APPENDIX 22: CLINICAL AND ECONOMIC EVIDENCE PROFILES

1 CLINICAL EVIDENCE PROFILES

1.1 PSYCHOSOCIAL INTERVENTIONS: PREVENTION (RISK FACTORS IDENTIFIED)

1.1.1 Depression: post-miscarriage self-help versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With depression: post-miscarriage self-help versus TAU**		**Risk with control**	**Risk difference with depression: post-miscarriage self-help versus TAU** (95% CI)
Depression mean symptoms post-treatment – ITT analysis (at-risk populations) (measured with: Brief Symptom Inventory (BSI): Depression; better indicated by lower values)											
228 (1 study) 5 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	113	115	-		The mean depression mean symptoms post-treatment – ITT analysis (at-risk populations) in the intervention groups was **0.64 standard deviations lower** (0.91 to 0.37 lower)

[1] Risk of bias due to statistically significant group differences at baseline
[2] Total population size is less than 400 (a threshold rule-of-thumb)

1.1.2 Depression: social support versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With depression: social support versus TAU**		**Risk with control**	**Risk difference with depression: social support versus TAU** (95% CI)
Depression diagnosis post-treatment – ITT analysis (at-risk populations) (assessed with: schedules for Clinical Assessment in Neuropsychiatry (SCAN))											
117 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	40/56 (71.4%)	37/61 (60.7%)	**RR 0.85** (0.65 to 1.1)	**Study population**	
										714 per 1000	**107 fewer per 1000** (from 250 fewer to 71 more)
										Moderate	
										714 per 1000	**107 fewer per 1000** (from 250 fewer to 71 more)
Depression diagnosis post-treatment – available case analysis (at-risk populations) (assessed with: schedules for Clinical Assessment in Neuropsychiatry (SCAN))											
65 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of	19/35 (54.3%)	6/30 (20%)	**RR 0.37** (0.17 to 0.8)	**Study population**	
										543 per 1000	**342 fewer per 1000** (from 109 fewer to 451 fewer)

						bias, imprecision			**Moderate**	
									543 per 1000	**342 fewer per 1000** (from 109 fewer to 451 fewer)

[1] Risk of bias due to non-blind outcome assessment

[2] Total number of events is less than 300 (a threshold rule-of-thumb)

[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.3 Depression: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With depression: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU**		**Risk with control**	**Risk difference with depression: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU** (95% CI)
Depression diagnosis post-treatment – ITT analysis (at-risk populations) (assessed with: schedules for Clinical Assessment in Neuropsychiatry (SCAN) or Structured Clinical Interview (SCID) or Structured Clinical Interview for Childhood Diagnoses (KID-SCID))											
360 (3 studies) 27 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	41/179 (22.9%)	29/181 (16%)	**RR 0.69** (0.45 to 1.05)	**Study population**	
										229 per 1000	**71 fewer per 1000** (from 126 fewer to 11 more)
										Moderate	
										333 per 1000	**103 fewer per 1000** (from 183 fewer to 17 more)

Depression diagnosis post-treatment – available case analysis (at-risk populations) (assessed with: schedules for Clinical Assessment in Neuropsychiatry (SCAN) or Structured Clinical Interview (SCID) or Structured Clinical Interview for Childhood Diagnoses (KID-SCID))										
320 (3 studies) 27 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	21/159 (13.2%) 9/161 (5.6%)	**RR 0.48** (0.23 to 1.01)	**Study population**	
									132 per 1000	**69 fewer per 1000** (from 102 fewer to 1 more)
									Moderate	
									227 per 1000	**118 fewer per 1000** (from 175 fewer to 2 more)
Depression symptomatology post-treatment – ITT analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥11/12)										
254 (2 studies) 27 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	38/127 (29.9%) 33/127 (26%)	**RR 0.85** (0.58 to 1.25)	**Study population**	
									299 per 1000	**45 fewer per 1000** (from 126 fewer to 75 more)
									Moderate	
									370 per 1000	**55 fewer per 1000** (from 155 fewer to 93 more)
Depression symptomatology post-treatment – available case analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥11/12)										
221 (2 studies) 27 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	20/109 (18.3%) 18/112 (16.1%)	**RR 0.88** (0.49 to 1.57)	**Study population**	
									183 per 1000	**22 fewer per 1000** (from 94 fewer to 105 more)
									Moderate	

										171 per 1000	**21 fewer per 1000** (from 87 fewer to 97 more)
Depression mean scores post-treatment – available case analysis (at-risk populations) (measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)											
33 (1 study)	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	18	**-**		The mean depression mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.06 standard deviations lower** (0.75 lower to 0.62 higher)
Depression diagnosis intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: structured Clinical Interview (SCID))											
45 (1 study) 20 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	8/21 (38.1%)	7/24 (29.2%)	**RR 0.77** (0.33 to 1.75)	**Study population**	
										381 per 1000	**88 fewer per 1000** (from 255 fewer to 286 more)
										Moderate	
										381 per 1000	**88 fewer per 1000** (from 255 fewer to 286 more)
Depression diagnosis intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: structured Clinical Interview (SCID))											
37 (1 study) 20 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	4/17 (23.5%)	3/20 (15%)	**RR 0.64** (0.17 to 2.46)	**Study population**	
										235 per 1000	**85 fewer per 1000** (from 195 fewer to 344 more)
										Moderate	

