to 11.09)	104 more per 1000 (from 38 more to 242 more)	⊕OOO VERY LOW	
								2.4%		104 more per 1000 (from 38 more to 242 more)		
Discontinuation due to adverse events (switch to combined antipsychotic + SSRI versus continuing TCA/SNRI) (follow-up 8-12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
2	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	39/389 (10%)	3/127 (2.4%)	RR 3.48 (1.06 to 11.44)	59 more per 1000 (from 1 more to 247 more)	⊕OOO VERY LOW	
								2.3%		57 more per 1000 (from 1 more to 240 more)		

[1] Unclear randomisation method and method of allocation concealment, and unclear blinding of intervention administration and outcome assessment, and unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used) in studies contributing >50% to weighting in analysis
[2] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 0.75)
[3] Funding from pharmaceutical companies
[4] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[5] Events<300
[6] I-squared>50%
[7] 95% CI crosses both line of no effect and threshold for clinically important harm (SMD 0.5)
[8] Unclear randomisation method and method of allocation concealment, and unclear blinding of intervention administration
[9] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

Switching to another antidepressant or non-antidepressant agent compared to augmenting with another antidepressant or non-antidepressant agent

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Switch to another antidepressant/non-antidepressant agent**	**Augmentation with another antidepressant/non-antidepressant agent**	**Relative (95% CI)**	**Absolute**		
Remission (switch to TeCA versus augmentation with TeCA [mianserin]) (follow-up mean 6 weeks; assessed with: Number of people scoring ≤8 on Hamilton Rating Scale for Depression (HAM-D))												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	12/33 (36.4%)	14/32 (43.8%)	RR 0.83 (0.46 to 1.51)	74 fewer per 1000 (from 236 fewer to 223 more) 74 fewer per 1000 (from 237 fewer to 223 more)	⊕OOO VERY LOW	
								43.8%				
Remission (switch to antipsychotic versus augmentation with antipsychotic) (follow-up 6-8 weeks; assessed with: Number of people scoring ≤10/<10 on Montgomery Asberg Depression Rating Scale (MADRS))												
2	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[6]	82/422 (19.4%)	127/427 (29.7%)	RR 0.65 (0.48 to 0.88)	104 fewer per 1000 (from 36 fewer to 155 fewer) 104 fewer per 1000 (from 36 fewer to 154 fewer)	⊕OOO VERY LOW	
								29.6%				
Remission (switch to antipsychotic versus augmentation with lithium) (follow-up mean 6 weeks; assessed with: Number of people scoring <10 on Montgomery Asberg Depression Rating Scale (MADRS))												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[8]	reporting bias[6]	53/225 (23.6%)	60/221 (27.1%)	RR 0.87 (0.63 to 1.19)	35 fewer per 1000 (from 100 fewer to 52 more) 35 fewer per 1000 (from 101 fewer to 52 more)	⊕OOO VERY LOW	
								27.2%				
Response (switch to TeCA versus augmentation with TeCA [mianserin]) (follow-up mean 6 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[8]	reporting bias[3]	16/33 (48.5%)	20/32 (62.5%)	RR 0.78 (0.5 to 1.21)	138 fewer per 1000 (from 312 fewer to 131 more) 138 fewer per 1000 (from	⊕OOO VERY LOW	
								62.5%				

										312 fewer to 131 more)		
Response (switch to antipsychotic versus augmentation with antipsychotic) (follow-up 6-8 weeks; assessed with: Number of people showing ≥50% improvement on Montgomery Asberg Depression Rating Scale (MADRS))												
2	randomised trials	serious[4]	very serious[9]	no serious indirectness	serious[8]	reporting bias[6]	165/422 (39.1%)	200/427 (46.8%)	RR 0.8 (0.53 to 1.2)	94 fewer per 1000 (from 220 fewer to 94 more)	⊕OOO VERY LOW	
								46.4%		93 fewer per 1000 (from 218 fewer to 93 more)		
Response (switch to antipsychotic versus augmentation with lithium) (follow-up mean 6 weeks; assessed with: Number of people showing ≥50% improvement on Montgomery Asberg Depression Rating Scale (MADRS))												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[6]	114/225 (50.7%)	112/221 (50.7%)	RR 1 (0.83 to 1.2)	0 fewer per 1000 (from 86 fewer to 101 more)	⊕OOO VERY LOW	
								50.7%		0 fewer per 1000 (from 86 fewer to 101 more)		
Response (switch to TeCA versus augmentation with TeCA [mianserin]) (follow-up mean 6 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[8]	reporting bias[3]	21/33 (63.6%)	23/32 (71.9%)	RR 0.89 (0.63 to 1.24)	79 fewer per 1000 (from 266 fewer to 173 more)	⊕OOO VERY LOW	
								71.9%		79 fewer per 1000 (from 266 fewer to 173 more)		
Response (switch to antipsychotic versus augmentation with antipsychotic) (follow-up mean 6 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[6]	139/225 (61.8%)	153/229 (66.8%)		53 fewer per 1000 (from		

									RR 0.92 (0.81 to 1.06)	127 fewer to 40 more) 53 fewer per 1000 (from 127 fewer to 40 more)	⊕OOO VERY LOW	
								66.8%				
Response (switch to antipsychotic versus augmentation with lithium) (follow-up mean 6 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[6]	139/225 (61.8%)	133/221 (60.2%)	RR 1.03 (0.88 to 1.19)	18 more per 1000 (from 72 fewer to 114 more) 18 more per 1000 (from 72 fewer to 114 more)	⊕OOO VERY LOW	
								60.2%				
Depression symptomatology (any switch versus any augmentation) (follow-up 6-8 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
2	randomised trials	serious[10]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[3]	230	230	-	SMD 0.39 higher (0.2 to 0.57 higher)	⊕⊕OO LOW	
Depression symptomatology (switch to TeCA versus augmentation with TeCA [mianserin]) (follow-up mean 6 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[11]	reporting bias[3]	33	32	-	SMD 0.41 higher (0.08 lower to 0.91 higher)	⊕OOO VERY LOW	
Depression symptomatology (switch to antipsychotic versus augmentation with antipsychotic) (follow-up mean 8 weeks; measured with: Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
1	randomised trials	serious[12]	no serious inconsistency	no serious indirectness	serious[13]	reporting bias[3]	197	198	-	SMD 0.38 higher (0.18 to 0.58 higher)	⊕OOO VERY LOW	

Discontinuation for any reason (switch to TeCA versus augmentation with TeCA [mianserin]) (follow-up mean 6 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))													
1	randomised trials	serious[14]	no serious inconsistency	no serious indirectness	serious[15]	reporting bias[3]	12/34 (35.3%)	6/32 (18.8%)	RR 1.88 (0.8 to 4.42)	165 more per 1000 (from 37 fewer to 641 more)	⊕OOO VERY LOW	
								18.8%		165 more per 1000 (from 38 fewer to 643 more)		
Discontinuation for any reason (switch to antipsychotic versus augmentation with antipsychotic) (follow-up 6-8 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))													
2	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[6]	121/427 (28.3%)	87/431 (20.2%)	RR 1.4 (1.11 to 1.78)	81 more per 1000 (from 22 more to 157 more)	⊕OOO VERY LOW	
								20.6%		82 more per 1000 (from 23 more to 161 more)		
Discontinuation for any reason (switch to antipsychotic versus augmentation with lithium) (follow-up mean 6 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))													
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	very serious[16]	reporting bias[6]	49/228 (21.5%)	47/229 (20.5%)	RR 1.05 (0.73 to 1.49)	10 more per 1000 (from 55 fewer to 101 more)	⊕OOO VERY LOW	
								20.5%		10 more per 1000 (from 55 fewer to 100 more)		
Discontinuation due to adverse events (switch to TeCA versus augmentation with TeCA [mianserin]) (follow-up mean 6 weeks; assessed with: Number of participants discontinuing due to adverse events)													
1	randomised trials	serious[14]	no serious inconsistency	no serious indirectness	serious[15]	reporting bias[3]	8/34 (23.5%)	2/32 (6.3%)	RR 3.76 (0.86 to 16.41)	172 more per 1000 (from 9 fewer to 963 more)	⊕OOO VERY LOW	

								6.3%		174 more per 1000 (from 9 fewer to 971 more)		
Discontinuation due to adverse events (switch to antipsychotic versus augmentation with antipsychotic) (follow-up 6-8 weeks; assessed with: Number of participants discontinuing due to adverse events)												
2	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[15]	reporting bias[6]	60/427 (14.1%)	50/431 (11.6%)	RR 1.21 (0.85 to 1.72)	24 more per 1000 (from 17 fewer to 84 more)	⊕OOO VERY LOW	
								11.7%		25 more per 1000 (from 18 fewer to 84 more)		
Discontinuation due to adverse events (switch to antipsychotic versus augmentation with lithium) (follow-up mean 6 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[15]	reporting bias[6]	28/228 (12.3%)	18/229 (7.9%)	RR 1.56 (0.89 to 2.74)	44 more per 1000 (from 9 fewer to 137 more)	⊕OOO VERY LOW	
								7.9%		44 more per 1000 (from 9 fewer to 137 more)		

[1] Unclear randomisation method and method of allocation concealment and unclear blinding of intervention administration. Risk of attrition bias was also unclear (drop-out>20% but difference between groups<20% and ITT analysis used). Outcome assessment was non-blind

[2] 95% CI crosses line of no effect and threshold for both clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)

[3] Funding from pharmaceutical companies

[4] Unclear method of allocation concealment and unclear blinding of, or non-blind, intervention administrator(s)

[5] Events<300

[6] Data cannot be extracted or is not reported for all outcomes and/or funding from pharmaceutical companies

[7] Unclear method of allocation concealment and non-blind intervention administrator(s)

[8] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 0.75)

[9] I-squared>80%

[10] Unclear randomisation method and method of allocation concealment, unclear blinding of intervention administrator(s), unclear blinding of (or non-blind) outcome assessment, and unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)

[11] 95% CI crosses both line of no effect and threshold for clinically important harm (SMD 0.5)

[12] Unclear randomisation method and method of allocation concealment, unclear blinding of intervention administrator(s) and outcome assessment, and unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)

[13] N<400

[14] Unclear randomisation method and method of allocation concealment and unclear blinding of intervention administration
[15] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 1.25)
[16] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

Switching to another antidepressant of the same class compared to switching to another antidepressant of a different class

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Switch to another antidepressant of the same class versus switch to another antidepressant of a different class	Control	Relative (95% CI)	Absolute		
Remission (switch to another SSRI versus switch to SNRI) (follow-up 12-14 weeks; assessed with: Number of people scoring ≤4/7 on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	75/440 (17%)	123/444 (27.7%)	RR 0.61 (0.45 to 0.83)	108 fewer per 1000 (from 47 fewer to 152 fewer)	⊕OOO VERY LOW	
								28.1%		110 fewer per 1000 (from 48 fewer to 155 fewer)		
Remission (switch to another SSRI versus switch to an atypical AD) (follow-up mean 14 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias[3]	42/238 (17.6%)	51/239 (21.3%)	RR 0.83 (0.57 to 1.19)	36 fewer per 1000 (from 92 fewer to 41 more)	⊕OOO VERY LOW	
								21.3%		36 fewer per 1000 (from 92 fewer to 40 more)		
Response (switch to another SSRI versus switch to SNRI) (follow-up mean 14 weeks; assessed with: Number of people showing ≥50% improvement on Quick Inventory of Depressive Symptomatology (QIDS))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	63/238 (26.5%)	70/250 (28%)		14 fewer per 1000 (from 81		

									RR 0.95 (0.71 to 1.26)	fewer to 73 more) 18 fewer per 1000 (from 106 fewer to 95 more)	⊕OOO VERY LOW	
								36.5%				
Response (switch to another SSRI versus switch to an atypical AD) (follow-up mean 14 weeks; assessed with: Number of people showing ≥50% improvement on Quick Inventory of Depressive Symptomatology (QIDS))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[6]	reporting bias[3]	63/238 (26.5%)	62/239 (25.9%)	RR 1.02 (0.76 to 1.38)	5 more per 1000 (from 62 fewer to 99 more) 5 more per 1000 (from 62 fewer to 98 more)	⊕OOO VERY LOW	
								25.9%				
Depression symptomatology (switch to another SSRI versus switch to SNRI) (follow-up mean 14 weeks; measured with: Quick Inventory of Depressive Symptomatology (QIDS; change score); Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[3]	238	250	-	SMD 0.08 lower (0.26 lower to 0.09 higher)	⊕⊕OO LOW	
Depression symptomatology (switch to another SSRI versus switch to an atypical AD) (follow-up mean 14 weeks; measured with: Quick Inventory of Depressive Symptomatology (QIDS; change score); Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[3]	238	239	-	SMD 0.12 lower (0.3 lower to 0.06 higher)	⊕⊕OO LOW	
Discontinuation for any reason (switch to another SSRI versus switch to SNRI) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[7]	reporting bias[3]	43/206 (20.9%)	49/200 (24.5%)	RR 0.85 (0.59 to 1.22)	37 fewer per 1000 (from 100 fewer to 54 more) 37 fewer per 1000 (from 100 fewer to 54 more)	⊕OOO VERY LOW	
								24.5%				

Discontinuation due to adverse events (switch to another SSRI versus switch to SNRI) (follow-up 12-14 weeks; assessed with: Number of participants discontinuing due to adverse events)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[8]	reporting bias[3]	61/443 (13.8%)	64/448 (14.3%)	RR 0.99 (0.72 to 1.35)	1 fewer per 1000 (from 40 fewer to 50 more) 1 fewer per 1000 (from 38 fewer to 47 more)	⊕OOO VERY LOW	
								13.4%				
Discontinuation due to adverse events (switch to another SSRI versus switch to an atypical AD) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[7]	reporting bias[3]	50/238 (21%)	65/239 (27.2%)	RR 0.77 (0.56 to 1.07)	63 fewer per 1000 (from 120 fewer to 19 more) 63 fewer per 1000 (from 120 fewer to 19 more)	⊕OOO VERY LOW	
								27.2%				

[1] Unclear randomisation method and method of allocation concealment, and unclear blinding of intervention administrator(s)
[2] Events<300
[3] Data cannot be extracted or is not reported for all outcomes and/or funding from pharmaceutical companies
[4] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 0.75)
[5] 95% CI crosses line of no effect and threshold for both clinically important harm (RR 0.75) and for clinically important benefit (RR 1.25)
[6] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[7] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 0.75)
[8] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

Switching to another antidepressant or non-antidepressant agent (head-to-head comparisons)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Switch to another antidepressant/non-antidepressant agent (head-to-head)	Control	Relative (95% CI)	Absolute		

Remission (switch to SSRI versus switch to non-SSRI AD) (follow-up 4-14 weeks; assessed with: Number of people scoring ≤4/≤7/<10 on Hamilton Rating Scale for Depression (HAM-D))												
4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[2]	102/587 (17.4%)	217/810 (26.8%)	RR 0.62 (0.5 to 0.77)	102 fewer per 1000 (from 62 fewer to 134 fewer)	⊕⊕OO LOW	
								31.4%		119 fewer per 1000 (from 72 fewer to 157 fewer)		
Remission (switch to SSRI versus switch to antipsychotic) (follow-up 8-12 weeks; assessed with: Number of people scoring ≤8 on Montgomery Asberg Depression Rating Scale (MADRS))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias[2]	29/198 (14.6%)	27/203 (13.3%)	RR 1.1 (0.68 to 1.8)	13 more per 1000 (from 43 fewer to 106 more)	⊕OOO VERY LOW	
								13.4%		13 more per 1000 (from 43 fewer to 107 more)		
Remission (switch to SNRI versus switch to atypical antidepressant) (follow-up 8-14 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[2]	83/300 (27.7%)	71/294 (24.1%)	RR 1.16 (0.89 to 1.52)	39 more per 1000 (from 27 fewer to 126 more)	⊕OOO VERY LOW	
								28.9%		46 more per 1000 (from 32 fewer to 150 more)		
Remission (switch to SSRI + antipsychotic versus switch to antipsychotic-only) (follow-up 8-12 weeks; assessed with: Number of people scoring ≤8 on Montgomery Asberg Depression Rating Scale (MADRS))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[2]	94/376 (25%)	27/203 (13.3%)	RR 1.63 (0.97 to 2.76)	84 more per 1000 (from 4 fewer to 234 more)	⊕OOO VERY LOW	
								13.4%		84 more per 1000 (from 4 fewer to 236 more)		
Remission (switch to SSRI + antipsychotic versus switch to SSRI-only) (follow-up 8-12 weeks; assessed with: Number of people scoring ≤8 on Montgomery Asberg Depression Rating Scale (MADRS))												

2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[2]	94/376 (25%)	29/198 (14.6%)	RR 1.45 (0.97 to 2.17)	66 more per 1000 (from 4 fewer to 171 more)	⊕OOO VERY LOW	
								15.6%		70 more per 1000 (from 5 fewer to 183 more)		
Response (switch to SSRI versus switch to non-SSRI AD) (follow-up 4-14 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D)/Quick Inventory of Depressive Symptomatology (QIDS)/≥50% improvement on HAM-D AND much/very much improved on CGI-I (score 1-2))												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[6]	reporting bias[2]	127/385 (33%)	196/616 (31.8%)	RR 0.91 (0.74 to 1.12)	29 fewer per 1000 (from 83 fewer to 38 more)	⊕OOO VERY LOW	
								45%		40 fewer per 1000 (from 117 fewer to 54 more)		
Response (switch to SSRI versus switch to antipsychotic) (follow-up 8-12 weeks; assessed with: Number of people showing ≥50% improvement on Montgomery Asberg Depression Rating Scale (MADRS))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[7]	reporting bias[2]	60/198 (30.3%)	43/203 (21.2%)	RR 1.43 (1.02 to 2.01)	91 more per 1000 (from 4 more to 214 more)	⊕OOO VERY LOW	
								22.4%		96 more per 1000 (from 4 more to 226 more)		
Response (switch to SNRI versus switch to atypical antidepressant) (follow-up 8-14 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D)/Quick Inventory of Depressive Symptomatology (QIDS))												
2	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[2]	102/300 (34%)	94/294 (32%)	RR 1.09 (0.88 to 1.35)	29 more per 1000 (from 38 fewer to 112 more)	⊕OOO VERY LOW	
								42.1%		38 more per 1000 (from 51 fewer to 147 more)		
Response (switch to SSRI + antipsychotic versus switch to antipsychotic-only) (follow-up 8-12 weeks; assessed with: Number of people showing ≥50% improvement on Montgomery Asberg Depression Rating Scale (MADRS))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[7]	reporting bias[2]	140/376 (37.2%)	43/203 (21.2%)	RR 1.54 (1.13 to 2.1)	114 more per 1000 (from 28 more to 233 more)	⊕OOO VERY LOW	

								22.4%		121 more per 1000 (from 29 more to 246 more)		
Response (switch to SSRI + antipsychotic versus switch to SSRI-only) (follow-up 8-12 weeks; assessed with: Number of people showing ≥50% improvement on Montgomery Asberg Depression Rating Scale (MADRS))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[2]	140/376 (37.2%)	60/198 (30.3%)	RR 1.09 (0.82 to 1.47)	27 more per 1000 (from 55 fewer to 142 more)	⊕OOO VERY LOW	
								31.4%		28 more per 1000 (from 57 fewer to 148 more)		
Response (switch to SSRI versus switch to SNRI) (follow-up mean 4 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[2]	36/55 (65.5%)	33/52 (63.5%)	RR 1.03 (0.78 to 1.37)	19 more per 1000 (from 140 fewer to 235 more)	⊕OOO VERY LOW	
								63.5%		19 more per 1000 (from 140 fewer to 235 more)		
Depression symptomatology (switch to SSRI versus switch to non-SSRI AD) (follow-up 4-14 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Quick Inventory of Depressive Symptomatology (QIDS; change score); Better indicated by lower values)												
3	randomised trials	serious[1]	serious[8]	no serious indirectness	no serious imprecision	reporting bias[2]	378	608	-	SMD 0.08 higher (0.18 lower to 0.34 higher)	⊕OOO VERY LOW	
Depression symptomatology (switch to SSRI versus switch to antipsychotic) (follow-up 8-12 weeks; measured with: Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[2]	202	206	-	SMD 0.27 lower (0.51 to 0.04 lower)	⊕⊕OO LOW	
Depression symptomatology (switch to SSRI + antipsychotic versus switch to antipsychotic-only) (follow-up 8-12 weeks; measured with: Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												

2	randomised trials	serious[1]	very serious[9]	no serious indirectness	serious[10]	reporting bias[2]	389	206	-	SMD 0.44 lower (0.91 lower to 0.03 higher)	⊕OOO VERY LOW	
Depression symptomatology (switch to SSRI + antipsychotic versus switch to SSRI-only) (follow-up 8-12 weeks; measured with: Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[2]	389	202	-	SMD 0.13 lower (0.35 lower to 0.1 higher)	⊕⊕OO LOW	
Discontinuation for any reason (switch to SSRI versus switch to non-SSRI AD) (follow-up 4-12 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
3	randomised trials	serious[11]	no serious inconsistency	no serious indirectness	serious[12]	reporting bias[2]	70/373 (18.8%)	75/345 (21.7%)	RR 0.86 (0.65 to 1.16)	30 fewer per 1000 (from 76 fewer to 35 more)	⊕OOO VERY LOW	
								20.2%		28 fewer per 1000 (from 71 fewer to 32 more)		
Discontinuation for any reason (switch to SSRI versus switch to antipsychotic) (follow-up 8-12 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[11]	no serious inconsistency	no serious indirectness	serious[12]	reporting bias[2]	40/202 (19.8%)	50/206 (24.3%)	RR 0.82 (0.56 to 1.18)	44 fewer per 1000 (from 107 fewer to 44 more)	⊕OOO VERY LOW	
								25.6%		46 fewer per 1000 (from 113 fewer to 46 more)		
Discontinuation for any reason (switch to SNRI versus switch to atypical antidepressant) (follow-up mean 8 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
1	randomised trials	very serious[13]	no serious inconsistency	no serious indirectness	very serious[14]	reporting bias[2]	9/50 (18%)	10/55 (18.2%)	RR 0.99 (0.44 to 2.24)	2 fewer per 1000 (from 102 fewer to 225 more)	⊕OOO VERY LOW	
								18.2%		2 fewer per 1000 (from 102 fewer to 226 more)		
Discontinuation for any reason (switch to SSRI + antipsychotic versus switch to antipsychotic-only) (follow-up 8-12 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												

2	randomised trials	serious[11]	no serious inconsistency	no serious indirectness	serious[12]	reporting bias[2]	90/389 (23.1%)	50/206 (24.3%)	RR 0.89 (0.65 to 1.21)	27 fewer per 1000 (from 85 fewer to 51 more)	⊕OOO VERY LOW	
								25.6%		28 fewer per 1000 (from 90 fewer to 54 more)		
Discontinuation for any reason (switch to SSRI + antipsychotic versus switch to SSRI-only) (follow-up 8-12 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
2	randomised trials	serious[11]	no serious inconsistency	no serious indirectness	serious[15]	reporting bias[2]	90/389 (23.1%)	40/202 (19.8%)	RR 1.12 (0.78 to 1.59)	24 more per 1000 (from 44 fewer to 117 more)	⊕OOO VERY LOW	
								19.9%		24 more per 1000 (from 44 fewer to 117 more)		
Discontinuation due to adverse events (switch to SSRI versus switch to non-SSRI AD) (follow-up 4-14 weeks; assessed with: Number of participants discontinuing due to adverse events)												
3	randomised trials	serious[11]	no serious inconsistency	no serious indirectness	serious[14]	reporting bias[2]	64/505 (12.7%)	134/748 (17.9%)	RR 0.87 (0.66 to 1.14)	23 fewer per 1000 (from 61 fewer to 25 more)	⊕OOO VERY LOW	
								8.2%		11 fewer per 1000 (from 28 fewer to 11 more)		
Discontinuation due to adverse events (switch to SSRI versus switch to antipsychotic) (follow-up 8-12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
2	randomised trials	serious[11]	no serious inconsistency	no serious indirectness	serious[7]	reporting bias[2]	7/202 (3.5%)	19/206 (9.2%)	RR 0.39 (0.16 to 0.91)	56 fewer per 1000 (from 8 fewer to 77 fewer)	⊕OOO VERY LOW	
								8.9%		54 fewer per 1000 (from 8 fewer to 75 fewer)		
Discontinuation due to adverse events (switch to SNRI versus switch to atypical antidepressant) (follow-up 8-14 weeks; assessed with: Number of participants discontinuing due to adverse events)												

2	randomised trials	serious[11]	no serious inconsistency	no serious indirectness	serious[14]	reporting bias[2]	53/300 (17.7%)	65/289 (22.5%)	RR 0.78 (0.57 to 1.07)	49 fewer per 1000 (from 97 fewer to 16 more)	⊕OOO VERY LOW	
								13.6%		30 fewer per 1000 (from 58 fewer to 10 more)		
Discontinuation due to adverse events (switch to SSRI + antipsychotic versus switch to antipsychotic-only) (follow-up 8-12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
2	randomised trials	serious[11]	no serious inconsistency	no serious indirectness	very serious[14]	reporting bias[2]	39/389 (10%)	19/206 (9.2%)	RR 0.98 (0.48 to 2.03)	2 fewer per 1000 (from 48 fewer to 95 more)	⊕OOO VERY LOW	
								8.9%		2 fewer per 1000 (from 46 fewer to 92 more)		
Discontinuation due to adverse events (switch to SSRI + antipsychotic versus switch to SSRI-only) (follow-up 8-12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
2	randomised trials	serious[11]	no serious inconsistency	no serious indirectness	serious[7]	reporting bias[2]	39/389 (10%)	7/202 (3.5%)	RR 2.41 (1.07 to 5.42)	49 more per 1000 (from 2 more to 153 more)	⊕OOO VERY LOW	
								3.9%		55 more per 1000 (from 3 more to 172 more)		

[1] Unclear randomisation method and method of allocation concealment, and unclear blinding of intervention administration and outcome assessment
[2] Data cannot be extracted or is not reported for all outcomes and/or funding from pharmaceutical companies
[3] 95% CI crosses line of no effect and threshold for both clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)
[4] Unclear (or high risk) randomisation method and unclear method of allocation concealment, and unclear blinding of intervention administrator(s)
[5] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[6] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 0.75)
[7] Events<300
[8] I-squared>50%
[9] I-squared>80%
[10] 95% CI crosses both line of no effect and threshold for clinically important benefit (SMD -0.5)
[11] Unclear randomisation method and method of allocation concealment and unclear blinding of intervention administrator(s)
[12] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 0.75)
[13] High risk of bias associated with randomisation method due to significant difference between groups at baseline and unclear blinding of intervention administrator(s)
[14] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)
[15] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 1.25)

Switching to a combined psychological and pharmacological intervention versus switching to a psychological intervention-only

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Switch to combined psych and pharm intervention versus switch to psych intervention-only	Control	Relative (95% CI)	Absolute		
Discontinuation for any reason (CBT individual [under 15 sessions] + antipsychotic versus CBT individual [under 15 sessions]-only) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing for any reason (including adverse events))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	1/11 (9.1%)	6/11 (54.5%)	RR 0.17 (0.02 to 1.17)	453 fewer per 1000 (from 535 fewer to 93 more)	⊕OOO VERY LOW	
								54.6%		453 fewer per 1000 (from 535 fewer to 93 more)		

[1] Unclear randomisation method and unclear blinding of intervention administrator(s)
[2] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 0.75)
[3] Efficacy data cannot be extracted and study funded by pharmaceutical company

Chronic depression (chapter 9)

Problem solving versus pill placebo for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Problem solving	Pill placebo	Relative (95% CI)	Absolute		
Remission (follow-up mean 11 weeks; assessed with: Number of people scoring <7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	32/63 (50.8%)	25/62 (40.3%)	RR 1.26 (0.85 to 1.86)	105 more per 1000 (from 60 fewer to 347 more)		

								40.3%		105 more per 1000 (from 60 fewer to 347 more)	⊕OOO VERY LOW	

[1] Intervention administrators and participants not blinded, although outcome assessment is blinded
[2] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[3] Medication and placebo supplied by pharmaceutical company and authors have some financial interests in pharmaceutical companies

Cognitive and cognitive behavioural therapies versus antidepressants for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Cognitive and cognitive behavioural therapies	Antidepressants	Relative (95% CI)	Absolute		
Remission (any cognitive or cognitive behavioural therapy versus any AD) (follow-up 8-12 weeks; assessed with: Number of people scoring <7/≤8 on Hamilton Rating Scale for Depression (HAM-D)/ ≤9 on Montgomery Asberg Depression Rating Scale (MADRS))												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	105/308 (34.1%)	95/307 (30.9%)	RR 1.1 (0.83 to 1.46)	31 more per 1000 (from 53 fewer to 142 more)	⊕OOO VERY LOW	
								29.1%		29 more per 1000 (from 49 fewer to 134 more)		
Remission (CBASP versus nefazodone) (follow-up mean 12 weeks; assessed with: Number of people scoring ≤8 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	72/216 (33.3%)	64/220 (29.1%)	RR 1.15 (0.87 to 1.52)	44 more per 1000 (from 38 fewer to 151 more)	⊕OOO VERY LOW	
								29.1%		44 more per 1000 (from 38 fewer to 151 more)		
Remission (CBASP versus escitalopram) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤9 on Montgomery Asberg Depression Rating Scale (MADRS))												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[6]	1/29 (3.4%)	5/30 (16.7%)	RR 0.21 (0.03 to 1.67)	132 fewer per 1000 (from 162 fewer to 112 more)		

								16.7%		132 fewer per 1000 (from 162 fewer to 112 more)	⊕OOO VERY LOW	
Remission (problem solving versus paroxetine) (follow-up mean 11 weeks; assessed with: Number of people scoring <7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[6]	32/63 (50.8%)	26/57 (45.6%)	RR 1.11 (0.77 to 1.62)	50 more per 1000 (from 105 fewer to 283 more)	⊕OOO VERY LOW	
								45.6%		50 more per 1000 (from 105 fewer to 283 more)		
Response (any cognitive or cognitive behavioural therapy versus any AD) (follow-up 8-12 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D) AND HAMD score 8-15)/≥50% improvement on Montgomery Asberg Depression Rating Scale (MADRS))												
2	randomised trials	serious[1]	serious[8]	no serious indirectness	very serious[5]	reporting bias[3]	33/245 (13.5%)	49/250 (19.6%)	RR 0.56 (0.21 to 1.49)	86 fewer per 1000 (from 155 fewer to 96 more)	⊕OOO VERY LOW	
								22.7%		100 fewer per 1000 (from 179 fewer to 111 more)		
Response (CBASP versus nefazodone) (follow-up mean 12 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D) AND HAMD score 8-15)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[9]	reporting bias[3]	31/216 (14.4%)	41/220 (18.6%)	RR 0.77 (0.5 to 1.18)	43 fewer per 1000 (from 93 fewer to 34 more)	⊕OOO VERY LOW	
								18.6%		43 fewer per 1000 (from 93 fewer to 33 more)		
Response (CBASP versus escitalopram) (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Montgomery Asberg Depression Rating Scale (MADRS))												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[9]	reporting bias[6]	2/29 (6.9%)	8/30 (26.7%)	RR 0.26 (0.06 to 1.12)	197 fewer per 1000 (from 251 fewer to 32 more)	⊕OOO VERY LOW	
								26.7%		198 fewer per 1000 (from 251 fewer to 32 more)		

Depression symptomatology (any cognitive or cognitive behavioural therapy versus any AD) (follow-up 12-16 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
2	randomised trials	serious[1]	very serious[10]	no serious indirectness	very serious[11]	reporting bias[3]	226	232	-	SMD 0.61 higher (0.54 lower to 1.76 higher)	⊕OOO VERY LOW	
Depression symptomatology (CBASP versus nefazodone) (follow-up mean 12 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[3]	216	220	-	SMD 0.11 higher (0.08 lower to 0.3 higher)	⊕⊕OO LOW	
Depression symptomatology (CBT versus fluoxetine) (follow-up mean 16 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[12]	no serious inconsistency	no serious indirectness	serious[13]	none	10	12	-	SMD 1.3 higher (0.36 to 2.24 higher)	⊕⊕OO LOW	
Discontinuation for any reason (any cognitive or cognitive behavioural therapy versus any AD) (follow-up 8-16 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
3	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	very serious[14]	reporting bias[3]	63/275 (22.9%)	67/270 (24.8%)	RR 0.92 (0.68 to 1.25)	20 fewer per 1000 (from 79 fewer to 62 more)	⊕OOO VERY LOW	
										18 fewer per 1000 (from 74 fewer to 58 more)		
								23.1%				
Discontinuation for any reason (CBASP versus nefazodone) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	very serious[14]	reporting bias[3]	55/228 (24.1%)	59/226 (26.1%)	RR 0.92 (0.67 to 1.27)	21 fewer per 1000 (from 86 fewer to 70 more)	⊕OOO VERY LOW	
										21 fewer per 1000 (from 86 fewer to 70 more)		
								26.1%				
Discontinuation for any reason (CBASP versus escitalopram) (follow-up mean 8 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												

1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[14]	reporting bias[6]	2/29 (6.9%)	5/31 (16.1%)	RR 0.43 (0.09 to 2.03)	92 fewer per 1000 (from 147 fewer to 166 more)	⊕OOO VERY LOW	
								16.1%		92 fewer per 1000 (from 147 fewer to 166 more)		
Discontinuation for any reason (CBT versus fluoxetine) (follow-up mean 16 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	serious[15]	no serious inconsistency	no serious indirectness	very serious[14]	none	6/18 (33.3%)	3/13 (23.1%)	RR 1.44 (0.44 to 4.74)	102 more per 1000 (from 129 fewer to 863 more)	⊕OOO VERY LOW	
								23.1%		102 more per 1000 (from 129 fewer to 864 more)		
Discontinuation due to adverse events (CBASP versus nefazodone) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[16]	reporting bias[3]	3/228 (1.3%)	31/226 (13.7%)	RR 0.1 (0.03 to 0.31)	123 fewer per 1000 (from 95 fewer to 133 fewer)	⊕OOO VERY LOW	
								13.7%		123 fewer per 1000 (from 95 fewer to 133 fewer)		

[1] Non-blind intervention administrator(s) and participants, although the outcome assessor was blinded. Unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)
[2] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[3] Funding from pharmaceutical company
[4] Unclear method of allocation concealment and non-blind intervention administrator(s) and participants, although the outcome assessor was blinded
[5] 95% CI crosses line of no effect and both threshold for clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)
[6] Data cannot be extracted or is not reported for all outcomes and funding from pharmaceutical company
[7] Non-blind intervention administrator(s) and participants, although outcome assessors are blinded
[8] I-squared=>50%
[9] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 0.75)
[10] I-squared>80%
[11] 95% CI crosses line of no effect and threshold for both clinically important benefit (SMD -0.5) and clinically important harm (SMD 0.5)
[12] Unclear randomisation method and method of allocation concealment, and non-blind intervention administrator(s) and participants (although outcome assessors are blinded). Unclear risk of attrition bias (drop-out>20% and completer analysis used but difference between groups<20%)
[13] N<400
[14] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

[15] Unclear randomisation method and method of allocation concealment, and non-blind intervention administrator(s) and participants (although outcome assessors are blinded)

[16] Events<300

CBASP versus other psychological intervention for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	CBASP	Other psych intervention	Relative (95% CI)	Absolute		
Remission (CBASP versus other psych intervention) (follow-up 16-20 weeks; assessed with: Number of people scoring ≤8 on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	35/138 (25.4%)	17/126 (13.5%)	RR 1.93 (1.14 to 3.26)	125 more per 1000 (from 19 more to 305 more)	⊕OOO VERY LOW	
								16.3%		152 more per 1000 (from 23 more to 368 more)		
Remission (CBASP versus IPT) (follow-up mean 16 weeks; assessed with: Number of people scoring ≤8 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[4]	none	8/14 (57.1%)	3/15 (20%)	RR 2.86 (0.94 to 8.66)	372 more per 1000 (from 12 fewer to 1000 more)	⊕⊕OO LOW	
								20%		372 more per 1000 (from 12 fewer to 1000 more)		
Remission (CBASP versus supportive psychotherapy) (follow-up mean 20 weeks; assessed with: Number of people scoring ≤8 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	27/124 (21.8%)	14/111 (12.6%)	RR 1.73 (0.95 to 3.12)	92 more per 1000 (from 6 fewer to 267 more)	⊕OOO VERY LOW	
								12.6%		92 more per 1000 (from 6 fewer to 267 more)		
Response (CBASP versus other psych intervention) (follow-up 16-20 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	57/138 (41.3%)	31/126 (24.6%)	RR 1.7 (1.18 to 2.44)	172 more per 1000 (from 44 more to 354 more)	⊕OOO VERY LOW	
								25.5%		179 more per 1000 (from 46 more to 367 more)		
Response (CBASP versus IPT) (follow-up mean 16 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D) AND HAMD score≤15))												