										235 per 1000	**85 fewer per 1000** (from 195 fewer to 343 more)
Depression symptomatology intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) >12)											
45 (1 study) 20 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	9/21 (42.9%)	12/24 (50%)	**RR 1.17** (0.62 to 2.2)	**Study population**	
										429 per 1000	**73 more per 1000** (from 163 fewer to 514 more)
										Moderate	
										429 per 1000	**73 more per 1000** (from 163 fewer to 515 more)
Depression symptomatology intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) >12)											
30 (1 study) 20 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	3/15 (20%)	3/15 (20%)	**RR 1** (0.24 to 4.18)	**Study population**	
										200 per 1000	**0 fewer per 1000** (from 152 fewer to 636 more)
										Moderate	
										200 per 1000	**0 fewer per 1000** (from 152 fewer to 636 more)
Depression mean scores intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)											
30 (1 study) 20 weeks	no serious	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2]	17	13	-		The mean depression mean scores intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk

	risk of bias					due to imprecision			populations) in the intervention groups was **0.02 standard deviations lower** (0.74 lower to 0.7 higher)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.4 Depression: psychoeducational booklet versus TAU or enhanced TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**
							With control	**With depression: psychoeducational booklet versus TAU or enhanced TAU**		**Risk with control** / **Risk difference with depression: psychoeducational booklet versus TAU or enhanced TAU** (95% CI)
Depression symptomatology post-treatment – ITT analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10/12)										
1140 (2 studies) 3 weeks	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⨁⨁⨁⊖ **MODERATE**[1] due to risk of bias	239/571 (41.9%)	216/569 (38%)	**RR 0.9** (0.79 to 1.03)	**Study population** **419 per 1000** / **42 fewer per 1000** (from 88 fewer to 13 more) **Moderate** **409 per 1000** / **41 fewer per 1000** (from 86 fewer to 12 more)
Depression symptomatology post-treatment – available case analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10/12)										

838 (2 studies) 3 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	87/419 (20.8%)	66/419 (15.8%)	**RR 0.73** (0.51 to 1.06)	**Study population**	
										208 per 1000	**56 fewer per 1000** (from 102 fewer to 12 more)
										Moderate	
										218 per 1000	**59 fewer per 1000** (from 107 fewer to 13 more)
Depression symptomatology short follow-up (9-16 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10)											
540 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	60/270 (22.2%)	53/270 (19.6%)	**RR 0.88** (0.64 to 1.23)	**Study population**	
										222 per 1000	**27 fewer per 1000** (from 80 fewer to 51 more)
										Moderate	
										222 per 1000	**27 fewer per 1000** (from 80 fewer to 51 more)
Depression symptomatology short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10)											
479 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊕⊖⊖ **LOW**[2,3] due to imprecision	32/242 (13.2%)	20/237 (8.4%)	**RR 0.64** (0.38 to 1.08)	**Study population**	
										132 per 1000	**48 fewer per 1000** (from 82 fewer to 11 more)
										Moderate	
										132 per 1000	**48 fewer per 1000** (from 82 fewer to 11 more)

Depression symptomatology intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10)										
540 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⨁⨀⨀ **LOW**[2,3] due to imprecision	90/270 (33.3%) 75/270 (27.8%)	**RR 0.83** (0.65 to 1.08)	**Study population**	
									333 per 1000	**57 fewer per 1000** (from 117 fewer to 27 more)
									Moderate	
									333 per 1000	**57 fewer per 1000** (from 117 fewer to 27 more)
Depression symptomatology intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10)										
423 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⨁⨀⨀ **LOW**[2,3] due to imprecision	29/209 (13.9%) 19/214 (8.9%)	**RR 0.64** (0.37 to 1.1)	**Study population**	
									139 per 1000	**50 fewer per 1000** (from 87 fewer to 14 more)
									Moderate	
									139 per 1000	**50 fewer per 1000** (from 88 fewer to 14 more)

[1] Risk of bias due to statistically significant group differences at baseline
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.5 Depression: non-mental health-focused education and support versus TAU or enhanced TAU

Quality assessment	Summary of findings

Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With depression: non-mental health-focused education and support versus TAU or enhanced TAU		Risk with control	Risk difference with depression: non-mental health-focused education and support versus TAU or enhanced TAU (95% CI)
Depression symptomatology post-treatment – ITT analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) >12)											
306 (2 studies) 6-13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	49/153 (32%)	34/153 (22.2%)	**RR 0.7** (0.44 to 1.14)	**Study population**	
										320 per 1000	**96 fewer per 1000** (from 179 fewer to 45 more)
										Moderate	
										316 per 1000	**95 fewer per 1000** (from 177 fewer to 44 more)
Depression symptomatology post-treatment – available case analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) >12)											
261 (2 studies) 6-13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	24/128 (18.8%)	14/133 (10.5%)	**RR 0.57** (0.31 to 1.05)	**Study population**	
										188 per 1000	**81 fewer per 1000** (from 129 fewer to 9 more)
										Moderate	
										188 per 1000	**81 fewer per 1000** (from 130 fewer to 9 more)

Depression mean scores post-treatment – ITT analysis (at-risk populations) (measured with: Center for Epidemiologic Studies Depression Scale (CES-D); better indicated by lower values)										
275 (1 study) 28 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	reporting bias strongly suspected[4]	⊕⊕⊖⊖ **LOW**[3,4] due to imprecision, publication bias	137	138	-	The mean depression mean scores post-treatment – ITT analysis (at-risk populations) in the intervention groups was **0.13 standard deviations lower** (0.37 lower to 0.1 higher)
Depression mean scores post-treatment – available case analysis (at-risk populations) (measured with: Beck Depression Inventory (BDI) or Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)										
370 (2 studies)	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊕⊖ **MODERATE**[3] due to imprecision	169	201	-	The mean depression mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.14 standard deviations lower** (0.34 lower to 0.07 higher)
Depression symptomatology short follow-up (9-16 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) >12)										
162 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	33/82 (40.2%)	22/80 (27.5%)	**RR 0.68** (0.44 to 1.06)	**Study population** **402 per 1000** | **129 fewer per 1000** (from 225 fewer to 24 more) **Moderate** **402 per 1000** | **129 fewer per 1000** (from 225 fewer to 24 more)

Depression symptomatology short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) – non-mental health-focused education and support (assessed with: Edinburgh postnatal Depression Scale (EPDS) >12)										
128 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	14/63 (22.2%)	7/65 (10.8%)	**RR 0.48** (0.21 to 1.12)	**Study population** **222 per 1000** **116 fewer per 1000** (from 176 fewer to 27 more) **Moderate** **222 per 1000** **115 fewer per 1000** (from 175 fewer to 27 more)
Depression mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)										
128 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊕⊖⊖ **LOW**[2,3] due to imprecision	63	65	-	The mean depression mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.21 standard deviations lower** (0.56 lower to 0.13 higher)
Depression symptomatology intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) >12)										
306 (2 studies) 20-24 weeks	no serious risk of bias	very serious[5]	no serious indirectness	very serious[1,2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,5] due to inconsistency, imprecision	45/153 (29.4%)	40/153 (26.1%)	**RR 0.91** (0.44 to 1.89)	**Study population** **294 per 1000** **26 fewer per 1000** (from 165 fewer to 262 more) **Moderate**

									290 per 1000	**26 fewer per 1000** (from 162 fewer to 258 more)
Depression symptomatology intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) >12)										
254 (2 studies) 20-24 weeks	no serious risk of bias	very serious[5]	no serious indirectness	very serious[1,2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,5] due to inconsistency, imprecision	18/126 (14.3%)	15/128 (11.7%)	**RR 0.84** (0.27 to 2.63)	**Study population**
									143 per 1000	**23 fewer per 1000** (from 104 fewer to 233 more)
									Moderate	
									142 per 1000	**23 fewer per 1000** (from 104 fewer to 231 more)
Depression mean scores intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)										
133 (1 study) 24 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊕⊖⊖ **LOW**[2,3] due to imprecision	65	68	**-**	The mean depression mean scores intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.3 standard deviations lower** (0.64 lower to 0.04 higher)
Depression symptomatology long follow-up (25-103 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) >12)										
									Study population	

162 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	34/82 (41.5%)	28/80 (35%)	**RR 0.84** (0.57 to 1.25)	**415 per 1000**	**66 fewer per 1000** (from 178 fewer to 104 more)
										Moderate	
										415 per 1000	**66 fewer per 1000** (from 178 fewer to 104 more)
Depression symptomatology long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) >12)											
123 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	12/60 (20%)	11/63 (17.5%)	**RR 0.87** (0.42 to 1.83)	**Study population**	
										200 per 1000	**26 fewer per 1000** (from 116 fewer to 166 more)
										Moderate	
										200 per 1000	**26 fewer per 1000** (from 116 fewer to 166 more)
Depression mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)											
123 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3] due to imprecision	60	63	-		The mean depression mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.08 standard deviations lower** (0.44 lower to 0.27 higher)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Total population size is less than 400 (a threshold rule-of-thumb)

[4] Paper omits data
[5] There is evidence of substantial heterogeneity of study effect sizes

1.1.6 Depression: home visits versus TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With depression: home visits versus TAU**		**Risk with control**	**Risk difference with depression: home visits versus TAU** (95% CI)
Depression symptomatology post-treatment – ITT analysis (at-risk populations) (assessed with: Center for Epidemiological Studies Depression Scale (CES-D) ≥21 or Hospital Anxiety and Depression Scale – Depression (HADS >7))											
204 (2 studies) 52-117 weeks	very serious[1]	very serious[2]	no serious indirectness	very serious[3,4]	reporting bias strongly suspected[5]	⨁⊖⊖⊖ **VERY LOW**[1,2,3,4,5] due to risk of bias, inconsistency, imprecision, publication bias	43/99 (43.4%)	42/105 (40%)	**RR 0.94** (0.45 to 1.96)	**Study population**	
										434 per 1000	**26 fewer per 1000** (from 239 fewer to 417 more)
										Moderate	
										429 per 1000	**26 fewer per 1000** (from 236 fewer to 412 more)
Depression symptomatology post-treatment – available case analysis (at-risk populations) (assessed with: Center for Epidemiological Studies Depression Scale (CES-D) ≥16/21 or Hospital Anxiety and Depression Scale – Depression (HADS >7))											
684 (3 studies)	very serious[1]	serious[6]	no serious indirectness	very serious[3,4]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,3,4,6] due to risk of bias,	97/292 (33.2%)	103/392 (26.3%)	**RR 0.78** (0.44 to 1.41)	**Study population**	
										332 per 1000	**73 fewer per 1000** (from 186 fewer to 136 more)

52-117 weeks						inconsistency, imprecision				**Moderate**	
										256 per 1000	**56 fewer per 1000** (from 143 fewer to 105 more)
Depression mean scores post-treatment – available case analysis (at-risk populations) (measured with: Center for Epidemiologic Studies Depression Scale (CES-D) or Hospital Anxiety and Depression Scale – Depression; better indicated by lower values)											
621 (2 studies) 52 weeks	very serious[1]	very serious[7]	no serious indirectness	no serious imprecision	undetected	⨁⊖⊖⊖ **VERY LOW**[1,7] due to risk of bias, inconsistency	260	361	-		The mean depression mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.38 standard deviations lower** (0.75 to 0.01 lower)
Depression symptomatology Very long follow-up (>104 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Hospital Anxiety and Depression Scale – Depression (HADS ≥8))											
120 (1 study) 104 weeks	very serious[1]	no serious inconsistency	no serious indirectness	very serious[3,4]	reporting bias strongly suspected[5]	⨁⊖⊖⊖ **VERY LOW**[1,3,4,5] due to risk of bias, imprecision, publication bias	27/59 (45.8%)	25/61 (41%)	**RR 0.90** (0.59 to 1.35)	**Study population**	
										458 per 1000	**46 fewer per 1000** (from 188 fewer to 160 more)
										Moderate	
										158 per 1000	**16 fewer per 1000** (from 65 fewer to 55 more)
Depression symptomatology Very long follow-up (>104 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Hospital Anxiety and Depression Scale – Depression (HADS ≥8))											
77 (1 study) 104 weeks	very serious[1]	no serious inconsistency	no serious indirectness	very serious[3,4]	reporting bias strongly suspected[5]	⨁⊖⊖⊖ **VERY LOW**[1,3,4,5] due to risk of bias,	6/38 (15.8%)	3/39 (7.7%)	**RR 0.49** (0.13 to 1.81)	**Study population**	
										158 per 1000	**81 fewer per 1000** (from 137 fewer to 128 more)