1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[4]	none	9/14 (64.3%)	4/15 (26.7%)	RR 2.41 (0.96 to 6.08)	376 more per 1000 (from 11 fewer to 1000 more)	⊕⊕OO LOW	
								26.7%		376 more per 1000 (from 11 fewer to 1000 more)		
Response (CBASP versus supportive psychotherapy) (follow-up mean 20 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	48/124 (38.7%)	27/111 (24.3%)	RR 1.59 (1.07 to 2.36)	144 more per 1000 (from 17 more to 331 more)	⊕OOO VERY LOW	
								24.3%		143 more per 1000 (from 17 more to 330 more)		
Depression symptomatology (CBASP versus other psych intervention) (follow-up 16-20 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	151	146	-	SMD 0.49 lower (0.98 lower to 0 higher)	⊕OOO VERY LOW	
Depression symptomatology (CBASP versus IPT) (follow-up mean 16 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[5]	none	14	15	-	SMD 0.89 lower (1.66 to 0.12 lower)	⊕⊕OO LOW	
Depression symptomatology (CBASP versus supportive psychotherapy) (follow-up mean 20 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	137	131	-	SMD 0.33 lower (0.58 to 0.09 lower)	⊕OOO VERY LOW	
Discontinuation for any reason (CBASP versus other psych intervention) (follow-up 16-20 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[6]	none	16/152 (10.5%)	24/146 (16.4%)	RR 0.64 (0.35 to 1.16)	59 fewer per 1000 (from 107 fewer to 26 more)	⊕OOO VERY LOW	
								15.1%		54 fewer per 1000 (from 98 fewer to 24 more)		
Discontinuation for any reason (CBASP versus IPT) (follow-up mean 16 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												

1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[7]	none	2/15 (13.3%)	2/15 (13.3%)	RR 1 (0.16 to 6.2)	0 fewer per 1000 (from 112 fewer to 693 more) 0 fewer per 1000 (from 112 fewer to 692 more)	⊕OOO VERY LOW	
								13.3%				
Discontinuation for any reason (CBASP versus supportive psychotherapy) (follow-up mean 20 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[6]	none	14/137 (10.2%)	22/131 (16.8%)	RR 0.61 (0.33 to 1.14)	65 fewer per 1000 (from 113 fewer to 24 more) 66 fewer per 1000 (from 113 fewer to 24 more)	⊕OOO VERY LOW	
								16.8%				

[1] High risk of bias associated with randomisation method due to significant difference between groups at baseline. Non-blind intervention administrator(s) and participants, although outcome assessors are blinded

[2] Events<300

[3] Non-blind intervention administrator(s) and participants, although outcome assessors are blinded

[4] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)

[5] N<400

[6] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 0.75)

[7] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

Cognitive and cognitive behavioural therapies + TAU/AD versus TAU/AD-only for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Cognitive and cognitive behavioural therapies + TAU/AD**	**TAU/AD-only**	**Relative (95% CI)**	**Absolute**		
Remission (MBCT+TAU versus TAU) (follow-up mean 8 weeks; assessed with: Number of people scoring ≤13 on Beck Depression Inventory II (BDI-II) AND ≥50% improvement on BDI-II/≤7 on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	12/52 (23.1%)	3/50 (6%)	RR 3.72 (1.1 to 12.54)	163 more per 1000 (from 6 more to 692 more) 169 more per 1000 (from 6 more to 715 more)	⊕OOO VERY LOW	
								6.2%				

Remission (CBASP + TAU/nefazodone versus TAU/nefazodone) (follow-up 8-52 weeks; assessed with: Number of people scoring ≤7/8 on Hamilton Rating Scale for Depression (HAM-D)/≤13 on Inventory of Depressive Symptoms (IDS))												
3	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[4]	131/328 (39.9%)	74/326 (22.7%) 11.3%	RR 1.71 (1.35 to 2.15)	161 more per 1000 (from 79 more to 261 more) 80 more per 1000 (from 40 more to 130 more)	⊕OOO VERY LOW	
Response (CBASP + TAU/nefazodone versus TAU/nefazodone) (follow-up 12-52 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D)/Inventory of Depressive Symptoms (IDS))												
2	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[4]	77/293 (26.3%)	57/292 (19.5%) 20.4%	RR 1.35 (1 to 1.83)	68 more per 1000 (from 0 more to 162 more) 71 more per 1000 (from 0 more to 169 more)	⊕OOO VERY LOW	
Depression symptomatology (MBCT+TAU versus TAU) (follow-up 8-12 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Beck Depression Inventory (BDI-II; change score); Better indicated by lower values)												
3	randomised trials	very serious[5]	very serious[6]	no serious indirectness	serious[7]	none	56	61	-	SMD 1.14 lower (2.1 to 0.19 lower)	⊕OOO VERY LOW	
Depression symptomatology (CBASP + TAU/nefazodone versus TAU/nefazodone) (follow-up 8-52 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Inventory of Depressive Symptoms (IDS; change score); Better indicated by lower values)												
3	randomised trials	very serious[8]	serious[9]	no serious indirectness	no serious imprecision	none	305	305	-	SMD 0.8 lower (1.13 to 0.47 lower)	⊕OOO VERY LOW	
Depression symptomatology (CBT [group] + TAU versus waitlist + TAU) (follow-up mean 10 weeks; measured with: Beck Depression Inventory (BDI; change score); Better indicated by lower values)												
1	randomised trials	very serious[10]	no serious inconsistency	no serious indirectness	serious[7]	none	48	40	-	SMD 0.85 lower (1.29 to 0.41 lower)	⊕OOO VERY LOW	
Discontinuation for any reason (MBCT+TAU versus TAU) (follow-up 8-12 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												

3	randomised trials	serious[11]	no serious inconsistency	no serious indirectness	serious[12]	none	15/66 (22.7%)	4/64 (6.3%)	RR 2.85 (0.84 to 9.66)	116 more per 1000 (from 10 fewer to 541 more) 124 more per 1000 (from 11 fewer to 580 more)	⊕⊕OO LOW	
								6.7%				
Discontinuation for any reason (CBASP + TAU/nefazodone versus TAU/nefazodone) (follow-up 8-52 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
3	randomised trials	very serious[13]	serious[9]	no serious indirectness	very serious[14]	none	74/329 (22.5%)	79/333 (23.7%)	RR 1.09 (0.54 to 2.17)	21 more per 1000 (from 109 fewer to 278 more) 23 more per 1000 (from 120 fewer to 305 more)	⊕OOO VERY LOW	
								26.1%				
Discontinuation for any reason (CBT [group] + TAU versus waitlist + TAU) (follow-up mean 10 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	serious[15]	no serious inconsistency	no serious indirectness	serious[2]	none	0/48 (0%)	8/48 (16.7%)	RR 0.06 (0 to 0.99)	157 fewer per 1000 (from 2 fewer to 167 fewer) 157 fewer per 1000 (from 2 fewer to 167 fewer)	⊕⊕OO LOW	
								16.7%				
Discontinuation due to adverse events (CBASP + nefazodone versus nefazodone) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[16]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[4]	16/227 (7%)	31/226 (13.7%)	RR 0.51 (0.29 to 0.91)	67 fewer per 1000 (from 12 fewer to 97 fewer) 67 fewer per 1000 (from 12 fewer to 97 fewer)	⊕OOO VERY LOW	
								13.7%				

[1] High risk of bias associated with randomisation method due to significant difference between groups at baseline. Non-blind intervention administrator(s) and participants, although outcome assessors are blind. Unclear risk of attrition bias (>20% difference in drop-out between groups but ITT analysis used)

[2] Events<300

[3] Non-blind intervention administrator(s) or participants, although outcome assessors are blinded. Unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)

[4] Funding from pharmaceutical company

[5] Non-blind intervention administrator(s) and participants and outcome assessment either non-blind or blinding unclear

[6] I-squared=>80%
[7] N<400
[8] High risk of bias associated with randomisation method due to significant difference between groups at baseline in studies contributing >50% to analysis. Non-blind intervention administrator(s) and participants, although outcome assessors are blind. Unclear risk of attrition bias (drop-out>20% or difference between groups>20%)
[9] I-squared>50%
[10] Unclear randomisation method and method of allocation concealment. Non-blind intervention administration and outcome assessment
[11] Unclear (or high risk associated with) randomisation method, and non-blind intervention administrator(s) and participants
[12] 95% CI crosses both the line of no effect and the threshold for clinically important harm (RR 1.25)
[13] High risk of bias associated with randomisation method due to significant difference between groups at baseline in studies contributing >50% to analysis. Non-blind intervention administrator(s) and participants
[14] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)
[15] Unclear randomisation method and method of allocation concealment and non-blind intervention administrator(s) and participants
[16] Non-blind intervention administrator(s) and participants

CBASP (maintenance treatment) versus assessment-only for relapse prevention in chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	CBASP (maintenance treatment)	Assessment-only	Relative (95% CI)	Absolute		
Relapse (follow-up mean 52 weeks; assessed with: Number of people scoring ≥16 on Hamilton Rating Scale for Depression (HAM-D) on 2 consecutive visits AND meeting DSM-IV criteria for a diagnosis of MDD)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	1/42 (2.4%)	8/40 (20%)	RR 0.12 (0.02 to 0.91)	176 fewer per 1000 (from 18 fewer to 196 fewer)	⊕OOO VERY LOW	
								20%		176 fewer per 1000 (from 18 fewer to 196 fewer)		
Depression symptomatology (follow-up mean 52 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias[3]	42	40	-	SMD 0.91 lower (1.37 to 0.45 lower)	⊕OOO VERY LOW	
Discontinuation for any reason (follow-up mean 52 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												

1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias[3]	10/42 (23.8%)	11/40 (27.5%)	RR 0.87 (0.41 to 1.81)	36 fewer per 1000 (from 162 fewer to 223 more)	⊕OOO VERY LOW	
								27.5%		36 fewer per 1000 (from 162 fewer to 223 more)		

[1] High risk of bias associated with randomisation method due to significant difference between groups at baseline, and method of allocation concealment is unclear. Non-blind intervention administrator(s) and participants, although outcome assessors are blinded. Unclear risk of attrition bias (drop-put>20% but difference between groups <20% and ITT analysis used)
[2] Events<300
[3] Funding from pharmaceutical company
[4] N<400
[5] High risk of bias associated with randomisation method due to significant difference between groups at baseline, and method of allocation concealment is unclear. Non-blind intervention administrator(s) and participants
[6] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

IPT versus sertraline for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**IPT**	**Sertraline**	**Relative (95% CI)**	**Absolute**		
Remission (follow-up mean 16 weeks; assessed with: Number of people scoring <7 on Hamilton Rating Scale for Depression (HAM-D) AND >50% improvement on HAMD AND GAF score>70)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	5/23 (21.7%)	10/24 (41.7%)	RR 0.52 (0.21 to 1.29)	200 fewer per 1000 (from 329 fewer to 121 more)	⊕OOO VERY LOW	
								41.7%		200 fewer per 1000 (from 329 fewer to 121 more)		
Response (follow-up 16-26 weeks; assessed with: Number of people showing ≥40% improvement on Montgomery Asberg Depression Rating Scale (MADRS)/≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	91/201 (45.3%)	131/220 (59.5%)	RR 0.76 (0.63 to 0.92)	143 fewer per 1000 (from 48 fewer to 220 fewer)	⊕OOO VERY LOW	
								59%		142 fewer per 1000 (from 47 fewer to 218 fewer)		

Depression symptomatology (follow-up 16-26 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
2	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[6]	201	220	-	SMD 0.49 higher (0.24 to 0.74 higher)	⊕OOO VERY LOW	
Discontinuation for any reason (follow-up mean 16 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[7]	reporting bias[3]	4/23 (17.4%)	5/24 (20.8%)	RR 0.83 (0.26 to 2.73)	35 fewer per 1000 (from 154 fewer to 360 more)	⊕OOO VERY LOW	
								20.8%		35 fewer per 1000 (from 154 fewer to 360 more)		

[1] Unclear randomisation and method of allocation concealment. Non-blind intervention administrator(s) and participants, although outcome assessors are blinded
[2] 95% CI crosses line of no effect and threshold for both clinically important harm (Rr 0.75) and clinically important benefit (RR 1.25)
[3] Study partially funded by pharmaceutical company
[4] High risk of bias associated with randomisation bias due to significant difference between groups at baseline. Non-blind intervention administrator(s) and participants, although outcome assessors are blinded
[5] Events<300
[6] N<400
[7] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

IPT versus brief supportive psychotherapy (BSP) for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	IPT versus brief supportive psychotherapy (BSP) for dysthymia	Control	Relative (95% CI)	Absolute		
Remission (follow-up mean 16 weeks; assessed with: Number of people scoring <7 on Hamilton Rating Scale for Depression (HAM-D) AND >50% improvement on HAMD AND GAF score>70)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	5/23 (21.7%)	3/26 (11.5%)	RR 1.88 (0.5 to 7.03)	102 more per 1000 (from 58 fewer to 696 more)	⊕OOO VERY LOW	
								11.5%		101 more per 1000 (from 58 fewer to 693 more)		

Response (follow-up mean 16 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	8/23 (34.8%)	8/26 (30.8%)	RR 1.13 (0.51 to 2.52)	40 more per 1000 (from 151 fewer to 468 more) 40 more per 1000 (from 151 fewer to 468 more)	⊕OOO VERY LOW	
								30.8%				
Depression symptomatology (follow-up mean 16 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	reporting bias[3]	23	26	-	SMD 0.06 lower (0.63 lower to 0.5 higher)	⊕OOO VERY LOW	
Discontinuation for any reason (follow-up mean 16 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[6]	reporting bias[3]	4/23 (17.4%)	11/26 (42.3%)	RR 0.41 (0.15 to 1.11)	250 fewer per 1000 (from 360 fewer to 47 more) 250 fewer per 1000 (from 360 fewer to 47 more)	⊕OOO VERY LOW	
								42.3%				

[1] Randomisation method and method of allocation concealment unclear. Non-blind intervention administrator(s) and participants, although outcome assessors are blinded. Unclear risk of attrition bias (drop-out>20% and difference between groups>20% but ITT analysis used)

[2] 95% CI crosses line of no effect and threshold for both clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)

[3] Study partially funded by pharmaceutical company

[4] 95% CI crosses line of no effect and threshold for both clinically important benefit (SMD -0.5) and clinically important harm (SMD 0.5)

[5] Randomisation method and method of allocation concealment unclear. Non-blind intervention administrator(s) and participants, although outcome assessors are blinded.

[6] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 0.75)

IPT + TAU/AD versus TAU/AD-only for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**IPT + TAU/AD**	**TAU/AD-only**	**Relative (95% CI)**	**Absolute**		
Remission (IPT + TAU/AD versus TAU/AD-only) (follow-up 5-16 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D)/<7 on HAMD-D >50% improvement on HAMD AND GAF score>70)												

2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	23/45 (51.1%)	16/45 (35.6%)	RR 1.43 (0.88 to 2.33)	153 more per 1000 (from 43 fewer to 473 more) 151 more per 1000 (from 42 fewer to 467 more)	⊕OOO VERY LOW	
								35.1%				
Remission (IPT + standard pharmacotherapy versus standard pharmacotherapy + clinical management) (follow-up mean 5 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[2]	none	12/24 (50%)	6/21 (28.6%)	RR 1.75 (0.8 to 3.84)	214 more per 1000 (from 57 fewer to 811 more) 215 more per 1000 (from 57 fewer to 812 more)	⊕⊕OO LOW	
								28.6%				
Remission (IPT + sertraline versus sertraline) (follow-up mean 16 weeks; assessed with: Number of people scoring <7 on Hamilton Rating Scale for Depression (HAM-D) AND >50% improvement on HAMD AND GAF score>70)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	11/21 (52.4%)	10/24 (41.7%)	RR 1.26 (0.67 to 2.35)	108 more per 1000 (from 138 fewer to 562 more) 108 more per 1000 (from 138 fewer to 563 more)	⊕OOO VERY LOW	
								41.7%				
Response (IPT + TAU/AD versus TAU/AD-only) (follow-up 5-26 weeks; assessed with: Number of people showing ≥40% improvement on Montgomery Asberg Depression Rating Scale (MADRS)/≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
3	randomised trials	very serious[6]	serious[7]	no serious indirectness	serious[2]	reporting bias[3]	151/257 (58.8%)	139/241 (57.7%)	RR 1.11 (0.79 to 1.56)	63 more per 1000 (from 121 fewer to 323 more) 64 more per 1000 (from 122 fewer to 326 more)	⊕OOO VERY LOW	
								58.3%				
Response (IPT + standard pharmacotherapy versus standard pharmacotherapy + clinical management) (follow-up mean 5 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[8]	none	17/24 (70.8%)	8/21 (38.1%)	RR 1.86 (1.02 to 3.4)	328 more per 1000 (from 8 more to 914 more)	⊕⊕OO LOW	

								38.1%		328 more per 1000 (from 8 more to 914 more)		
Response (IPT + sertraline versus sertraline) (follow-up 16-26 weeks; assessed with: Number of people showing ≥40% improvement on Montgomery Asberg Depression Rating Scale (MADRS)/≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	very serious[6]	no serious inconsistency	no serious indirectness	serious[8]	reporting bias[9]	134/233 (57.5%)	131/220 (59.5%)	RR 0.97 (0.83 to 1.13)	18 fewer per 1000 (from 101 fewer to 77 more)	⊕OOO VERY LOW	
								59%		18 fewer per 1000 (from 100 fewer to 77 more)		
Depression symptomatology (IPT + TAU/AD versus TAU/AD-only) (follow-up 5-26 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
4	randomised trials	very serious[6]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[9]	268	254	-	SMD 0.16 lower (0.43 lower to 0.11 higher)	⊕OOO VERY LOW	
Depression symptomatology (IPT + standard pharmacotherapy versus standard pharmacotherapy + clinical management) (follow-up mean 5 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[10]	none	24	21	-	SMD 0.71 lower (1.32 to 0.1 lower)	⊕⊕OO LOW	
Depression symptomatology (IPT + moclobemide versus moclobemide + clinical management) (follow-up mean 12 weeks; measured with: Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
1	randomised trials	very serious[11]	no serious inconsistency	no serious indirectness	very serious[12]	none	11	13	-	SMD 0.03 lower (0.83 lower to 0.77 higher)	⊕OOO VERY LOW	
Depression symptomatology (IPT + sertraline versus sertraline) (follow-up 16-26 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
2	randomised trials	very serious[6]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[9]	233	220	-	SMD 0.06 lower (0.24 lower to 0.12 higher)	⊕OOO VERY LOW	
Discontinuation for any reason (IPT + TAU/AD versus TAU/AD-only) (follow-up 5-16 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
3	randomised trials	serious[13]	no serious inconsistency	no serious indirectness	very serious[14]	none	16/61 (26.2%)	18/64 (28.1%)	RR 0.95 (0.45 to 1.99)	14 fewer per 1000 (from 155 fewer to 278 more)		

								20.8%		10 fewer per 1000 (from 114 fewer to 206 more)	⊕OOO VERY LOW	
Discontinuation for any reason (IPT + standard pharmacotherapy versus standard pharmacotherapy + clinical management) (follow-up mean 5 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[14]	none	6/24 (25%)	2/21 (9.5%)	RR 2.62 (0.59 to 11.64)	154 more per 1000 (from 39 fewer to 1000 more)	⊕OOO VERY LOW	
								9.5%		154 more per 1000 (from 39 fewer to 1000 more)		
Discontinuation for any reason (IPT + moclobemide versus moclobemide + clinical management) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	serious[13]	no serious inconsistency	no serious indirectness	very serious[14]	none	6/16 (37.5%)	11/19 (57.9%)	RR 0.65 (0.31 to 1.36)	203 fewer per 1000 (from 399 fewer to 208 more)	⊕OOO VERY LOW	
								57.9%		203 fewer per 1000 (from 400 fewer to 208 more)		
Discontinuation for any reason (IPT + sertraline versus sertraline) (follow-up mean 16 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	very serious[6]	no serious inconsistency	no serious indirectness	very serious[14]	reporting bias[3]	4/21 (19%)	5/24 (20.8%)	RR 0.91 (0.28 to 2.97)	19 fewer per 1000 (from 150 fewer to 410 more)	⊕OOO VERY LOW	
								20.8%		19 fewer per 1000 (from 150 fewer to 410 more)		

[1] Randomisation method is unclear and unclear method of allocation concealment. Non-blind intervention administrator(s) and participants, although outcome assessors are blinded
[2] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[3] Study partially funded by pharmaceutical company
[4] Baseline group comparability is unclear and unclear method of allocation concealment. Non-blind intervention administrator(s) and participants, although outcome assessors are blinded
[5] 95% CI crosses line of no effect and threshold for both clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)
[6] High risk associated with randomisation method due to significant difference between groups at baseline. Non-blind intervention administrator(s) and participants, although outcome assessors are blinded
[7] I-squared>50%
[8] Events<300
[9] Data cannot be extracted or is not reported for all outcomes and study partially funded by pharmaceutical company
[10] N<400
[11] Unclear randomisation method and method of allocation concealment. Non-blind intervention administrator(s) and participants, although outcome assessors are blinded. High risk of attrition bias

(drop-out>20% and difference between groups>20%)

[12] 95% CI crosses line of no effect and threshold for both clinically important benefit (SMD -0.5) and clinically important harm (SMD 0.5)

[13] Unclear randomisation method and method of allocation concealment. Non-blind intervention administrator(s) and participants

[14] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

Brief supportive psychotherapy (BSP) versus sertraline for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Brief supportive psychotherapy (BSP)	Sertraline	Relative (95% CI)	Absolute		
Remission (follow-up mean 16 weeks; assessed with: Number of people scoring <7 on Hamilton Rating Scale for Depression (HAM-D) AND >50% improvement on HAMD AND GAF score>70)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	3/26 (11.5%)	10/24 (41.7%)	RR 0.28 (0.09 to 0.89)	300 fewer per 1000 (from 46 fewer to 379 fewer)	⊕OOO VERY LOW	
								41.7%		300 fewer per 1000 (from 46 fewer to 379 fewer)		
Response (follow-up mean 16 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias[3]	8/26 (30.8%)	14/24 (58.3%)	RR 0.53 (0.27 to 1.03)	274 fewer per 1000 (from 426 fewer to 17 more)	⊕OOO VERY LOW	
								58.3%		274 fewer per 1000 (from 426 fewer to 17 more)		
Depression symptomatology (follow-up mean 16 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	26	24	-	SMD 0.77 higher (0.19 to 1.34 higher)	⊕OOO VERY LOW	
Discontinuation for any reason (follow-up mean 16 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												

1	randomised trials	very serious[6]	no serious inconsistency	no serious indirectness	serious[7]	reporting bias[3]	11/26 (42.3%)	5/24 (20.8%)	RR 2.03 (0.83 to 4.99)	215 more per 1000 (from 35 fewer to 831 more)	⊕OOO VERY LOW	
										214 more per 1000 (from 35 fewer to 830 more)		
								20.8%				

[1] High risk of bias associated with randomisation method due to significant difference between groups at baseline, and method of allocation concealment is unclear. Non-blind intervention administrator(s) and participants, although outcome assessors are blinded. High risk of attrition bias (>20% drop-out and difference between groups >20%), although ITT analysis used

[2] Events<300

[3] Study partially funded by pharmaceutical company

[4] No explanation was provided

[5] N<400

[6] High risk of bias associated with randomisation method due to significant difference between groups at baseline, and method of allocation concealment is unclear. Non-blind intervention administrator(s) and participants

[7] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 1.25)

SSRIs versus placebo for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**SSRIs**	**Placebo**	**Relative (95% CI)**	**Absolute**		
Remission (any SSRI) (follow-up 11-13 weeks; assessed with: Number of people scoring <7/≤4/7/8 on Hamilton Rating Scale for Depression (HAM-D))												
5	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	137/301 (45.5%)	85/277 (30.7%)	RR 1.47 (1.15 to 1.87)	144 more per 1000 (from 46 more to 267 more)	⊕OOO VERY LOW	
								25.6%		120 more per 1000 (from 38 more to 223 more)		
Remission (sertraline) (follow-up mean 12 weeks; assessed with: Number of people scoring ≤4 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	63/134 (47%)	45/140 (32.1%)	RR 1.46 (1.08 to 1.98)	148 more per 1000 (from 26 more to 315 more)	⊕OOO VERY LOW	
								32.1%		148 more per 1000 (from 26 more to 315 more)		
Remission (fluoxetine) (follow-up mean 13 weeks; assessed with: Number of people scoring ≤7 on Hamilton Rating Scale for Depression (HAM-D))												

1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[6]	reporting bias[7]	32/72 (44.4%)	10/39 (25.6%)	RR 1.73 (0.96 to 3.14)	187 more per 1000 (from 10 fewer to 549 more)	⊕OOO VERY LOW	
								25.6%		187 more per 1000 (from 10 fewer to 548 more)		
Remission (escitalopram) (follow-up mean 12 weeks; assessed with: Number of people scoring ≤4 on Hamilton Rating Scale for Depression (HAM-D) AND HAMD item # 1 (depressed mood) score=0)												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[6]	reporting bias[3]	4/17 (23.5%)	1/17 (5.9%)	RR 4 (0.5 to 32.2)	176 more per 1000 (from 29 fewer to 1000 more)	⊕OOO VERY LOW	
								5.9%		177 more per 1000 (from 30 fewer to 1000 more)		
Remission (paroxetine) (follow-up 11-12 weeks; assessed with: Number of people scoring <7/≤8 on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	serious[8]	serious[9]	no serious indirectness	very serious[10]	reporting bias[11]	38/78 (48.7%)	29/81 (35.8%)	RR 1.58 (0.68 to 3.66)	208 more per 1000 (from 115 fewer to 952 more)	⊕OOO VERY LOW	
								30.7%		178 more per 1000 (from 98 fewer to 817 more)		
Response (any SSRI) (follow-up 8-13 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D) AND HAMD score≤10/AND much/very much improved on CGI-I (score 1-2)/ AND/OR much/very much improved on CGI-I (score 1-2))												
6	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	175/294 (59.5%)	100/264 (37.9%)	RR 1.62 (1.29 to 2.03)	235 more per 1000 (from 110 more to 390 more)	⊕OOO VERY LOW	
								30.9%		192 more per 1000 (from 90 more to 318 more)		
Response (sertraline) (follow-up mean 12 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D) AND HAMD score≤10/Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
2	randomised trials	serious[1]	serious[9]	no serious indirectness	serious[6]	reporting bias[3]	102/168 (60.7%)	72/173 (41.6%)	RR 1.61 (0.99 to 2.64)	254 more per 1000 (from 4 fewer to 683 more)	⊕OOO VERY LOW	
								37.3%		228 more per 1000 (from 4 fewer to 612 more)		
Response (fluoxetine) (follow-up 8-13 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D) AND much/very much improved on CGI-I (score 1-2))												
2	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[7]	52/88 (59.1%)	17/55 (30.9%)	RR 1.96 (1.05 to 3.64)	297 more per 1000 (from 15 more to 816 more)		

								27.3%		262 more per 1000 (from 14 more to 721 more)	⊕OOO VERY LOW	
Response (escitalopram) (follow-up mean 12 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D) AND much/very much improved on CGI-I (score 1-2))												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	very serious[10]	reporting bias[3]	7/17 (41.2%)	5/17 (29.4%)	RR 1.4 (0.55 to 3.55)	118 more per 1000 (from 132 fewer to 750 more)	⊕OOO VERY LOW	
								29.4%		118 more per 1000 (from 132 fewer to 750 more)		
Response (paroxetine) (follow-up mean 12 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D) AND/OR much/very much improved on CGI-I (score 1-2))												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[2]	none	14/21 (66.7%)	6/19 (31.6%)	RR 2.11 (1.02 to 4.37)	351 more per 1000 (from 6 more to 1000 more)	⊕⊕OO LOW	
								31.6%		351 more per 1000 (from 6 more to 1000 more)		
Depression symptomatology (any SSRI) (follow-up 8-13 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
6	randomised trials	serious[1]	serious[9]	no serious indirectness	no serious imprecision	reporting bias[3]	293	263	-	SMD 0.69 lower (1.02 to 0.35 lower)	⊕OOO VERY LOW	
Depression symptomatology (sertraline) (follow-up mean 12 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
2	randomised trials	serious[1]	very serious[12]	no serious indirectness	serious[13]	reporting bias[3]	167	172	-	SMD 0.61 lower (1.3 lower to 0.07 higher)	⊕OOO VERY LOW	
Depression symptomatology (fluoxetine) (follow-up 8-13 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
2	randomised trials	serious[5]	very serious[12]	no serious indirectness	serious[13]	reporting bias[7]	88	55	-	SMD 0.8 lower (1.81 lower to 0.21 higher)	⊕OOO VERY LOW	
Depression symptomatology (escitalopram) (follow-up mean 12 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												

1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[14]	reporting bias[3]	17	17	-	SMD 0.9 lower (1.61 to 0.19 lower)	⊕OOO VERY LOW	
Depression symptomatology (paroxetine) (follow-up mean 12 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[14]	none	21	19	-	SMD 0.77 lower (1.41 to 0.12 lower)	⊕⊕OO LOW	
Discontinuation for any reason (any SSRI) (follow-up 8-13 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
6	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	46/318 (14.5%)	59/275 (21.5%)	RR 0.64 (0.42 to 0.96)	77 fewer per 1000 (from 9 fewer to 124 fewer)	⊕OOO VERY LOW	
								22.3%		80 fewer per 1000 (from 9 fewer to 129 fewer)		
Discontinuation for any reason (sertraline) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	25/168 (14.9%)	42/174 (24.1%)	RR 0.62 (0.4 to 0.97)	92 fewer per 1000 (from 7 fewer to 145 fewer)	⊕OOO VERY LOW	
								23.9%		91 fewer per 1000 (from 7 fewer to 143 fewer)		
Discontinuation for any reason (fluoxetine) (follow-up 8-13 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
2	randomised trials	serious[5]	serious[9]	no serious indirectness	very serious[15]	reporting bias[7]	15/110 (13.6%)	13/65 (20%)	RR 1.17 (0.11 to 12.85)	34 more per 1000 (from 178 fewer to 1000 more)	⊕OOO VERY LOW	
								13.3%		23 more per 1000 (from 118 fewer to 1000 more)		
Discontinuation for any reason (escitalopram) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	very serious[15]	reporting bias[3]	3/19 (15.8%)	0/17 (0%)	RR 6.3 (0.35 to 113.81)	-	⊕OOO VERY LOW	
								0%		-		
Discontinuation for any reason (paroxetine) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	very serious[15]	none	3/21 (14.3%)	4/19 (21.1%)	RR 0.68 (0.17 to 2.65)	67 fewer per 1000 (from 175 fewer to 347 more)	⊕OOO VERY LOW	
								21.1%		68 fewer per 1000 (from 175 fewer to 348 more)		

Discontinuation due to adverse events (any SSRI) (follow-up 8-12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[15]	reporting bias[3]	10/193 (5.2%)	5/192 (2.6%)	RR 1.83 (0.69 to 4.86)	22 more per 1000 (from 8 fewer to 101 more)	⊕OOO VERY LOW	
								0%		-		
Discontinuation due to adverse events (sertraline) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[15]	reporting bias[3]	8/134 (6%)	5/140 (3.6%)	RR 1.67 (0.56 to 4.98)	24 more per 1000 (from 16 fewer to 142 more)	⊕OOO VERY LOW	
								3.6%		24 more per 1000 (from 16 fewer to 143 more)		
Discontinuation due to adverse events (fluoxetine) (follow-up mean 8 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	very serious[15]	reporting bias[3]	1/19 (5.3%)	0/16 (0%)	RR 2.55 (0.11 to 58.6)	-	⊕OOO VERY LOW	
								0%		-		
Discontinuation due to adverse events (escitalopram) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	very serious[15]	reporting bias[3]	1/19 (5.3%)	0/17 (0%)	RR 2.7 (0.12 to 62.17)	-	⊕OOO VERY LOW	
								0%		-		
Discontinuation due to adverse events (paroxetine) (follow-up mean 12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[2]	none	0/21 (0%)	0/19 (0%)	not pooled	not pooled	⊕⊕OO LOW	
								0%		not pooled		

[1] Unclear (or high risk of bias associated with) randomisation method and unclear method of allocation concealment. Unclear blinding of intervention administration and outcome assessment
[2] Events<300
[3] Funding from pharmaceutical company
[4] High risk of bias associated with randomisation method due to significant differences between groups at baseline and unclear method of allocation concealment. Unclear blinding of intervention administration and outcome assessment. Unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)
[5] Unclear randomisation method and method of allocation concealment. Unclear blinding of intervention administration and outcome assessment
[6] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[7] Data is not reported for all outcomes
[8] Unclear blinding of intervention administrator(s)
[9] I-squared>50%
[10] 95% CI crosses line of no effect and threshold for both clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)
[11] Data is not reported for all outcomes and funding from pharmaceutical company
[12] I-squared>80%

[13] 95% CI crosses both line of no effect and threshold for clinically important benefit (SMD -0.5)
[14] N<400
[15] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

Sertraline versus imipramine for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Sertraline	Imipramine	Relative (95% CI)	Absolute		
Remission (follow-up mean 12 weeks; assessed with: Number of people scoring ≤4 on Hamilton Rating Scale for Depression (HAM-D)/≤7 on HAM-D AND much/very much improved on CGI-I (score 1-2))												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	133/555 (24%)	88/338 (26%)	RR 1.11 (0.89 to 1.39)	29 more per 1000 (from 29 fewer to 102 more)	⊕OOO VERY LOW	
								28.2%		31 more per 1000 (from 31 fewer to 110 more)		
Response (follow-up mean 12 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D) AND HAMD≤15 AND much/very much improved on CGI-I (score 1-2) AND CGI-S≤3 (mildly ill)/Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
2	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[3]	299/555 (53.9%)	191/338 (56.5%)	RR 0.97 (0.86 to 1.1)	17 fewer per 1000 (from 79 fewer to 57 more)	⊕⊕OO LOW	
								57.7%		17 fewer per 1000 (from 81 fewer to 58 more)		
Depression symptomatology (follow-up mean 12 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	134	136	-	SMD 0.05 higher (0.19 lower to 0.29 higher)	⊕OOO VERY LOW	
Discontinuation for any reason (follow-up mean 12 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												

2	randomised trials	serious[4]	serious[6]	no serious indirectness	serious[7]	reporting bias[3]	97/560 (17.3%)	95/345 (27.5%)	RR 0.61 (0.39 to 0.95)	107 fewer per 1000 (from 14 fewer to 168 fewer)	⊕OOO VERY LOW	
								28.5%		111 fewer per 1000 (from 14 fewer to 174 fewer)		
Discontinuation due to adverse events (follow-up mean 12 weeks; assessed with: Number of participants discontinuing due to adverse events)												
2	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[7]	reporting bias[3]	35/560 (6.3%)	50/345 (14.5%)	RR 0.45 (0.29 to 0.71)	80 fewer per 1000 (from 42 fewer to 103 fewer)	⊕OOO VERY LOW	
								15.2%		84 fewer per 1000 (from 44 fewer to 108 fewer)		

[1] High risk of bias associated with randomisation method due to significant difference between groups at baseline and method of allocation concealment unclear. Unclear blinding of intervention administration and outcome assessment. Unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)
[2] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[3] Funding from pharmaceutical company
[4] Unclear randomisation method and method of allocation concealment. Blinding of intervention administration and outcome assessment is unclear
[5] N<400
[6] I-squared>50%
[7] Events<300

Sertraline + IPT versus IPT-only for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Sertraline + IPT**	**IPT-only**	**Relative (95% CI)**	**Absolute**		
Remission (follow-up mean 16 weeks; assessed with: Number of people scoring <7 on Hamilton Rating Scale for Depression (HAM-D) AND >50% improvement on HAMD AND GAF score>70)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	11/21 (52.4%)	5/23 (21.7%)	RR 2.41 (1 to 5.79)	307 more per 1000 (from 0 more to 1000 more)	⊕OOO VERY LOW	
								21.7%		306 more per 1000 (from 0 more to 1000 more)		

Response (follow-up 16-26 weeks; assessed with: Number of people showing ≥40% improvement on Montgomery Asberg Depression Rating Scale (MADRS)/≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
2	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[5]	134/233 (57.5%)	91/201 (45.3%)	RR 1.26 (1.05 to 1.52)	118 more per 1000 (from 23 more to 235 more) 106 more per 1000 (from 20 more to 212 more)	⊕OOO VERY LOW	
								40.7%				
Depression symptomatology (follow-up 16-26 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
2	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[5]	233	201	-	SMD 0.5 lower (0.7 to 0.31 lower)	⊕OOO VERY LOW	
Discontinuation for any reason (follow-up mean 16 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias[5]	4/21 (19%)	4/23 (17.4%)	RR 1.1 (0.31 to 3.84)	17 more per 1000 (from 120 fewer to 494 more) 17 more per 1000 (from 120 fewer to 494 more)	⊕OOO VERY LOW	
								17.4%				

[1] High risk of bias associated with randomisation method due to significant difference between groups at baseline, and method of allocation concealment is unclear. Non-blind intervention administrator(s) and participants, although outcome assessors are blind

[2] Events<300

[3] Study partially funded by pharmaceutical company

[4] High risk of bias associated with randomisation method due to significant difference between groups at baseline. Non-blind intervention administrator(s) and participants, although outcome assessors are blind