						imprecision, publication bias			**Moderate**	
									158 per 1000	**81 fewer per 1000** (from 137 fewer to 128 more)
Depression mean scores Very long follow-up (>104 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Hospital Anxiety and Depression Scale – Depression; better indicated by lower values)										
77 (1 study) 104 weeks	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4,8]	reporting bias strongly suspected[5]	⨁⊖⊖⊖ **VERY LOW**[1,4,5,8] due to risk of bias, imprecision, publication bias	38	39	-	The mean depression mean scores very long follow-up (>104 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.37 standard deviations lower** (0.82 lower to 0.08 higher)

[1] Risk of bias due to statistically significant group differences at baseline
[2] There is evidence of considerable heterogeneity of study effect sizes
[3] Total number of events is less than 300 (a threshold rule-of-thumb)
[4] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[5] Paper omits data
[6] There is evidence of moderate heterogeneity of study effect sizes
[7] There is evidence of substantial heterogeneity of study effect sizes
[8] Total population size is less than 400 (a threshold rule-of-thumb)

1.1.7 Depression: post-delivery discussion versus enhanced TAU

Quality assessment							**Summary of findings**			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**
							With control	**With depression: post-delivery**		**Risk with control**	**Risk difference with depression: post-delivery**

								discussion versus enhanced TAU			**discussion versus enhanced TAU** (95% CI)
Depression symptomatology post-treatment – ITT analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥13)											
1041 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⨁⨁⨁⨀ **MODERATE**[1] due to imprecision	137/521 (26.3%)	134/520 (25.8%)	**RR 0.98** (0.8 to 1.2)	**Study population**	
										263 per 1000	**5 fewer per 1000** (from 53 fewer to 53 more)
										Moderate	
										263 per 1000	**5 fewer per 1000** (from 53 fewer to 53 more)
Depression symptomatology post-treatment – available case analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥13)											
916 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	65/449 (14.5%)	81/467 (17.3%)	**RR 1.2** (0.89 to 1.62)	**Study population**	
										145 per 1000	**29 more per 1000** (from 16 fewer to 90 more)
										Moderate	
										145 per 1000	**29 more per 1000** (from 16 fewer to 90 more)
Depression mean scores post-treatment – available case analysis (at-risk populations) (measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)											
916 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⨁⨁⨁⨁ **HIGH**	449	467	-		The mean depression mean scores post-treatment – available case analysis (at-risk populations) in the intervention

										groups was **0.08 standard deviations higher** (0.05 lower to 0.21 higher)
Depression symptomatology Very long follow-up (>104 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥13)										
1041 (1 study) 208-312 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	296/521 (56.8%) 298/520 (57.3%)	**RR 1.01** (0.91 to 1.12)	**Study population**	
									568 per 1000	**6 more per 1000** (from 51 fewer to 68 more)
									Moderate	
									568 per 1000	**6 more per 1000** (from 51 fewer to 68 more)
Depression symptomatology Very long follow-up (>104 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥13)										
534 (1 study) 208-312 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	45/270 (16.7%) 42/264 (15.9%)	**RR 0.95** (0.65 to 1.4)	**Study population**	
									167 per 1000	**8 fewer per 1000** (from 58 fewer to 67 more)
									Moderate	
									167 per 1000	**8 fewer per 1000** (from 58 fewer to 67 more)
Depression mean scores Very long follow-up (>104 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)										
534 (1 study)	no serious	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	270 264	-		The mean depression mean scores very long follow-up (>104 weeks post-intervention) –

208-312 weeks	risk of bias									available case analysis (at-risk populations) in the intervention groups was **0.08 standard deviations lower** (0.25 lower to 0.09 higher)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.8 Depression: mother–infant relationship interventions versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With depression: mother–infant relationship interventions versus TAU**		**Risk with control**	**Risk difference with depression: mother–infant relationship interventions versus TAU** (95% CI)
Depression diagnosis post-treatment – ITT analysis (at-risk populations) (assessed with: structured Clinical Interview (SCID))											
449 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	74/229 (32.3%)	71/220 (32.3%)	**RR 1** (0.76 to 1.31)	**Study population**	
										323 per 1000	**0 fewer per 1000** (from 78 fewer to 100 more)
										Moderate	
										323 per 1000	**0 fewer per 1000** (from 78 fewer to 100 more)
Depression diagnosis post-treatment – available case analysis (at-risk populations) (assessed with: structured Clinical Interview (SCID))											

354 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	29/184 (15.8%)	21/170 (12.4%)	**RR 0.78** (0.47 to 1.32)	**Study population**	
										158 per 1000	**35 fewer per 1000** (from 84 fewer to 50 more)
										Moderate	
										158 per 1000	**35 fewer per 1000** (from 84 fewer to 51 more)
Depression symptomatology post-treatment – ITT analysis (at-risk populations) (assessed with: Center for Epidemiologic Studies Depression Scale (CES-D) ≥16)											
106 (1 study) 27 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to risk of bias, imprecision	10/50 (20%)	17/56 (30.4%)	**RR 1.52** (0.77 to 3)	**Study population**	
										200 per 1000	**104 more per 1000** (from 46 fewer to 400 more)
										Moderate	
										200 per 1000	**104 more per 1000** (from 46 fewer to 400 more)
Depression symptomatology post-treatment – available case analysis (at-risk populations) (assessed with: Center for Epidemiologic Studies Depression Scale (CES-D) ≥16)											
87 (1 study) 27 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to risk of bias, imprecision	2/42 (4.8%)	6/45 (13.3%)	**RR 2.8** (0.6 to 13.11)	**Study population**	
										48 per 1000	**86 more per 1000** (from 19 fewer to 577 more)
										Moderate	
										48 per 1000	**86 more per 1000** (from 19 fewer to 581 more)

Depression mean scores post-treatment – available case analysis (at-risk populations) (measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)										
417 (2 studies) 15-26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	215	202	-	The mean depression mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.22 standard deviations lower** (0.41 to 0.02 lower)
Depression mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)										
63 (1 study) 28 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,4]	undetected	⊕⊕⊖⊖ **LOW**[2,4] due to imprecision	31	32	-	The mean depression mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.3 standard deviations lower** (0.8 lower to 0.19 higher)
Depression diagnosis long follow-up (25-103 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: structured Clinical Interview (SCID))										
449 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	76/229 (33.2%)	73/220 (33.2%)	**RR 1** (0.77 to 1.3)	**Study population**
										332 per 1000 — **0 fewer per 1000** (from 76 fewer to 100 more)
										Moderate
										332 per 1000 — **0 fewer per 1000** (from 76 fewer to 100 more)

Depression diagnosis long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: structured Clinical Interview (SCID))										
346 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	28/181 (15.5%)	18/165 (10.9%)	**RR 0.71** (0.41 to 1.23)	**Study population**
										155 per 1000 | **45 fewer per 1000** (from 91 fewer to 36 more)
										Moderate
										155 per 1000 | **45 fewer per 1000** (from 91 fewer to 36 more)
Depression symptomatology long follow-up (25-103 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Center for Epidemiologic Studies Depression Scale (CES-D) ≥16)										
106 (1 study) 53 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	18/50 (36%)	19/56 (33.9%)	**RR 0.94** (0.56 to 1.58)	**Study population**
										360 per 1000 | **22 fewer per 1000** (from 158 fewer to 209 more)
										Moderate
										360 per 1000 | **22 fewer per 1000** (from 158 fewer to 209 more)
Depression symptomatology long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Center for Epidemiologic Studies Depression Scale (CES-D) ≥16)										
80 (1 study) 53 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	6/38 (15.8%)	5/42 (11.9%)	**RR 0.75** (0.25 to 2.27)	**Study population**
										158 per 1000 | **39 fewer per 1000** (from 118 fewer to 201 more)
										Moderate

									158 per 1000	**40 fewer per 1000** (from 119 fewer to 201 more)
Depression mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)										
354 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⨁⨁⨁⊖ **MODERATE**[4] due to imprecision	184	170	-	The mean depression mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.14 standard deviations lower** (0.35 lower to 0.06 higher)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Risk of bias due to statistically significant group differences at baseline
[4] Total population size is less than 400 (a threshold rule-of-thumb)

1.1.9 Depression: case management and individualised treatment

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With depression: case management and individualised treatment versus TAU**		**Risk with control**	**Risk difference with depression: case management and individualised treatment versus TAU** (95% CI)
Depression symptomatology post-treatment – ITT analysis (at-risk populations) (assessed with: Beck Depression Inventory (BDI) ≥9)											
										Study population	

34 (1 study) 5 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	7/16 (43.8%)	2/18 (11.1%)	**RR 0.25** (0.06 to 1.05)	**438 per 1000**	**328 fewer per 1000** (from 411 fewer to 22 more)
										Moderate	
										438 per 1000	**329 fewer per 1000** (from 412 fewer to 22 more)
Depression symptomatology post-treatment – available case analysis (at-risk populations) (assessed with: Beck Depression Inventory (BDI) ≥9)											
34 (1 study) 5 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	7/16 (43.8%)	2/18 (11.1%)	**RR 0.25** (0.06 to 1.05)	**Study population**	
										438 per 1000	**328 fewer per 1000** (from 411 fewer to 22 more)
										Moderate	
										438 per 1000	**329 fewer per 1000** (from 412 fewer to 22 more)

[1] Risk of bias due to statistically significant group differences at baseline

[2] Total number of events is less than 300 (a threshold rule-of-thumb)

[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.10 Anxiety: post-miscarriage self-help versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Anxiety: post-miscarriage self-help versus TAU**		**Risk with control**	**Risk difference with Anxiety: post-miscarriage self-help versus TAU** (95% CI)

Anxiety mean scores post-treatment – ITT analysis (at-risk populations) (measured with: Brief Symptom Inventory (BSI): Anxiety; better indicated by lower values)										
228 (1 study) 5 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to risk of bias, imprecision	113	115	-	The mean anxiety mean scores post-treatment – ITT analysis (at-risk populations) in the intervention groups was **0.47 standard deviations lower** (0.73 to 0.2 lower)

[1] Risk of bias due to statistically significant group differences at baseline

[2] Total population size is less than 400 (a threshold rule-of-thumb)

1.1.11 Anxiety: non-mental health-focused education and support versus TAU or enhanced TAU

Quality assessment							**Summary of findings**				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Anxiety: non-mental health-focused education and support versus TAU or enhanced TAU**		**Risk with control**	**Risk difference with Anxiety: non-mental health-focused education and support versus TAU or enhanced TAU** (95% CI)
Anxiety symptomatology post-treatment – ITT analysis (at-risk populations) (assessed with: Hospital Anxiety and Depression Scale – Anxiety (above unspecified threshold))											
162 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to imprecision, publication bias	25/82 (30.5%)	18/80 (22.5%)	**RR 0.74** (0.44 to 1.24)	**Study population**	
										305 per 1000	**79 fewer per 1000** (from 171 fewer to 73 more)
										Moderate	