[5] Data is not reported for all outcomes and funding from pharmaceutical company

[6] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

TCAs versus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**TCAs**	**Placebo**	**Relative (95% CI)**	**Absolute**		
Remission (imipramine) (follow-up 6-26 weeks; assessed with: Number of people scoring ≤4/6 on Hamilton Rating Scale for Depression (HAM-D)/<8 on Montgomery Asberg Depression Rating Scale (MADRS))												

4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	109/332 (32.8%)	80/335 (23.9%)	RR 1.38 (1.02 to 1.86)	91 more per 1000 (from 5 more to 205 more)	⊕⊕OO LOW	
								19.2%		73 more per 1000 (from 4 more to 165 more)		
Response (any TCA) (follow-up 6-26 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I)/Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
5	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	267/410 (65.1%)	152/421 (36.1%)	RR 1.85 (1.51 to 2.26)	307 more per 1000 (from 184 more to 455 more)	⊕⊕⊕O MODERATE	
								33.3%		283 more per 1000 (from 170 more to 420 more)		
Response (imipramine) (follow-up 6-26 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I)/Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
4	randomised trials	serious[1]	serious[3]	no serious indirectness	no serious imprecision	none	212/321 (66%)	125/337 (37.1%)	RR 1.86 (1.43 to 2.4)	319 more per 1000 (from 159 more to 519 more)	⊕⊕OO LOW	
								33.8%		291 more per 1000 (from 145 more to 473 more)		
Response (amineptine) (follow-up mean 13 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[2]	none	55/89 (61.8%)	27/84 (32.1%)	RR 1.92 (1.35 to 2.73)	296 more per 1000 (from 113 more to 556 more)	⊕⊕OO LOW	
								32.1%		295 more per 1000 (from 112 more to 555 more)		
Depression symptomatology (any TCA) (follow-up 8-13 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score)/Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
3	randomised trials	serious[1]	serious[3]	no serious indirectness	no serious imprecision	none	337	342	-	SMD 0.63 lower (0.95 to 0.3 lower)	⊕⊕OO LOW	
Depression symptomatology (imipramine) (follow-up 8-12 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
2	randomised trials	serious[1]	very serious[5]	no serious indirectness	no serious imprecision	none	230	237	-	SMD 0.64 lower (1.21 to 0.08 lower)	⊕OOO VERY LOW	
Depression symptomatology (amineptine) (follow-up mean 13 weeks; measured with: Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												

1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[6]	none	107	105	-	SMD 0.61 lower (0.88 to 0.33 lower)	⊕⊕OO LOW	
Discontinuation for any reason (any TCA) (follow-up 6-26 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
6	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[8]	none	147/468 (31.4%)	135/467 (28.9%)	RR 1.06 (0.85 to 1.31)	17 more per 1000 (from 43 fewer to 90 more)	⊕⊕OO LOW	
								31.6%		19 more per 1000 (from 47 fewer to 98 more)		
Discontinuation for any reason (imipramine) (follow-up 6-26 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
5	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[8]	none	107/357 (30%)	93/359 (25.9%)	RR 1.11 (0.83 to 1.49)	28 more per 1000 (from 44 fewer to 127 more)	⊕⊕OO LOW	
								24.3%		27 more per 1000 (from 41 fewer to 119 more)		
Discontinuation for any reason (amineptine) (follow-up mean 13 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	serious[9]	no serious inconsistency	no serious indirectness	very serious[10]	none	40/111 (36%)	42/108 (38.9%)	RR 0.93 (0.66 to 1.31)	27 fewer per 1000 (from 132 fewer to 121 more)	⊕OOO VERY LOW	
								38.9%		27 fewer per 1000 (from 132 fewer to 121 more)		
Discontinuation due to adverse events (any TCA) (follow-up 6-26 weeks; assessed with: Number of participants discontinuing due to adverse events)												
6	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[2]	none	63/468 (13.5%)	10/467 (2.1%)	RR 5.77 (3.09 to 10.79)	102 more per 1000 (from 45 more to 210 more)	⊕⊕OO LOW	
								1.4%		67 more per 1000 (from 29 more to 137 more)		
Discontinuation due to adverse events (imipramine) (follow-up 6-26 weeks; assessed with: Number of participants discontinuing due to adverse events)												
5	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[2]	none	58/357 (16.2%)	9/359 (2.5%)	RR 5.87 (3.05 to 11.29)	122 more per 1000 (from 51 more to 258 more)	⊕⊕OO LOW	
								1.9%		93 more per 1000 (from 39 more to 196 more)		
Discontinuation due to adverse events (amineptine) (follow-up mean 13 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[9]	no serious inconsistency	no serious indirectness	very serious[10]	none	5/111 (4.5%)	1/108 (0.9%)	RR 4.86 (0.58 to 40.96)	36 more per 1000 (from 4 fewer to 370 more)		

								0.9%		35 more per 1000 (from 4 fewer to 360 more)	⊕OOO VERY LOW	

[1] Unclear (or high risk of bias associated with) randomisation method and unclear method of allocation concealment. Unclear blinding of intervention administration and outcome assessment. Unclear risk of attrition bias (drop-out>20% and/or difference between groups>20% but ITT analysis used)
[2] Events<300
[3] I-squared>50%
[4] Unclear randomisation method and method of allocation concealment. Blinding of intervention administration and outcome assessment is unclear. Unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)
[5] I-squared>80%
[6] N<400
[7] Unclear (or high risk of bias associated with) randomisation method and unclear method of allocation concealment. Unclear blinding of intervention administration
[8] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 1.25)
[9] Unclear randomisation method and method of allocation concealment. Blinding of intervention administration unclear
[10] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

TCA versus antipsychotic for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**TCA**	**Antipsychotic**	**Relative (95% CI)**	**Absolute**		
Remission (imipramine versus amisulpride) (follow-up mean 26 weeks; assessed with: Number of people scoring <8 on Montgomery Asberg Depression Rating Scale (MADRS))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	24/73 (32.9%)	26/73 (35.6%)	RR 0.92 (0.59 to 1.45)	28 fewer per 1000 (from 146 fewer to 160 more)	⊕OOO VERY LOW	
								35.6%		28 fewer per 1000 (from 146 fewer to 160 more)		
Response (any TCA versus amisulpride) (follow-up 13-26 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
2	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[5]	none	101/162 (62.3%)	101/150 (67.3%)	RR 0.92 (0.78 to 1.09)	54 fewer per 1000 (from 148 fewer to 61 more)	⊕⊕OO LOW	
								67.3%		54 fewer per 1000 (from 148 fewer to 61 more)		
Response (amineptine versus amisulpride) (follow-up mean 13 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												

1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[6]	none	55/89 (61.8%)	54/77 (70.1%)	RR 0.88 (0.71 to 1.1)	84 fewer per 1000 (from 203 fewer to 70 more)	⊕⊕OO LOW	
								70.1%		84 fewer per 1000 (from 203 fewer to 70 more)		
Response (imipramine versus amisulpride) (follow-up mean 26 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	46/73 (63%)	47/73 (64.4%)	RR 0.98 (0.77 to 1.25)	13 fewer per 1000 (from 148 fewer to 161 more)	⊕OOO VERY LOW	
								64.4%		13 fewer per 1000 (from 148 fewer to 161 more)		
Depression symptomatology (amineptine versus amisulpride) (follow-up mean 13 weeks; measured with: Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[7]	none	107	101	-	SMD 0.06 higher (0.21 lower to 0.33 higher)	⊕⊕OO LOW	
Discontinuation for any reason (any TCA versus amisulpride) (follow-up 13-26 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
2	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	serious[9]	none	75/184 (40.8%)	67/177 (37.9%)	RR 1.09 (0.84 to 1.4)	34 more per 1000 (from 61 fewer to 151 more)	⊕⊕OO LOW	
								38.3%		34 more per 1000 (from 61 fewer to 153 more)		
Discontinuation for any reason (amineptine versus amisulpride) (follow-up mean 13 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	very serious[10]	none	40/111 (36%)	37/104 (35.6%)	RR 1.01 (0.71 to 1.45)	4 more per 1000 (from 103 fewer to 160 more)	⊕OOO VERY LOW	
								35.6%		4 more per 1000 (from 103 fewer to 160 more)		
Discontinuation for any reason (imipramine versus amisulpride) (follow-up mean 26 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	serious[9]	reporting bias[3]	35/73 (47.9%)	30/73 (41.1%)	RR 1.17 (0.81 to 1.68)	70 more per 1000 (from 78 fewer to 279 more)	⊕OOO VERY LOW	
								41.1%		70 more per 1000 (from 78 fewer to 279 more)		

Discontinuation due to adverse events (any TCA versus amisulpride) (follow-up 13-26 weeks; assessed with: Number of participants discontinuing due to adverse events)												
2	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias[3]	22/184 (12%)	10/177 (5.6%) 6.4%	RR 2.16 (1.08 to 4.35)	66 more per 1000 (from 5 more to 189 more) 74 more per 1000 (from 5 more to 214 more)	⊕OOO VERY LOW	
Discontinuation due to adverse events (amineptine versus amisulpride) (follow-up mean 13 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	very serious[10]	none	5/111 (4.5%)	2/104 (1.9%) 1.9%	RR 2.34 (0.46 to 11.81)	26 more per 1000 (from 10 fewer to 208 more) 25 more per 1000 (from 10 fewer to 205 more)	⊕OOO VERY LOW	
Discontinuation due to adverse events (imipramine versus amisulpride) (follow-up mean 26 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	serious[9]	reporting bias[3]	17/73 (23.3%)	8/73 (11%) 11%	RR 2.12 (0.98 to 4.61)	123 more per 1000 (from 2 fewer to 396 more) 123 more per 1000 (from 2 fewer to 397 more)	⊕OOO VERY LOW	

[1] Randomisation method and method of allocation concealment is unclear. Blinding of intervention administrator is also unclear and there is an unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)
[2] 95% CI crosses line of no effect and threshold for both clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)
[3] Data is not reported or cannot be extracted for all outcomes
[4] Randomisation method and method of allocation concealment is unclear. Blinding of intervention administration and outcome assessment is also unclear and there is an unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)
[5] Events<300
[6] 95% CI crosses both the line of no effect and the threshold for clinically important harm (RR 0.75)
[7] N<400
[8] Randomisation method and method of allocation concealment is unclear. Blinding of intervention administration is also unclear
[9] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 1.25)
[10] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

Duloxetine versus placebo for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Duloxetine**	**Placebo**	**Relative (95% CI)**	**Absolute**		

Remission (follow-up mean 10 weeks; assessed with: Number of people scoring ≤4 on Hamilton Rating Scale for Depression (HAM-D) AND HAMD item # 1 (depressed mood) score=0)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	16/29 (55.2%)	4/28 (14.3%)	RR 3.86 (1.47 to 10.13)	409 more per 1000 (from 67 more to 1000 more)	⊕OOO VERY LOW	
								14.3%		409 more per 1000 (from 67 more to 1000 more)		
Response (follow-up mean 10 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D) AND much/very much improved on CGI-I (score 1-2))												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	19/29 (65.5%)	7/28 (25%)	RR 2.62 (1.31 to 5.24)	405 more per 1000 (from 77 more to 1000 more)	⊕OOO VERY LOW	
								25%		405 more per 1000 (from 77 more to 1000 more)		
Depression symptomatology (follow-up mean 10 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias[3]	29	28	-	SMD 1.31 lower (1.89 to 0.74 lower)	⊕OOO VERY LOW	

[1] High risk of bias associated with randomisation method due to significant group difference at baseline and method of allocation concealment is unclear. Blinding of intervention administration and outcome assessment is also unclear

[2] Events<300

[3] Data cannot be extracted or is not reported for all outcomes and funding from pharmaceutical company

[4] N<400

Phenelzine versus placebo for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Phenelzine**	**Placebo**	**Relative (95% CI)**	**Absolute**		
Response (follow-up mean 6 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	7/12 (58.3%)	9/27 (33.3%)	RR 1.75 (0.85 to 3.58)	250 more per 1000 (from 50 fewer to 860 more)	⊕OOO VERY LOW	
								33.3%		250 more per 1000 (from 50 fewer to 859 more)		

[1] Unclear randomisation method and method of allocation concealment, and unclear blinding of intervention administrator(s)
[2] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[3] Data is not reported or cannot be extracted for all outcomes

Phenelzine versus imipramine for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Phenelzine	Imipramine	Relative (95% CI)	Absolute		
Response (follow-up mean 6 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	7/12 (58.3%)	14/18 (77.8%)	RR 0.75 (0.44 to 1.28)	194 fewer per 1000 (from 436 fewer to 218 more)	⊕OOO VERY LOW	
								77.8%		195 fewer per 1000 (from 436 fewer to 218 more)		
Depression symptomatology (follow-up mean 6 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D at endpoint); Better indicated by lower values)												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[5]	none	16	16	-	SMD 0.73 lower (1.45 to 0.01 lower)	⊕⊕OO LOW	
Discontinuation for any reason (follow-up mean 6 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	none	3/19 (15.8%)	4/20 (20%)	RR 0.79 (0.2 to 3.07)	42 fewer per 1000 (from 160 fewer to 414 more)	⊕OOO VERY LOW	
								20%		42 fewer per 1000 (from 160 fewer to 414 more)		
Discontinuation due to adverse events (follow-up mean 6 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	none	3/19 (15.8%)	4/20 (20%)	RR 0.79 (0.2 to 3.07)	42 fewer per 1000 (from 160 fewer to 414 more)	⊕OOO VERY LOW	
								20%		42 fewer per 1000 (from 160 fewer to 414 more)		

[1] Unclear randomisation method and method of allocation concealment, and unclear blinding of intervention administrator(s)
[2] 95% CI crosses line of no effect and threshold for both clinically important harm (RR 0.75) and clinically important benefit (RR 1.25)
[3] Data is not reported or cannot be extracted for all outcomes
[4] Unclear randomisation method and method of allocation concealment, and unclear blinding of intervention administration and outcome assessment

[5] N<400

[6] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

Moclobemide versus placebo for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Moclobemide	Placebo	Relative (95% CI)	Absolute		
Remission (follow-up mean 8 weeks; assessed with: Number of people scoring ≤4 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	33/104 (31.7%)	16/97 (16.5%)	RR 1.92 (1.13 to 3.27)	152 more per 1000 (from 21 more to 374 more)	⊕⊕OO LOW	
								16.5%		152 more per 1000 (from 21 more to 375 more)		
Response (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	74/104 (71.2%)	29/97 (29.9%)	RR 2.38 (1.71 to 3.31)	413 more per 1000 (from 212 more to 691 more)	⊕⊕OO LOW	
								29.9%		413 more per 1000 (from 212 more to 691 more)		
Depression symptomatology (follow-up mean 8 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	104	97	-	SMD 1.03 lower (1.33 to 0.74 lower)	⊕⊕OO LOW	
Discontinuation for any reason (follow-up mean 8 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	13/108 (12%)	15/104 (14.4%)	RR 0.83 (0.42 to 1.67)	25 fewer per 1000 (from 84 fewer to 97 more)	⊕OOO VERY LOW	
								14.4%		24 fewer per 1000 (from 84 fewer to 96 more)		
Discontinuation due to adverse events (follow-up mean 8 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	7/108 (6.5%)	2/104 (1.9%)	RR 3.37 (0.72 to 15.85)	46 more per 1000 (from 5 fewer to 286 more)		

								1.9%		45 more per 1000 (from 5 fewer to 282 more)	⊕OOO VERY LOW	

[1] Unclear randomisation method and method of allocation concealment, and unclear blinding of intervention administration and outcome assessment
[2] Events<300
[3] N<400
[4] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

Moclobemide versus imipramine for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Moclobemide	Imipramine	Relative (95% CI)	Absolute		
Remission (follow-up mean 8 weeks; assessed with: Number of people scoring ≤4 on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	33/104 (31.7%)	19/94 (20.2%)	RR 1.57 (0.96 to 2.56)	115 more per 1000 (from 8 fewer to 315 more)	⊕⊕OO LOW	
								20.2%		115 more per 1000 (from 8 fewer to 315 more)		
Response (follow-up mean 8 weeks; assessed with: Number of people showing ≥50% improvement on Hamilton Rating Scale for Depression (HAM-D))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	74/104 (71.2%)	65/94 (69.1%)	RR 1.03 (0.86 to 1.23)	21 more per 1000 (from 97 fewer to 159 more)	⊕⊕OO LOW	
								69.2%		21 more per 1000 (from 97 fewer to 159 more)		
Depression symptomatology (follow-up mean 8 weeks; measured with: Hamilton Rating Scale for Depression (HAM-D; change score); Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	104	94	-	SMD 0.16 lower (0.44 lower to 0.12 higher)	⊕⊕OO LOW	
Discontinuation for any reason (follow-up mean 8 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	none	13/108 (12%)	15/103 (14.6%)	RR 0.83 (0.41 to 1.65)	25 fewer per 1000 (from 86 fewer to 95 more)	⊕OOO VERY LOW	
								14.6%		25 fewer per 1000 (from 86 fewer to 95 more)		

Discontinuation due to adverse events (follow-up mean 8 weeks; assessed with: Number of participants discontinuing due to adverse events)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	none	7/108 (6.5%) 10.7%	11/103 (10.7%) 10.7%	RR 0.61 (0.24 to 1.51)	42 fewer per 1000 (from 81 fewer to 54 more) 42 fewer per 1000 (from 81 fewer to 55 more)	⊕OOO VERY LOW	

[1] Unclear randomisation method and method of allocation concealment, and unclear blinding of intervention administration and outcome assessment
[2] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[3] Events<300
[4] N<400
[5] 95% CI crosses line of no effect and threshold for both clinically important benefit (RR 0.75) and clinically important harm (RR 1.25)

Amisulpride versus placebo for chronic depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Amisulpride**	**Placebo**	**Relative (95% CI)**	**Absolute**		
Remission (follow-up mean 26 weeks; assessed with: Number of people scoring <8 on Montgomery Asberg Depression Rating Scale (MADRS))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	26/73 (35.6%) 21.9%	16/73 (21.9%) 21.9%	RR 1.62 (0.95 to 2.77)	136 more per 1000 (from 11 fewer to 388 more) 136 more per 1000 (from 11 fewer to 388 more)	⊕OOO VERY LOW	
Response (follow-up 13-26 weeks; assessed with: Number of people rated as much or very much improved on Clinical Global Impressions scale (CGI-I))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	101/150 (67.3%) 33.2%	52/157 (33.1%) 33.2%	RR 2.03 (1.59 to 2.61)	341 more per 1000 (from 195 more to 533 more) 342 more per 1000 (from 196 more to 535 more)	⊕⊕OO LOW	
Depression symptomatology (follow-up mean 13 weeks; measured with: Montgomery Asberg Depression Rating Scale (MADRS; change score); Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	101	105	-	SMD 0.68 lower (0.97 to 0.4 lower)	⊕⊕OO LOW	
Discontinuation for any reason (follow-up 13-26 weeks; assessed with: Number of participants discontinuing for any reason including adverse events)												