									305 per 1000	**79 fewer per 1000** (from 171 fewer to 73 more)
Anxiety symptomatology post-treatment – available case analysis (at-risk populations) (assessed with: Hospital Anxiety and Depression Scale – Anxiety (above unspecified threshold))										
131 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	6/63 (9.5%)	6/68 (8.8%)	**RR 0.93** (0.32 to 2.72)	**Study population**
									95 per 1000	**7 fewer per 1000** (from 65 fewer to 164 more)
									Moderate	
									95 per 1000	**7 fewer per 1000** (from 65 fewer to 163 more)
Anxiety mean scores post-treatment – available case analysis (at-risk populations) (measured with: state-Trait Anxiety Inventory (STAI)-State or Hospital Anxiety and Depression Scale – Anxiety; better indicated by lower values)										
370 (2 studies) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	168	202	-	The mean anxiety mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.1 standard deviations lower** (0.3 lower to 0.11 higher)
Anxiety symptomatology short follow-up (9-16 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Hospital Anxiety and Depression Scale – Anxiety (above unspecified threshold))										
162 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	23/82 (28%)	15/80 (18.8%)	**RR 0.67** (0.38 to 1.19)	**Study population**
									280 per 1000	**93 fewer per 1000** (from 174 fewer to 53 more)
									Moderate	

									281 per 1000	**93 fewer per 1000** (from 174 fewer to 53 more)
Anxiety symptomatology short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Hospital Anxiety and Depression Scale – Anxiety (above unspecified threshold))										
128 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	4/63 (6.3%)	0/65 (0%)	**RR 0.11** (0.01 to 1.96)	**Study population**
									63 per 1000	**57 fewer per 1000** (from 63 fewer to 61 more)
									Moderate	
									64 per 1000	**57 fewer per 1000** (from 63 fewer to 61 more)
Anxiety mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Hospital Anxiety and Depression Scale – Anxiety; better indicated by lower values)										
128 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,4]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[2,3,4] due to imprecision, publication bias	63	65	-	The mean anxiety mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.2 standard deviations lower** (0.54 lower to 0.15 higher)
Anxiety symptomatology intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Hospital Anxiety and Depression Scale – Anxiety (above unspecified threshold))										
162 (1 study) 24 weeks	no serious	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to	23/82 (28%)	17/80 (21.3%)	**RR 0.76** (0.44 to 1.31)	**Study population**
									280 per 1000	**67 fewer per 1000** (from 157 fewer to 87 more)

	risk of bias					imprecision, publication bias			**Moderate**	
									281 per 1000	**67 fewer per 1000** (from 157 fewer to 87 more)
Anxiety symptomatology intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Hospital Anxiety and Depression Scale – Anxiety (above unspecified threshold))										
130 (1 study) 24 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	4/63 (6.3%)	4/67 (6%)	**RR 0.94** (0.25 to 3.6)	**Study population**
										63 per 1000 **4 fewer per 1000** (from 48 fewer to 165 more)
										Moderate
										64 per 1000 **4 fewer per 1000** (from 48 fewer to 166 more)
Anxiety mean scores intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Hospital Anxiety and Depression Scale – Anxiety; better indicated by lower values)										
130 (1 study) 24 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,4]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[2,3,4] due to imprecision, publication bias	63	67	-	The mean anxiety mean scores intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.26 standard deviations lower** (0.6 lower to 0.09 higher)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)

[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

[3] Paper omits data

[4] Total population size is less than 400 (a threshold rule-of-thumb)

1.1.12 Anxiety: home visits versus TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Anxiety: home visits versus TAU**		**Risk with control**	**Risk difference with Anxiety: home visits versus TAU** (95% CI)
Anxiety symptomatology post-treatment – ITT analysis (at-risk populations) (assessed with: Hospital Anxiety and Depression Scale – Anxiety (HADS >7))											
120 (1 study) 52 weeks	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias strongly suspected[3]	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	37/59 (62.7%)	24/61 (39.3%)	**RR 0.63** (0.43 to 0.91)	**Study population**	
										627 per 1000	**232 fewer per 1000** (from 56 fewer to 357 fewer)
										Moderate	
										627 per 1000	**232 fewer per 1000** (from 56 fewer to 357 fewer)
Anxiety symptomatology post-treatment – available case analysis (at-risk populations) (assessed with: Hospital Anxiety and Depression Scale – Anxiety (HADS >7))											
90 (1 study) 52 weeks	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias strongly suspected[3]	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	21/43 (48.8%)	10/47 (21.3%)	**RR 0.44** (0.23 to 0.82)	**Study population**	
										488 per 1000	**273 fewer per 1000** (from 88 fewer to 376 fewer)
										Moderate	
										488 per 1000	**273 fewer per 1000** (from 88 fewer to 376 fewer)

Anxiety mean scores post-treatment – available case analysis (at-risk populations) (measured with: Hospital Anxiety and Depression Scale – Anxiety; better indicated by lower values)										
90 (1 study) 52 weeks	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	43	47	-	The mean anxiety mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.89 standard deviations lower** (1.33 to 0.46 lower)
Anxiety symptomatology long follow-up (25-103 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Hospital Anxiety and Depression Scale – Anxiety (HADS ≥8))										
120 (1 study) 104 weeks	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	42/59 (71.2%)	32/61 (52.5%)	**RR 0.74** (0.55 to 0.98)	**Study population** **712 per 1000**: **185 fewer per 1000** (from 14 fewer to 320 fewer) **Moderate** **712 per 1000**: **185 fewer per 1000** (from 14 fewer to 320 fewer)
Anxiety symptomatology long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Hospital Anxiety and Depression Scale – Anxiety (HADS ≥8))										
77 (1 study) 104 weeks	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	21/38 (55.3%)	10/39 (25.6%)	**RR 0.46** (0.25 to 0.85)	**Study population** **553 per 1000**: **298 fewer per 1000** (from 83 fewer to 414 fewer) **Moderate** **553 per 1000**: **299 fewer per 1000** (from 83 fewer to 415 fewer)

Anxiety mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Hospital Anxiety and Depression Scale – Anxiety; better indicated by lower values)										
77 (1 study) 104 weeks	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	38	39	-		The mean anxiety mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.61 standard deviations lower** (1.06 to 0.15 lower)

[1] Risk of bias due to statistcially significant group differences at baseline

[2] Total number of events is less than 300 (a threshold rule-of-thumb)

[3] Paper omits data

[4] Total population size is less than 400 (a threshold rule-of-thumb)

1.1.13 PTSD: post-miscarriage self-help versus TAU

Quality assessment							**Summary of findings**				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With PTSD: post-miscarriage self-help versus TAU**		**Risk with control**	**Risk difference with PTSD: post-miscarriage self-help versus TAU** (95% CI)
PTSD symptomatology post-treatment – ITT analysis (at-risk populations) (assessed with: Impact of Events Scale-Revised (IES-R) ≥35)											
228 (1 study) 5 weeks	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	35/113 (31%)	12/115 (10.4%)	**RR 0.34** (0.18 to 0.62)	**Study population**	
										310 per 1000	**204 fewer per 1000** (from 118 fewer to 254 fewer)
										Moderate	

									310 per 1000	**205 fewer per 1000** (from 118 fewer to 254 fewer)
PTSD mean scores post-treatment – ITT analysis (at-risk populations) (measured with: Impact of Events Scale-Revised (IES-R); better indicated by lower values)										
228 (1 study) 5 weeks	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⨁⨀⨀⨀ **VERY LOW**[1,3] due to risk of bias, imprecision	113	115	-	The mean ptsd mean scores post-treatment – ITT analysis (at-risk populations) in the intervention groups was **0.88 standard deviations lower** (1.15 to 0.61 lower)

[1] Risk of bias due to statistically significant group differences at baseline

[2] Total number of events is less than 300 (a threshold rule-of-thumb)

[3] Total population size is less than 400 (a threshold rule-of-thumb)

1.1.14 General mental health: post-miscarriage self-help versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With General mental health: post-miscarriage self-help versus TAU**		**Risk with control**	**Risk difference with General mental health: post-miscarriage self-help versus TAU** (95% CI)
General mental health mean scores post-treatment – ITT analysis (at-risk populations) (measured with: Brief Symptom Inventory (BSI): Global severity index (Mental health); better indicated by lower values)											
228 (1 study) 5 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to risk of bias, imprecision	113	115	-		The mean general mental health mean scores post-treatment – ITT analysis (at-risk populations) in the intervention groups was **0.61 standard deviations lower** (0.87 to 0.34 lower)

[1] High risk of selection bias due to unclear allocation concealment and statistically significant difference in baseline intrusion subscale of the IES-R (19.2 in control group and 17.4 in intervention group)

[2] Total population size is less than 400 (a threshold rule-of-thumb)

1.1.15 General mental health: home visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With General mental health: home visits versus TAU**		**Risk with control**	**Risk difference with General mental health: home visits versus TAU** (95% CI)
General mental health mean scores post-treatment – available case analysis (at-risk populations) (measured with: General Health Questionnaire (GHQ); better indicated by lower values)											
207 (2 studies) 78 weeks	no serious risk of bias	very serious[1]	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to inconsistency, imprecision, publication bias	101	106	-		The mean general mental health mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.18 standard deviations lower** (0.7 lower to 0.33 higher)

[1] There is evidence of substantial heterogeneity of study effect sizes
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Total population size is less than 400 (a threshold rule-of-thumb)
[4] Paper omits data

1.1.16 General mental health: post-delivery discussion versus enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With General mental health: post-delivery discussion versus enhanced TAU**		**Risk with control**	**Risk difference with General mental health: post-delivery discussion versus enhanced TAU** (95% CI)
General mental health mean scores post-treatment – available case analysis (at-risk populations) (measured with: sF-36 – Mental health; better indicated by lower values)											

917 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	450	467	-		The mean general mental health mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.08 standard deviations lower** (0.21 lower to 0.05 higher)
General mental health mean scores Very long follow-up (>104 weeks post-intervention) – available case analysis (at-risk populations) (measured with: sF-36 – Mental health; better indicated by lower values)											
534 (1 study) 208-312 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	270	264	-		The mean general mental health mean scores very long follow-up (>104 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.17 standard deviations higher** (0 to 0.34 higher)

1.1.17 General mental health: mother–infant relationship interventions versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With General mental health: mother–infant relationship interventions versus TAU**		**Risk with control**	**Risk difference with General mental health: mother–infant relationship interventions versus TAU** (95% CI)
General mental health mean scores post-treatment – available case analysis (at-risk populations) (measured with: General Health Questionnaire (GHQ-28); better indicated by lower values)											
125 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	61	64	-		The mean general mental health mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was

										0.18 standard deviations higher (0.17 lower to 0.53 higher)
General mental health mean scores long follow-up (25-104 weeks post-intervention) – available case analysis (at-risk populations) (measured with: General Health Questionnaire (GHQ-28); better indicated by lower values)										
88 (1 study) 104 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	39	49	-	The mean general mental health mean scores long follow-up (25-104 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.09 standard deviations lower** (0.52 lower to 0.33 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)

[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.18 Mother–infant attachment: non-mental health-focused education and support versus TAU or enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Mother–infant attachment: non-mental health-focused education and support versus TAU or enhanced TAU**		**Risk with control**	**Risk difference with Mother–infant attachment: non-mental health-focused education and support versus TAU or enhanced TAU** (95% CI)
Mother–infant attachment problems post-treatment – ITT analysis (at-risk populations) (assessed with: Green scale: mother–infant attachment problems (above unspecified threshold))											
162 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to imprecision, publication bias	41/82 (50%)	36/80 (45%)	**RR 0.9** (0.65 to 1.25)	**Study population**	
										500 per 1000	**50 fewer per 1000** (from 175 fewer to 125 more)
										Moderate	