2	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	serious[7]	none	67/177 (37.9%)	78/181 (43.1%)	RR 0.87 (0.68 to 1.12)	56 fewer per 1000 (from 138 fewer to 52 more) 57 fewer per 1000 (from 141 fewer to 53 more)	⊕⊕OO LOW	
								44.1%				
Discontinuation due to adverse events (follow-up 13-26 weeks; assessed with: Number of participants discontinuing due to adverse events)												
2	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	serious[8]	reporting bias[3]	10/177 (5.6%)	3/181 (1.7%)	RR 3.31 (0.92 to 11.9)	38 more per 1000 (from 1 fewer to 181 more) 42 more per 1000 (from 1 fewer to 196 more)	⊕OOO VERY LOW	
								1.8%				

[1] Unclear randomisation method and method of allocation concealment and unclear blinding of intervention administrator(s). Unclear risk of attrition bias (drop-out>20% but difference between groups<20% and ITT analysis used)
[2] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 1.25)
[3] Data is not reported or cannot be extracted for all outcomes
[4] Events<300
[5] N<400
[6] Unclear randomisation method and method of allocation concealment and unclear blinding of intervention administrator(s).
[7] 95% CI crosses both line of no effect and threshold for clinically important benefit (RR 0.75)
[8] 95% CI crosses both line of no effect and threshold for clinically important harm (RR 1.25)

Complex depression (chapter 10)

CBT/behavioural therapies versus psychodynamic therapies

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**CBT/behavioural therapies**	**Psychodynamic therapies**	**Relative (95% CI)**	**Absolute**		
Depression symptomatology at endpoint (measured with: BDI; Better indicated by lower values)												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	26	25	-	MD 6.35 lower (13.18 lower to 0.47 higher)	⊕OOO VERY LOW	CRITICAL
Depression symptomatology (follow-up 12 weeks; measured with: BDI; Better indicated by lower values)												

2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	26	25	-	MD 0.3 lower (0.86 lower to 0.25 higher)	⊕OOO VERY LOW	CRITICAL
Depression symptomatology (follow-up 24 weeks; measured with: BDI; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	12	12	-	MD 9.00 lower (16.09 to 1.91 lower)	⊕OOO VERY LOW	CRITICAL
Depression symptomatology (follow-up 36 weeks; measured with: BDI; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	12	12	-	MD 3.00 lower (11.84 lower to 5.84 higher)	⊕OOO VERY LOW	CRITICAL
Depression symptomatology (follow-up 1 years; measured with: BDI; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	14	13	-	MD 0.25 higher (6.87 lower to 7.37 higher)	⊕OOO VERY LOW	CRITICAL
Suicide attempts (follow-up 24 weeks)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	3/12 (25%) 33.3%	4/12 (33.3%) 33.3%	RR 0.75 (0.21 to 2.66)	83 fewer per 1000 (from 263 fewer to 553 more) 83 fewer per 1000 (from 263 fewer to 553 more)	⊕OOO VERY LOW	CRITICAL
Suicide attempts (2 year follow-up) (follow-up 2 years)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	5/12 (41.7%) 50%	6/12 (50%) 50%	RR 0.83 (0.35 to 2.00)	85 fewer per 1000 (from 325 fewer to 500 more) 85 fewer per 1000 (from 325 fewer to 500 more)	⊕OOO VERY LOW	CRITICAL
Discontinuations for any reason												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	7/36 (19.4%)	10/37 (27%)	RR 0.73 (0.33 to 1.60)	73 fewer per 1000 (from 181 fewer to 162 more)	⊕OOO VERY LOW	CRITICAL
								27%		73 fewer per 1000 (from 181 fewer to 162 more)		

[1] High ROB across multiple domains
[2] 95% CI crosses one clinical decision threshold
[3] OIS not met (<400 participants)
[4] 95% CI crosses two clinical decision thresholds

Pharmacotherapy versus combination therapy (pharmacotherapy + SPSP)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Pharmacotherapy versus combi therapy (pharm + SPSP)**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression symptomatology (measured with: HAM-D 17; Better indicated by lower values)												
2	randomised trials	very serious[1]	very serious[2]	no serious indirectness	very serious[3]	none	46	58	-	MD 8 higher (1.35 lower to 17.34 higher)	⊕OOO VERY LOW	CRITICAL
Depression symptomatology at endpoint (pharm protocol versus pharm + SPSP) (follow-up mean 24 weeks; measured with: HAM-D 17; Better indicated by lower values)												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	none	36	49	-	MD 3.79 higher (0.36 to 7.22 higher)	⊕OOO VERY LOW	CRITICAL
Depression symptomatology (lofepramine alone versus lofepramine + RET) (Better indicated by lower values)												
1	randomised trials	very serious[6]	no serious inconsistency	no serious indirectness	serious[7]	none	10	9	-	MD 13.4 higher (5.92 to 20.88 higher)	⊕OOO VERY LOW	CRITICAL

Remission at endpoint (follow-up mean 24 weeks; assessed with: HAM-D 17)												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	none	7/36 (19.4%)	23/49 (46.9%)	RR 0.41 (0.2 to 0.86)	277 fewer per 1000 (from 66 fewer to 376 fewer)	⊕OOO VERY LOW	CRITICAL
								0%		-		
Discontinuations for any reason												
1	randomised trials	very serious[6]	no serious inconsistency	no serious indirectness	very serious[3]	none	0/10 (0%)	1/10 (10%)	RR 0.33 (0.02 to 7.32)	67 fewer per 1000 (from 98 fewer to 632 more)	⊕OOO VERY LOW	CRITICAL
								10%		67 fewer per 1000 (from 98 fewer to 632 more)		

[1] High or unclear ROB across multiple domains
[2] I2 >80%
[3] 95% CI crosses two clinical decision thresholds
[4] High risk of bias for selective outcome reporting and allocation concealment unlikely to affect results, however unclear effect of bias from missing outcome data
[5] Confidence intervals cross 1 minimally important difference. Sample size less than optimal information size (<400 for continuous outcomes or <300 events for dichotomous outcomes).
[6] High ROB across multiple domains
[7] OIS not met (<400 participants)

Psychotic depression (chapter 10)

Antidepressants versus other pharmacological interventions

Antidepressants versus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Antidepressant versus placebo**	**Control**	**Relative (95% CI)**	**Absolute**		
Depressive symptoms at endpoint (HAMD 17) - TCA versus placebo (Better indicated by lower values)												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	69	67	-	MD 3 lower (4.71 to 1.29 lower)	⊕⊕OO LOW	CRITICAL
Remission - TCA versus placebo												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	none	4/10 (40%)	0/10 (0%)	RR 9 (0.55 to 147.95)	-	⊕OOO VERY LOW	CRITICAL
								0%		-		
Response - TCA versus placebo												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	53/69 (76.8%)	15/67 (22.4%)	not pooled	not pooled	⊕⊕OO LOW	CRITICAL
								22.4%		not pooled		
Discontinuation - TCA versus placebo												
2	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	none	7/86 (8.1%)	3/87 (3.4%)	RR 1.88 (0.4 to 8.82)	30 more per 1000 (from 21 fewer to 270 more)	⊕OOO VERY LOW	CRITICAL
								11.5%		101 more per 1000 (from 69 fewer to 899 more)		

[1] Unclear ROB across multiple domians
[2] OIS not met (<400 participants)
[3] High ROB in one domain and unclear in several others
[4] 95% CI crosses two clinical decision thresholds
[5] OIS not met (<300 events)

Antidepressants versus antidepressants

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Antidepressant versus antidepressant**	**Control**	**Relative (95% CI)**	**Absolute**		
Depressive symptoms at endpoint - TCA versus SNRI (Better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	17	12	-	MD 1.1 higher (1.47 lower to 3.67 higher)	⊕⊕OO LOW	CRITICAL

Depressive symmptoms at endpoint - TCA (clomipramine) versus TCA (imipramine) (Better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	12	10	-	MD 0.3 higher (8.72 lower to 9.32 higher)	⊕⊕OO LOW	CRITICAL
Remission - SSRI versus SNRI												
1	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	9/11 (81.8%)	6/11 (54.5%)	RR 1.5 (0.82 to 2.75)	273 more per 1000 (from 98 fewer to 955 more)	⊕⊕OO LOW	CRITICAL
								54.6%		273 more per 1000 (from 98 fewer to 956 more)		
Remission - SSRI (sertraline) versus SSRI (paroxetine)												
1	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	13/18 (72.2%)	3/14 (21.4%)	RR 3.37 (1.19 to 9.57)	508 more per 1000 (from 41 more to 1000 more)	⊕⊕OO LOW	CRITICAL
								21.4%		507 more per 1000 (from 41 more to 1000 more)		
Remission - TCA versus SNRI												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	none	15/20 (75%)	11/12 (91.7%)	RR 0.82 (0.6 to 1.11)	165 fewer per 1000 (from 367 fewer to 101 more)	⊕⊕⊕O MODERATE	CRITICAL
								91.7%		165 fewer per 1000 (from 367 fewer to 101 more)		
Response - TCA versus atypical ADM												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[1]	none	9/15 (60%)	7/15 (46.7%)	RR 1.29 (0.65 to 2.54)	135 more per 1000 (from 163 fewer to 719 more)	⊕OOO VERY LOW	CRITICAL
								46.7%		135 more per 1000 (from 163 fewer to 719 more)		
Response - TCA versus SNRI												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	none	16/20 (80%)	12/13 (92.3%)	RR 0.87 (0.66 to 1.13)	120 fewer per 1000 (from 314 fewer to 120 more) 120 fewer per 1000 (from 314 fewer to 120 more)	⊕⊕⊕O MODERATE	CRITICAL
								92.3%				
Response - TCA versus SSRI												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[3]	none	16/25 (64%)	7/25 (28%)	RR 2.29 (1.14 to 4.58)	361 more per 1000 (from 39 more to 1000 more) 361 more per 1000 (from 39 more to 1000 more)	⊕⊕OO LOW	CRITICAL
								28%				
Discontinuation - TCA versus atypical antidepressant												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[1]	none	4/15 (26.7%)	8/15 (53.3%)	RR 0.5 (0.19 to 1.31)	267 fewer per 1000 (from 432 fewer to 165 more) 266 fewer per 1000 (from 432 fewer to 165 more)	⊕OOO VERY LOW	CRITICAL
								53.3%				
Discontinuation - TCA versus SSRI												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[1]	none	4/25 (16%)	2/25 (8%)	RR 2 (0.4 to 9.95)	80 more per 1000 (from 48 fewer to 716 more) 80 more per 1000 (from 48 fewer to 716 more)	⊕OOO VERY LOW	CRITICAL
								8%				
Discontinuation - TCA versus SNRI												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	none	3/20 (15%)	1/13 (7.7%)	RR 1.95 (0.23 to 16.79)	73 more per 1000 (from 59 fewer to 1000 more) 73 more per 1000 (from 59 fewer to 1000 more)	⊕⊕OO LOW	CRITICAL
								7.7%				

Discontinuation - TCA (clomipramine) versus TCA (imipramine)												
1	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	very serious[1]	none	0/12 (0%) 16.7%	2/12 (16.7%)	RR 0.2 (0.01 to 3.77)	133 fewer per 1000 (from 165 fewer to 462 more) 134 fewer per 1000 (from 165 fewer to 463 more)	⊕OOO VERY LOW	CRITICAL
Discontinuation - SSRI (sertraline) versus SSRI (paroxetine)												
1	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	0/18 (0%) 35.7%	5/14 (35.7%)	RR 0.07 (0 to 1.2)	332 fewer per 1000 (from 357 fewer to 71 more) 332 fewer per 1000 (from 357 fewer to 71 more)	⊕⊕OO LOW	CRITICAL
Discontinuation - SSRI versus SNRI												
1	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	very serious[1]	none	0/11 (0%) 18.2%	2/11 (18.2%)	RR 0.2 (0.01 to 3.74)	145 fewer per 1000 (from 180 fewer to 498 more) 146 fewer per 1000 (from 180 fewer to 499 more)	⊕OOO VERY LOW	CRITICAL
Discontinuation due to side effects - TCA (clomipramine) versus TCA (imipramine)												
1	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	very serious[1]	none	0/12 (0%) 16.7%	2/12 (16.7%)	RR 0.2 (0.01 to 3.77)	133 fewer per 1000 (from 165 fewer to 462 more) 134 fewer per 1000 (from 165 fewer to 463 more)	⊕OOO VERY LOW	CRITICAL

[1] 95% CI crosses two clinical decision thresholds
[2] Unclear ROB across multiple domains
[3] 95% CI crosses one clinical decision threshold
[4] High ROB in at least one domain and unclear in several others
[5] No explanation was provided

Antidepressants versus antipsychotics

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antidepressant versus antipsychotic	Control	Relative (95% CI)	Absolute		
Remission - TCA versus antipsychotic												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	7/19 (36.8%)	3/17 (17.6%)	not pooled	not pooled	⊕⊕OO LOW	CRITICAL
								17.7%		not pooled		
Discontinuation - TCA versus antipsychotic												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	2/19 (10.5%)	1/17 (5.9%)	not pooled	not pooled	⊕⊕OO LOW	CRITICAL
								5.9%		not pooled		

[1] 95% CI crosses two clinical decision thresholds

Antidepressants versus combined antipsychotic and antidepressants

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antidepressant versus antipsychotic + antidepressant	Control	Relative (95% CI)	Absolute		
Depression symptomatology at endpoint (HAMD-17) - SNRI versus antipsychotic + SNRI (Better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	12	24	-	MD 0.3 lower (2.44 lower to 1.84 higher)	⊕⊕OO LOW	CRITICAL
Depression symptomatology at endpoint (HAMD-17) - Tetracyclic versus antipsychotic +TCA (Better indicated by lower values)												

1	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	very serious[1]	none	17	18	-	MD 0.9 higher (5 lower to 6.8 higher)	⊕OOO VERY LOW	CRITICAL
Depression symptomatology at endpoint (HAMD-17) - TCA versus antipsychotic + SNRI (Better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	17	24	-	MD 1.4 lower (4.12 lower to 1.32 higher)	⊕⊕OO LOW	CRITICAL
Remission - TCA versus TCA + antipsychotic												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	none	7/17 (41.2%)	14/18 (77.8%)	RR 0.53 (0.28 to 0.98)	366 fewer per 1000 (from 16 fewer to 560 fewer)	⊕⊕⊕O MODERATE	CRITICAL
								77.8%		366 fewer per 1000 (from 16 fewer to 560 fewer)		
Remission - SNRI versus antipsychotic + SNRI												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	none	11/12 (91.7%)	20/24 (83.3%)	RR 1.1 (0.86 to 1.41)	83 more per 1000 (from 117 fewer to 342 more)	⊕⊕⊕O MODERATE	CRITICAL
								83.3%		83 more per 1000 (from 117 fewer to 342 more)		
Remission - TCA versus antipsychotic + SNRI												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	none	15/17 (88.2%)	20/24 (83.3%)	RR 1.06 (0.83 to 1.36)	50 more per 1000 (from 142 fewer to 300 more)	⊕⊕⊕O MODERATE	CRITICAL
								83.3%		50 more per 1000 (from 142 fewer to 300 more)		
Response - SNRI versus antipsychotic + SNRI												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	none	12/12 (100%)	23/24 (95.8%)	RR 1.02 (0.88 to 1.18)	19 more per 1000 (from 115 fewer to 172 more)	⊕⊕⊕O MODERATE	CRITICAL

								95.8%		19 more per 1000 (from 115 fewer to 172 more)		
Response - Tetracyclic versus antipsychotic + TCA													
1	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	12/17 (70.6%)	17/18 (94.4%)	RR 0.75 (0.54 to 1.04)	236 fewer per 1000 (from 434 fewer to 38 more)	⊕OOO VERY LOW	CRITICAL
								94.4%		236 fewer per 1000 (from 434 fewer to 38 more)		
Response - TCA versus antipsychotic + SNRI													
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	none	16/17 (94.1%)	23/24 (95.8%)	RR 0.98 (0.85 to 1.14)	19 fewer per 1000 (from 144 fewer to 134 more)	⊕⊕⊕O MODERATE	CRITICAL
								95.8%		19 fewer per 1000 (from 144 fewer to 134 more)		
Discontinuation - SNRI versus antipsychotic + SNRI													
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	1/13 (7.7%)	2/26 (7.7%)	RR 1 (0.1 to 10.04)	0 fewer per 1000 (from 69 fewer to 695 more)	⊕⊕OO LOW	CRITICAL
								7.7%		0 fewer per 1000 (from 69 fewer to 696 more)		
Discontinuation - Tetracyclic versus antipsychotic + TCA													
1	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	very serious[1]	none	9/21 (42.9%)	7/25 (28%)	RR 1.53 (0.69 to 3.4)	148 more per 1000 (from 87 fewer to 672 more)	⊕OOO VERY LOW	CRITICAL
								28%		148 more per 1000 (from 87 fewer to 672 more)		
Discontinuation - TCA versus antipsychotic + SNRI													

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	3/20 (15%)	2/26 (7.7%)	RR 1.95 (0.36 to 10.58)	73 more per 1000 (from 49 fewer to 737 more) 73 more per 1000 (from 49 fewer to 738 more)	⊕⊕OO LOW	CRITICAL
								7.7%				
Discontinuation - TCA versus antipsychotic + TCA												
2	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	very serious[1]	none	16/68 (23.5%)	17/67 (25.4%)	RR 0.92 (0.51 to 1.66)	20 fewer per 1000 (from 124 fewer to 167 more) 19 fewer per 1000 (from 115 fewer to 155 more)	⊕OOO VERY LOW	CRITICAL
								23.5%				
Discontinuation due to side effects - TCA versus antipsychotic + TCA												
2	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	very serious[1]	none	5/68 (7.4%)	10/67 (14.9%)	RR 0.52 (0.19 to 1.39)	72 fewer per 1000 (from 121 fewer to 58 more) 64 fewer per 1000 (from 109 fewer to 52 more)	⊕OOO VERY LOW	CRITICAL
								13.4%				

[1] 95% CI crosses two clinical decision thresholds
[2] High or unclear ROB in most domains
[3] 95% CI crosses one clinical decision threshold
[4] OIS not met (<300 participants)
[5] Unclear ROB across multiple domains

Combined antidepressants and antipsychotics versus other pharmacological interventions

Antidepressants plus antipsychotics versus antidepressants plus placebo

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antidepressant + antipsychotic versus antidepressant + placebo	Control	Relative (95% CI)	Absolute		
Depression symptomatology at endpoint (HAMD-17) - TCA + antipsychotic versus TCA + placebo (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	14	16	-	MD 1 higher (4.24 lower to 6.24 higher)	⊕OOO VERY LOW	CRITICAL
Remission - TCA + antipsychotic versus TCA + placebo												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	7/14 (50%)	7/16 (43.8%)	RR 1.14 (0.53 to 2.45)	61 more per 1000 (from 206 fewer to 634 more)	⊕OOO VERY LOW	CRITICAL
								43.8%		61 more per 1000 (from 206 fewer to 635 more)		
Treatment discontinuation - TCA + antipsychotic versus TCA + placebo												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	3/17 (17.6%)	3/19 (15.8%)	RR 1.12 (0.26 to 4.81)	19 more per 1000 (from 117 fewer to 602 more)	⊕OOO VERY LOW	CRITICAL
								15.8%		19 more per 1000 (from 117 fewer to 602 more)		

[1] High ROB in one domain, unclear ROB in several others

[2] 95% CI crosses two clinical decision thresholds

Antidepressants plus antipsychotics versus antipsychotics plus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antidepressant + antipsychotic versus antipsychotic + placebo	Control	Relative (95% CI)	Absolute		
Remission - SSRI + antipsychotic versus antipsychotic + placebo												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	54/81 (66.7%)	31/61 (50.8%)	RR 1.31 (0.98 to 1.75)	158 more per 1000 (from 10 fewer to 381 more)	⊕⊕⊕O MODERATE	CRITICAL
								50.8%		157 more per 1000 (from 10 fewer to 381 more)		
Treatment discontinuation - SSRI + antipsychotic versus antipsychotic + placebo												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	48/129 (37.2%)	69/130 (53.1%)	RR 0.7 (0.53 to 0.92)	159 fewer per 1000 (from 42 fewer to 249 fewer)	⊕⊕⊕O MODERATE	CRITICAL
								53.1%		159 fewer per 1000 (from 42 fewer to 250 fewer)		

[1] 95% CI crosses one clinical decision threshold

Antipsychotics versus other pharmacological interventions

Antipsychotics versus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antipsychotic versus placebo	Control	Relative (95% CI)	Absolute		
Response - Olanzapine versus placebo												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	32/63 (50.8%)	28/53 (52.8%)	RR 0.94 (0.67 to 1.31)	32 fewer per 1000 (from 174 fewer to 164 more)	⊕OOO VERY LOW	CRITICAL
								55.2%		33 fewer per 1000 (from 182 fewer to 171 more)		

Treatment discontinuation - Olanzapine versus placebo													
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	38/101 (37.6%)	47/100 (47%)	RR 0.8 (0.58 to 1.09)	94 fewer per 1000 (from 197 fewer to 42 more)	⊕⊕OO LOW	CRITICAL
								47.2%		94 fewer per 1000 (from 198 fewer to 42 more)		

[1] Unclear ROB in most domains and high ROB in one
[2] 95% CI crosses two clinical decision thresholds
[3] 95% CI crosses one clinical decision threshold

Antipsychotics versus antipsychotics plus antidepressants

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Antipsychotic versus antipsychotic + antidepressant**	**Control**	**Relative (95% CI)**	**Absolute**		
Response - antipsychotic versus SSRI + antipsychotic												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	15/35 (42.9%)	14/14 (100%)	RR 0.45 (0.3 to 0.66)	550 fewer per 1000 (from 340 fewer to 700 fewer)	⊕⊕OO LOW	CRITICAL
								100%		550 fewer per 1000 (from 340 fewer to 700 fewer)		
Treatment discontinuation - antipsychotic versus antipsychotic +SSRI												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	13/48 (27.1%)	11/25 (44%)	RR 0.62 (0.32 to 1.17)	167 fewer per 1000 (from 299 fewer to 75 more)	⊕⊕OO LOW	CRITICAL
								44%		167 fewer per 1000 (from 299 fewer to 75 more)		

[1] Unclear ROB in most domains, and high ROB in one
[2] OIS not met (<300 participants)
[3] 95% CI crosses one clinical decision threshold

Benzodiazepines versus other pharmacological interventions

Benzodiazepines versus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Benzodiazepines versus placebo	Control	Relative (95% CI)	Absolute		
Depression symptomatology at endpoint (HAMD-17) - Lorazepam versus placebo (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	59	67	-	MD 3.7 lower (5.6 to 1.8 lower)	⊕⊕OO LOW	CRITICAL
Depression symptomatology at endpoint (HAMD-17) - Alprazolam versus placebo (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	62	67	-	MD 3.2 lower (5.03 to 1.37 lower)	⊕⊕OO LOW	CRITICAL
Response - Lorazepam versus placebo												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	40/59 (67.8%)	15/67 (22.4%)	RR 3.03 (1.88 to 4.89)	454 more per 1000 (from 197 more to 871 more)	⊕⊕OO LOW	CRITICAL
								22.4%		455 more per 1000 (from 197 more to 871 more)		
Response - Alprazolam versus placebo												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	41/62 (66.1%)	15/67 (22.4%)	RR 2.95 (1.83 to 4.77)	437 more per 1000 (from 186 more to 844 more)	⊕⊕OO LOW	CRITICAL
								22.4%		437 more per 1000 (from 186 more to 844 more)		
Treatment discontinuation - Lorazepam versus placebo												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	7/66 (10.6%)	7/74 (9.5%)	RR 1.12 (0.42 to 3.03)	11 more per 1000 (from 55 fewer to 192 more) 11 more per 1000 (from 55 fewer to 193 more)	⊕OOO VERY LOW	CRITICAL
								9.5%				
Treatment discontinuation - Alprazolam versus placebo												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	8/70 (11.4%)	7/74 (9.5%)	RR 1.21 (0.46 to 3.16)	20 more per 1000 (from 51 fewer to 204 more) 20 more per 1000 (from 51 fewer to 205 more)	⊕OOO VERY LOW	CRITICAL
								9.5%				
Discontinuation due to side effects - Lorazepam versus placebo												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	1/66 (1.5%)	0/74 (0%)	RR 3.36 (0.14 to 81.05)	-	⊕OOO VERY LOW	CRITICAL
								0%		-		
Discontinuation due to side effects - Alprazolam versus placebo												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	3/70 (4.3%)	0/74 (0%)	RR 7.39 (0.39 to 140.62)	-	⊕OOO VERY LOW	CRITICAL
								0%		-		

[1] Unclear ROB in most domains
[2] OIS not met (<400 participants)
[3] OIS not met (<300 events)
[4] 95% CI crosses two clinical decision thresholds

Benzodiazepines versus antidepressants

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Benzodiazepines versus antidepressants**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression symptomatology at endpoint (HAMD-17) - Lorazepam versus TCA (Better indicated by lower values)												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	59	69	-	MD 0.7 lower (2.59 lower to 1.19 higher)	⊕OOO VERY LOW	CRITICAL
Depression symptomatology at endpoint (HAMD-17) - Alprazolam versus TCA (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	62	69	-	MD 0.2 lower (2.02 lower to 1.62 higher)	⊕OOO VERY LOW	CRITICAL
Response - Lorazepam versus TCA												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	40/59 (67.8%)	53/69 (76.8%)	RR 0.88 (0.71 to 1.1)	92 fewer per 1000 (from 223 fewer to 77 more)	⊕⊕OO LOW	CRITICAL
								76.8%		92 fewer per 1000 (from 223 fewer to 77 more)		
Response - Alprazolam versus TCA												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	41/62 (66.1%)	53/69 (76.8%)	RR 0.86 (0.69 to 1.07)	108 fewer per 1000 (from 238 fewer to 54 more)	⊕⊕OO LOW	CRITICAL
								76.8%		108 fewer per 1000 (from 238 fewer to 54 more)		
Treatment discontinuation - Lorazepam versus TCA												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	7/66 (10.6%)	3/72 (4.2%)	RR 2.55 (0.69 to 9.44)	65 more per 1000 (from 13 fewer to 352 more)	⊕OOO VERY LOW	CRITICAL
								4.2%		65 more per 1000 (from 13 fewer to 354 more)		
Treatment discontinuation - Alprazolam versus TCA												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	8/70 (11.4%)	3/72 (4.2%)	RR 2.74 (0.76 to 9.92)	73 more per 1000 (from 10 fewer to 372 more)	⊕⊕OO LOW	CRITICAL

								4.2%		73 more per 1000 (from 10 fewer to 375 more)		
Discontinuation due to side effects - Lorazepam versus TCA												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	1/66 (1.5%)	0/72 (0%)	RR 3.27 (0.14 to 78.87)	-	⊕OOO VERY LOW	CRITICAL
								0%		-		
Discontinuation due to side effects - Alprazolam versus TCA												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	3/70 (4.3%)	0/72 (0%)	RR 7.2 (0.38 to 136.84)	-	⊕OOO VERY LOW	CRITICAL
								0%		-		

[1] Unclear ROB in most domains
[2] 95% CI crosses two clinical decision thresholds
[3] 95% CI crosses one clinical decision threshold

Benzodiazepines versus benzodiazepines

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Benzodiazepines versus benzodiazepines**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression symptomatology at endpoint (HAMD-17) - Lorazepam versus alprazolam (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	59	62	-	MD 0.5 lower (2.5 lower to 1.5 higher)	⊕OOO VERY LOW	CRITICAL
Response - Lorazepam versus alprazolam												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	40/59 (67.8%)	41/62 (66.1%)	RR 1.03 (0.8 to 1.32)	20 more per 1000 (from 132 fewer to 212 more)	⊕⊕OO LOW	CRITICAL
								66.1%		20 more per 1000 (from 132 fewer to 212 more)		

Treatment discontinuation - Lorazepam versus alprazolam												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	7/66 (10.6%)	8/70 (11.4%)	RR 0.93 (0.36 to 2.42)	8 fewer per 1000 (from 73 fewer to 162 more) 8 fewer per 1000 (from 73 fewer to 162 more)	⊕OOO VERY LOW	CRITICAL
								11.4%				
Discontinuation due to side effects - Lorazepam versus alprazolam												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	1/66 (1.5%)	3/70 (4.3%)	RR 0.35 (0.04 to 3.31)	28 fewer per 1000 (from 41 fewer to 99 more) 28 fewer per 1000 (from 41 fewer to 99 more)	⊕OOO VERY LOW	CRITICAL
								4.3%				

[1] Unclear ROB across most domains
[2] 95% CI crosses two clinical decision thresholds
[3] 95% CI crosses one clinical decision threshold

Relapse prevention (chapter 11)

Psychological interventions

Psychological interventions versus control

CBT/CT versus control for relapse prevention

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**CBT/CT**	**Control**	**Relative (95% CI)**	**Absolute**		
Relapse (follow-up 12 months)												

4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	95/246 (38.6%)	124/225 (55.1%)	RR 0.71 (0.53 to 0.95)	160 fewer per 1000 (from 28 fewer to 259 fewer) 161 fewer per 1000 (from 28 fewer to 261 fewer)	⊕⊕OO LOW	CRITICAL
								55.5%				
Relapse (follow-up 24 months)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	131/224 (58.5%)	144/202 (71.3%)	RR 0.82 (0.69 to 0.98)	128 fewer per 1000 (from 14 fewer to 221 fewer) 133 fewer per 1000 (from 15 fewer to 229 fewer)	⊕⊕OO LOW	CRITICAL
								73.9%				

[1] ROB unclear or high in 1-2 domains for each study
[2] 95% CI crosses one clinical decision threshold

MBCT versus control for relapse prevention

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**MBCT versus control**	**Control**	**Relative (95% CI)**	**Absolute**		
Relapse (follow-up 12 months)												
9	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	247/525 (47%)	281/475 (59.2%)	RR 0.79 (0.7 to 0.89)	124 fewer per 1000 (from 65 fewer to 177 fewer) 125 fewer per 1000 (from 65 fewer to 178 fewer)	⊕⊕OO LOW	CRITICAL
								59.4%				
Relapse (follow-up 24 months)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	148/313 (47.3%)	161/314 (51.3%)	RR 0.92 (0.79 to 1.08)	41 fewer per 1000 (from 108 fewer to 41 more) 43 fewer per 1000 (from 112 fewer to 43 more)	⊕⊕⊕O MODERATE	CRITICAL
								53.5%				

[1] ROB unclear or high in 1-2 domains for most studies
[2] 95% CI crosses one clinical decision threshold

IPT versus control for relapse prevention

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	IPT	Control	Relative (95% CI)	Absolute		
Relapse (follow-up 12 months)												
3	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	69/118 (58.5%)	57/75 (76%)	RR 0.77 (0.63 to 0.95)	175 fewer per 1000 (from 38 fewer to 281 fewer) 180 fewer per 1000 (from 39 fewer to 290 fewer)	⊕OOO VERY LOW	CRITICAL
								78.3%				
Relapse (follow-up 24 months)												
3	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	75/116 (64.7%)	46/71 (64.8%)	RR 0.89 (0.74 to 1.07)	71 fewer per 1000 (from 168 fewer to 45 more) 79 fewer per 1000 (from 188 fewer to 51 more)	⊕OOO VERY LOW	CRITICAL
								72.2%				

[1] ROB unclear or high across multiple domains in most included studies
[2] 95% CI crosses one clinical decision threshold

'Other' psychological interventions versus control for relapse prevention

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Other psychological interventions	Control	Relative (95% CI)	Absolute		

CBASP vs control - Relapse (follow-up 12 months)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	1/42 (2.4%)	8/40 (20%)	RR 0.12 (0.02 to 0.91)	176 fewer per 1000 (from 18 fewer to 196 fewer) 176 fewer per 1000 (from 18 fewer to 196 fewer)	⊕OOO VERY LOW	CRITICAL
								20%				

[1] ROB high or unclear across multiple domains
[2] 95% CI crosses one clinical decision threshold

Psychological interventions versus psychological interventions

CBT versus psychoeducation

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	CBT	Psychoeducation	Relative (95% CI)	Absolute		
Relapse (follow-up 12 months)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	46/90 (51.1%)	54/90 (60%)	RR 0.85 (0.65 to 1.11)	90 fewer per 1000 (from 210 fewer to 66 more) 90 fewer per 1000 (from 210 fewer to 66 more)	⊕⊕OO LOW	CRITICAL
								60%				

[1] ROB unclear or high in 1-2 domains
[2] 95% CI crosses one clinical decision threshold

IPT versus IPT

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	IPT	IPT	Relative (95% CI)	Absolute		
Relapse - Weekly IPT vs Bi-monthly IPT (follow-up 24 months)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	23/43 (53.5%)	19/44 (43.2%)	RR 1.24 (0.8 to 1.92)	104 more per 1000 (from 86 fewer to 397 more)	⊕⊕OO LOW	CRITICAL
								43.2%		104 more per 1000 (from 86 fewer to 397 more)		
Relapse - Weekly IPT vs Monthly IPT (follow-up 24 months)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	23/43 (53.5%)	21/44 (47.7%)	RR 1.12 (0.74 to 1.7)	57 more per 1000 (from 124 fewer to 334 more)	⊕OOO VERY LOW	CRITICAL
								47.7%		57 more per 1000 (from 124 fewer to 334 more)		
Relapse - Bi-monthly IPT vs monthly IPT (follow-up 24 months)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	19/44 (43.2%)	21/44 (47.7%)	RR 0.9 (0.57 to 1.43)	48 fewer per 1000 (from 205 fewer to 205 more)	⊕OOO VERY LOW	CRITICAL
								47.7%		48 fewer per 1000 (from 205 fewer to 205 more)		

[1] ROB high or unclear across 1-2 domains
[2] 95% CI crosses one clinical decision threshold
[3] 95% CI crosses two clinical decision thresholds

Pharmacological interventions

Antidepressant versus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antidepressant	Placebo	Relative (95% CI)	Absolute		
Relapse- All												

48	randomised trials	very serious[1]	serious[2]	no serious indirectness	no serious imprecision	none	1505/4880 (30.8%)	2216/4225 (52.4%)	RR 0.59 (0.55 to 0.65)	215 fewer per 1000 (from 184 fewer to 236 fewer)	⊕OOO VERY LOW	CRITICAL
								52.4%		215 fewer per 1000 (from 183 fewer to 236 fewer)		

[1] ROB low in only one or two domains
[2] I2 >50% <80%

Antidepressant (full dose) versus antidepressant (half dose)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antidepressant (full dose)	Antidepressant (half dose)	Relative (95% CI)	Absolute		
Relapse												
3	randomised trials	very serious[1]	serious[2]	no serious indirectness	serious[3]	none	155/513 (30.2%)	190/511 (37.2%)	RR 0.81 (0.60 to 108)	71 fewer per 1000 (from 149 fewer to 1000 more)	⊕OOO VERY LOW	CRITICAL
TCA (full dose) versus TCA (half dose)												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[3]	none	110/385 (28.6%)	136/382 (35.6%)	RR 0.80 (0.65 to 0.99)	71 fewer per 1000 (from 4 fewer to 125 fewer)	⊕⊕OO LOW	CRITICAL
SSRI (full dose) versus SSRI (half dose)												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	45/128 (35.2%)	54/129 (41.9%)	RR 0.73 (0.35 to 1.54)	113 fewer per 1000 (from 272 fewer to 226 more)	⊕OOO VERY LOW	CRITICAL

[1] ROB high or unclear across multiple domains
[2] I2 >50%<80%
[3] 95% CI crosses one clinical decision threshold
[4] ROB unclear in several domains

Antidepressant versus lithium

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antidepressant	Lithium alone	Relative (95% CI)	Absolute		
Relapse - Amitriptyline vs lithium												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	32/57 (56.1%)	39/50 (78%)	RR 0.72 (0.55 to 0.95)	218 fewer per 1000 (from 39 fewer to 351 fewer)	⊕OOO VERY LOW	CRITICAL
								78%		218 fewer per 1000 (from 39 fewer to 351 fewer)		

[1] ROB high or unclear across multiple domains
[2] 95% CI crosses one clinical decision threshold

Lithium augmentation of antidepressants versus placebo augmentation

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Lithium augmentation + AD	Placebo + AD	Relative (95% CI)	Absolute		
Relapse												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	19/80 (23.8%)	33/80 (41.3%)	RR 0.62 (0.35 to 1.12)	157 fewer per 1000 (from 268 fewer to 49 more)	⊕⊕OO LOW	CRITICAL
								38.5%		146 fewer per 1000 (from 250 fewer to 46 more)		

[1] ROB high or unclear in several domains
[2] 95% CI crosses one clinical decision threshold

Risperidone augmentation of antidepressants versus placebo augmentation

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Risperidone augmentation + AD	Placebo + AD	Relative (95% CI)	Absolute		
Relapse												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	65/122 (53.3%)	65/119 (54.6%)	RR 0.98 (0.77 to 1.23)	11 fewer per 1000 (from 126 fewer to 126 more) 11 fewer per 1000 (from 126 fewer to 126 more)	⊕⊕OO LOW	CRITICAL
								54.6%				

[1] ROB unclear across several domains

[2] OIS not met (<300 events)

Antipsychotics versus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antipsychotics versus placebo	Control	Relative (95% CI)	Absolute		
Relapse - Quetiapine versus placebo												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	54/387 (14%)	127/384 (33.1%)	RR 0.42 (0.32 to 0.56)	192 fewer per 1000 (from 146 fewer to 225 fewer) 192 fewer per 1000 (from 146 fewer to 225 fewer)	⊕⊕OO LOW	CRITICAL
								33.1%				

[1] ROB high or unclear across multiple domains

Combination interventions

Combination psychological plus pharmacological versus pharmacological interventions

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Combination pharm + Psych	Pharm- 12 month	Relative (95% CI)	Absolute		
Imipramine + IPT vs Imipramine												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	4/25 (16%)	11/28 (39.3%)	RR 0.41 (0.15 to 1.12)	232 fewer per 1000 (from 334 fewer to 47 more)	⊕OOO VERY LOW	CRITICAL
								39.3%		232 fewer per 1000 (from 334 fewer to 47 more)		
MBCT + AD vs AD												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	17/33 (51.5%)	20/35 (57.1%)	RR 0.9 (0.58 to 1.4)	57 fewer per 1000 (from 240 fewer to 229 more)	⊕OOO VERY LOW	CRITICAL
								57.1%		57 fewer per 1000 (from 240 fewer to 228 more)		
Paroxetine + IPT vs paroxetine												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[3]	none	11/28 (39.3%)	16/35 (45.7%)	OR 0.86 (0.42 to 1.4)	37 fewer per 1000 (from 196 fewer to 84 more)	⊕OOO VERY LOW	CRITICAL
								45.7%		37 fewer per 1000 (from 196 fewer to 84 more)		
CBT vs AD vs AD alone												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	36/88 (40.9%)	42/89 (47.2%)	RR 0.86 (0.62 to 1.21)	66 fewer per 1000 (from 179 fewer to 99 more)	⊕OOO VERY LOW	CRITICAL

CT + fluoxetine versus fluoxetine alone												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	27/66 (40.9%)	29/66 (43.9%)	RR 0.93 (0.62 to 1.39)	31 fewer per 1000 (from 167 fewer to 171 more)	⊕OOO VERY LOW	CRITICAL
CBT + AD vs AD												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[3]	none	9/22 (40.9%)	13/23 (56.5%)	RR 0.72 (0.3 to 1.23)	158 fewer per 1000 (from 396 fewer to 130 more)	⊕OOO VERY LOW	CRITICAL
								56.5%		158 fewer per 1000 (from 396 fewer to 130 more)		

[1] ROB high or unclear across multiple domains
[2] 95% CI crosses one clinical decision threshold
[3] 95% CI crosses two clinical decision thresholds
[4] ROB high or unclear across 1-2 domains

Combination psychological plus pharmacological versus psychological interventions

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**COMBINATION PHARM + PSYCH**	**PSYCH-12 month**	**Relative (95% CI)**	**Absolute**		
CBT + fluoxetine vs CBT												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	4/11 (36.4%)	6/13 (46.2%)	RR 0.79 (0.3 to 2.09)	97 fewer per 1000 (from 323 fewer to 503 more)	⊕OOO VERY LOW	CRITICAL
								46.2%		97 fewer per 1000 (from 323 fewer to 504 more)		
IPT + Imipramine vs IPT												

1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[4]	none	4/25 (16%)	14/26 (53.8%)	RR 0.3 (0.11 to 0.78)	377 fewer per 1000 (from 118 fewer to 479 fewer)	⊕OOO VERY LOW	CRITICAL
								53.9%		377 fewer per 1000 (from 119 fewer to 480 fewer)		
IPT + nortriptyline vs IPT												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	11/29 (37.9%)	18/31 (58.1%)	RR 0.65 (0.38 to 1.14)	203 fewer per 1000 (from 360 fewer to 81 more)	⊕⊕OO LOW	CRITICAL
								58.1%		203 fewer per 1000 (from 360 fewer to 81 more)		
IPT + paroxetine vs IPT												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	11/28 (39.3%)	23/35 (65.7%)	RR 0.6 (0.36 to 1)	263 fewer per 1000 (from 421 fewer to 0 more)	⊕⊕OO LOW	CRITICAL
								65.7%		263 fewer per 1000 (from 420 fewer to 0 more)		
MBCT +mADM vs MBCT												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[4]	none	85/121 (70.2%)	105/128 (82%)	RR 0.86 (0.74 to 0.99)	115 fewer per 1000 (from 8 fewer to 213 fewer)	⊕OOO VERY LOW	CRITICAL
								82%		115 fewer per 1000 (from 8 fewer to 213 fewer)		

[1] ROB high or unclear across 1-2 domains
[2] 95% CI crosses two clinical decision thresholds
[3] ROB high or unclear aoss multiple domains
[4] 95% CI crosses one clinical decision threshold

Antidepressants plus antipsychotics versus antidepressants plus placebo

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	ADM + antipsychotic versus ADM + placebo	Control	Relative (95% CI)	Absolute		
Relapses - TCA + antipsychotic versus TCA + placebo												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	5/15 (33.3%)	2/13 (15.4%)	RR 2.17 (0.5 to 9.35)	180 more per 1000 (from 77 fewer to 1000 more)	⊕OOO VERY LOW	CRITICAL
								15.4%		180 more per 1000 (from 77 fewer to 1000 more)		

[1] Unclear ROB in most domains

[2] 95% CI crosses two clinical decision thresholds

ECT plus antidepressants versus antidepressants (with or without lithium augmentation)

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	ECT + ADM versus ADM (+/- Li)	Control	Relative (95% CI)	Absolute	Quality	Importance
Relapses												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	4/27 (14.8%)	6/27 (22.2%)	RR 0.65 (0.22 to 1.91)	78 fewer per 1000 (from 173 fewer to 202 more)	⊕OOO VERY LOW	CRITICAL
								25.9%		91 fewer per 1000 (from 202 fewer to 236 more)		
Relapses - ECT + TCA versus TCA												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	1/16 (6.3%)	2/17 (11.8%)	RR 0.53 (0.05 to 5.31)	55 fewer per 1000 (from 112 fewer to 507 more)		CRITICAL

								11.8%		55 fewer per 1000 (from 112 fewer to 509 more)	⊕OOO VERY LOW	
Relapses - ECT + ADM versus ADM (+/- Li augmentation)													
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	3/11 (27.3%)	4/10 (40%)	RR 0.68 (0.2 to 2.33)	128 fewer per 1000 (from 320 fewer to 532 more)	⊕OOO VERY LOW	CRITICAL
								40%		128 fewer per 1000 (from 320 fewer to 532 more)		

[1] High ROB in one domain and unclear in several others
[2] 95% CI crosses two clinical decision thresholds

Light therapy

Which therapy is most effective for relapse prevention of depression with a seasonal pattern/SAD?

Quality assessment							Summary of findings					Importance
							No of patients		Effect		Quality	
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Relapse Prevention	Control	Relative (95% CI)	Absolute		
Leaving study early for any reason - Bright white light visor vs no treatment control												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[1]	none	4/18 (22.2%)	1/10 (10%)	RR 2.22 (0.29 to 17.27)	12 more per 100 (from 7 fewer to 163 more)	⊕⊕OO LOW	
								10%		12 more per 100 (from 7 fewer to 163 more)		
Leaving study early for any reason - Bright white light visor vs dim red light visor												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	very serious[1]	none	4/18 (22.2%)	3/18 (16.7%)	RR 1.33 (0.35 to 5.13)	6 more per 100 (from 11 fewer to 69 more)	⊕⊕OO LOW	
								16.7%		6 more per 100 (from 11 fewer to 69 more)		
Relapse during course of study (BDI>=13 for 2 consecutive wks) - Bright white light visor vs no treatment control												

1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	9/18 (50%)	8/10 (80%) 80%	RR 0.63 (0.36 to 1.09)	30 fewer per 100 (from 51 fewer to 7 more) 30 fewer per 100 (from 51 fewer to 7 more)	⊕⊕⊕O MODERATE	
Relapse during course of study (BDI>=13 for 2 consecutive wks) - Bright white light visor vs dim red light visor												
1	randomised trials	no serious limitations	no serious inconsistency	no serious indirectness	serious[2]	none	9/18 (50%)	4/18 (22.2%) 22.2%	RR 2.25 (0.84 to 5.99)	28 more per 100 (from 4 fewer to 111 more) 28 more per 100 (from 4 fewer to 111 more)	⊕⊕⊕O MODERATE	

[1] Inconclusive effect size; single study
[2] Single study

Non-light therapy

Is relapse prevention effective for depression with a seasonal pattern/SAD? (Buspirone versus placebo)

Quality assessment							Summary of findings					
							No. of patients		Effect		Quality	Importance
No. of studies	**Design**	**Limitations**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Buspirone-prevention**	**Placebo**	**Relative (95% CI)**	**Absolute**		
Relapse Prevention - Number of patients experiencing a recurrence												
3	randomised trial	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	92/542 (17%)	153/519 (29.5%) 31.9%	RR 0.58 (0.46 to 0.72)	12 fewer per 100 (from 8 fewer to - 16 fewer) 13 fewer per 100	⊕⊕⊕⊕ HIGH	

Access to services (chapter 12)

Telephone administered psychological interventions versus usual care

Clinic based telepsychiatry using a video-webcam versus usual care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Clinic-based telepsychiatry using a video Webcam versus TAU	Control	Relative (95% CI)	Absolute		
Number of subjects who made a mental health appointment (follow-up mean 6 months; assessed with: Not reported)												
1	randomised trials	serious[1]	no serious inconsistency	serious[2]	serious[3]	none	77/80 (96.3%)	29/87 (33.3%)	RR 2.89 (2.14 to 3.9)	630 more per 1000 (from 380 more to 967 more)	⊕OOO VERY LOW	
								33.3%		629 more per 1000 (from 380 more to 966 more)		
Number of subjects who made a primary care appointment (follow-up mean 6 months; assessed with: Not reported)												
1	randomised trials	serious[1]	no serious inconsistency	serious[2]	serious[3]	none	56/80 (70%)	76/87 (87.4%)	RR 0.8 (0.68 to 0.94)	175 fewer per 1000 (from 52 fewer to 280 fewer)	⊕OOO VERY LOW	
								87.4%		175 fewer per 1000 (from 52 fewer to 280 fewer)		
Number used antidepressants (follow-up mean 6 months; assessed with: Not reported)												
1	randomised trials	serious[1]	no serious inconsistency	serious[2]	serious[3]	none	56/80 (70%)	40/87 (46%)	RR 1.52 (1.16 to 1.99)	239 more per 1000 (from 74 more to 455 more)		

								46%		239 more per 1000 (from 74 more to 455 more)	⊕OOO VERY LOW	
Mean number of completed mental health appointments (follow-up mean 6 months; measured with: Not reported; Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	serious[2]	serious[4]	none	77	29	-	MD 0.5 higher (0.94 lower to 1.94 higher)	⊕OOO VERY LOW	
Mean number of completed primary care appointments (follow-up mean 6 months; measured with: Not reported; Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	serious[2]	serious[5]	none	56	76	-	MD 0 higher (1.17 lower to 1.17 higher)	⊕OOO VERY LOW	
Satisfaction (follow-up mean 6 months; measured with: Visit Specific Satisfaction Questionnaire (VSQ-9); range of scores: 0-36; Better indicated by higher values)												
1	randomised trials	serious[6]	no serious inconsistency	serious[2]	serious[5]	none	80	87	-	MD 0.2 higher (0.16 lower to 0.56 higher)	⊕OOO VERY LOW	

[1] Unclear blinding of outcome assessment
[2] US study with potential applicability issues
[3] Events<300
[4] 95% CI crosses both line of no effect and threshold for clinically significant benefit (SMD 0.5)
[5] N<400
[6] Non-blind outcome assessment (self-report)

Telephone CBT versus enhanced usual care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Telephone CBT**	**Enhanced usual care**	**Relative (95% CI)**	**Absolute**		
Number reporting they were staisfied with the treatment provided												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	24/64 (37.5%)	12/33 (36.4%)	RR 1.03 (0.59 to 1.79)	11 more per 1000 (from 149 fewer to 287 more)	⊕OOO VERY LOW	CRITICAL

[1] High ROB in one domain and unclear ROB in two others
[2] 95% CI crosses two clinical decision thresholds

Telephone-administered monitoring interventions versus usual care

Telephone disease management versus usual care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Telephone disease management versus usual care	Control	Relative (95% CI)	Absolute		
Number completing at least one mental health/substance abuse appointment (follow-up mean 4 months; assessed with: Self-report)												
1	randomised trials	serious[1]	no serious inconsistency	serious[2]	serious[2,3]	none	19/46 (41.3%)	5/51 (9.8%)	RR 4.21 (1.71 to 10.37)	315 more per 1000 (from 70 more to 919 more)	⊕OOO VERY LOW	
								9.8%		315 more per 1000 (from 70 more to 918 more)		

[1] Non-blind outcome assessment (self-report)
[2] US study with potential applicability issues and veteran population so may not be applicable to all men
[3] Events<300

Close monitoring versus usual care

Quality assessment							No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Close monitoring versus usual care	Control	Relative (95% CI)	Absolute		
Number attending primary care visits during study period (follow-up mean 6 months; assessed with: Case review)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	serious[3]	none	92/130 (70.8%)	62/93 (66.7%)	RR 1.06 (0.89 to 1.27)	40 more per 1000 (from 73 fewer to 180 more)	⊕OOO VERY LOW	
								66.7%		40 more per 1000 (from 73 fewer to 180 more)		
Number who had any MH care (including behavioral health specialist) during the study period (follow-up mean 6 months; assessed with: Case review)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	serious[4]	none	43/130 (33.1%)	6/93 (6.5%)	RR 5.13 (2.28 to 11.54)	266 more per 1000 (from 83 more to 680 more)	⊕OOO VERY LOW	
								6.5%		268 more per 1000 (from 83 more to 685 more)		
Number who started an antidepressant during the study period (follow-up mean 6 months; assessed with: Case review)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	serious[3]	none	21/130 (16.2%)	9/93 (9.7%)	RR 1.67 (0.8 to 3.48)	65 more per 1000 (from 19 fewer to 240 more)	⊕OOO VERY LOW	
								9.7%		65 more per 1000 (from 19 fewer to 241 more)		

[1] Outcome assessment was non-blind and there were statistically significant baseline differences between groups (more males, more financial troubles, more subjects with trauma exposure, more with a past history of depression and more with a GAD diagnosis in the intervention group)
[2] US study with potential applicability issues and veteran population so may not be applicable to all men
[3] 95% CI crosses both line of no effect and threshold for clinically significant benefit (RR 1.25)
[4] Events<300

Simple collaborative care versus usual care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Simple collaborative care versus usual care	Control	Relative (95% CI)	Absolute		
Number who attended ≥1 appointment with mental health specialist (follow-up mean 12 months; assessed with: Database review)												

2	randomised trials	serious[1]	serious[2]	serious[3]	serious[4]	none	138/357 (38.7%)	120/372 (32.3%) 32.3%	RR 1.2 (0.77 to 1.86)	65 more per 1000 (from 74 fewer to 277 more) 65 more per 1000 (from 74 fewer to 278 more)	⊕OOO VERY LOW	
Number who have had a depression-related primary care visit (follow-up mean 12 months; assessed with: Database review)												
1	randomised trials	serious[1]	no serious inconsistency	serious[3]	serious[5]	none	141/168 (83.9%)	106/186 (57%) 57%	RR 1.47 (1.28 to 1.7)	268 more per 1000 (from 160 more to 399 more) 268 more per 1000 (from 160 more to 399 more)	⊕OOO VERY LOW	
Number of patients whose unhelpful medications (those potentially exacerbating depression) were terminated												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[7]	none	23/100 (23%)	17/75 (22.7%)	RR 1.01 (0.58 to 1.76)	2 more per 1000 (from 95 fewer to 172 more)	⊕OOO VERY LOW	CRITICAL
Received ≥ 90 days of therapy with a minimally therapeutic dosage of antidepressant (follow-up mean 12 months; assessed with: Database review)												
2	randomised trials	serious[1]	serious[2]	serious[3]	serious[4]	none	224/324 (69.1%)	182/301 (60.5%) 61%	RR 1.13 (0.95 to 1.35)	79 more per 1000 (from 30 fewer to 212 more) 79 more per 1000 (from 31 fewer to 214 more)	⊕OOO VERY LOW	
Number of adults starting an antidepressant												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	serious[5]	none	26/100 (26%)	6/75 (8%)	RR 3.25 (1.41 to 7.5)	180 more per 1000 (from 33 more to 520 more)	⊕⊕OO LOW	CRITICAL
Number of patients for whom a psychiatric consultation was sought												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[7]	none	12/100 (12%)	11/75 (14.7%)	RR 0.82 (0.38 to 1.75)	26 fewer per 1000 (from 91 fewer to 110 more)	⊕OOO VERY LOW	CRITICAL

[1] Statistically significant group differences at baseline in Hedrick 2003 (more subjects with previous depression in intervention group)
[2] I-squared > 50%
[3] US study with potential applicability issues and veteran population so may not be applicable to all men
[4] 95% CI crosses both line of no effect and threshold for clinically significant benefit (RR 1.25)
[5] Events<300
[6] Unclear ROB in multiple domains
[7] 95% CI crosses two clinical decision thresholds

Co-located versus geographically separate services

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Co-located services	Geographically separate services	Relative (95% CI)	Absolute		
Number of patient who engaged with treatment												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	481/640 (75.2%)	338/657 (51.4%)	RR 1.46 (1.34 to 1.59)	237 more per 1000 (from 175 more to 304 more)	⊕⊕⊕O MODERATE	CRITICAL
Number of treatment visits (Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	687	703	-	MD 1.28 higher (0.87 to 1.69 higher)	⊕⊕OO LOW	CRITICAL

[1] Unclear ROB in multiple domains
[2] 95% CI crosses one clinical decision threshold

Culturally-adapted psychological interventions versus usual care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Culturally-adapted CBT	TAU	Relative (95% CI)	Absolute		

Number of participants stating that they were 'very satisfied' with treatment												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	50/69 (72.5%)	32/68 (47.1%)	RR 1.54 (1.15 to 2.06)	254 more per 1000 (from 71 more to 499 more)	⊕OOO VERY LOW	CRITICAL

[1] High ROB in multiple domains
[2] 95% CI crosses one clinical decision threshold