									500 per 1000	**50 fewer per 1000** (from 175 fewer to 125 more)
Mother–infant attachment problems post-treatment – available case analysis (at-risk populations) (assessed with: Green scale: mother–infant attachment problems (above unspecified threshold))										
133 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	23/64 (35.9%)	25/69 (36.2%)	**RR 1.01** (0.64 to 1.59)	**Study population**
										359 per 1000 **4 more per 1000** (from 129 fewer to 212 more)
										Moderate
										359 per 1000 **4 more per 1000** (from 129 fewer to 212 more)
Positive mother–infant interaction mean scores post-treatment – available case analysis (at-risk populations) (measured with: Index of Parental Behavior in the NICU: positive interaction with quiet alert infant; better indicated by lower values)										
211 (1 study)	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected[3]	⊕⊕⊖⊖ **LOW**[3,4] due to imprecision, publication bias	95	116	-	The mean positive mother–infant interaction mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.57 standard deviations higher** (0.29 to 0.85 higher)
Maternal sensitivity mean scores post-treatment – available case analysis (at-risk populations) (measured with: Index of Parental Behavior in the NICU: sensitivity to needs of infant in NICU; better indicated by lower values)										
199 (1 study)	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[3,4] due to imprecision, publication bias	87	112	-	The mean maternal sensitivity mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.3 standard deviations higher** (0.02 to 0.58 higher)

Maternal confidence mean scores post-treatment – available case analysis (at-risk populations) (measured with: parental Belief Scale-NICU: parent role confidence; better indicated by lower values)										
241 (1 study)	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected[3]	⨁⨁⨀⨀ **LOW**[3,4] due to imprecision, publication bias	107	134	-	The mean maternal confidence mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.15 standard deviations higher** (0.1 lower to 0.41 higher)
Mother–infant attachment problems short follow-up (9-16 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Green scale: mother–infant attachment problems (above unspecified threshold))										
162 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to imprecision, publication bias	38/82 (46.3%)	40/80 (50%)	**RR 1.08** (0.78 to 1.49)	**Study population**
										463 per 1000 — **37 more per 1000** (from 102 fewer to 227 more)
										Moderate
										463 per 1000 — **37 more per 1000** (from 102 fewer to 227 more)
Mother–infant attachment problems short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Green scale: mother–infant attachment problems (above unspecified threshold))										
126 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to imprecision, publication bias	18/62 (29%)	24/64 (37.5%)	**RR 1.29** (0.78 to 2.13)	**Study population**
										290 per 1000 — **84 more per 1000** (from 64 fewer to 328 more)
										Moderate
										290 per 1000 — **84 more per 1000** (from 64 fewer to 328 more)
Mother–infant attachment problems intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Green scale: mother–infant attachment problems (above unspecified threshold))										

162 (1 study) 24 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	48/82 (58.5%)	40/80 (50%)	**RR 0.85** (0.64 to 1.14)	**Study population**	
										585 per 1000	**88 fewer per 1000** (from 211 fewer to 82 more)
										Moderate	
										585 per 1000	**88 fewer per 1000** (from 211 fewer to 82 more)
Mother–infant attachment problems intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Green scale: mother–infant attachment problems (above unspecified threshold))											
127 (1 study) 24 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	27/61 (44.3%)	26/66 (39.4%)	**RR 0.89** (0.59 to 1.34)	**Study population**	
										443 per 1000	**49 fewer per 1000** (from 181 fewer to 150 more)
										Moderate	
										443 per 1000	**49 fewer per 1000** (from 182 fewer to 151 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data
[4] Total population size is less than 400 (a threshold rule-of-thumb)

1.1.19 Mother–infant attachment: home visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Mother–infant attachment: home visits versus TAU**		**Risk with control**	**Risk difference with Mother–infant attachment: home visits versus TAU** (95% CI)
Maternal sensitivity mean scores post-treatment – available case analysis (at-risk populations) (measured with: CARE Index scale – Maternal sensitivity; better indicated by lower values)											

121 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	59	62	-		The mean maternal sensitivity mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.36 standard deviations higher** (0 to 0.72 higher)
Infant involvement mean scores post-treatment – available case analysis (at-risk populations) (measured with: CARE Index scale – Infant cooperativeness; better indicated by lower values)											
121 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,3] due to imprecision, publication bias	59	62	-		The mean infant involvement mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.42 standard deviations higher** (0.06 to 0.78 higher)
Discontinued breastfeeding <6 months – ITT analysis (at-risk populations) (assessed with: Breastfeeding – discontinued before 6 months)											
131 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,4]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[2,3,4] due to imprecision, publication bias	24/63 (38.1%)	20/68 (29.4%)	**RR 0.77** (0.48 to 1.25)	**Study population**	
										381 per 1000	**88 fewer per 1000** (from 198 fewer to 95 more)
										Moderate	
										381 per 1000	**88 fewer per 1000** (from 198 fewer to 95 more)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data
[4] Total number of events is less than 300 (a threshold rule-of-thumb)

1.1.20 Mother–infant attachment: mother–infant relationship interventions versus TAU

Quality assessment	Summary of findings

Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	**With Mother–infant attachment: mother–infant relationship interventions versus TAU**		**Risk with control**	**Risk difference with Mother–infant attachment: mother–infant relationship interventions versus TAU** (95% CI)
Mother–infant attachment problems post-treatment – ITT analysis (at-risk populations) (assessed with: Ainsworth Strange Situation: Insecure)											
449 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁◯◯ **LOW**[1,2] due to imprecision	127/229 (55.5%)	104/220 (47.3%)	**RR 0.85** (0.71 to 1.02)	**Study population**	
										555 per 1000	**83 fewer per 1000** (from 161 fewer to 11 more)
										Moderate	
										555 per 1000	**83 fewer per 1000** (from 161 fewer to 11 more)
Mother–infant attachment problems post-treatment – available case analysis (at-risk populations) (assessed with: Ainsworth Strange Situation: Insecure)											
318 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⨁⨁◯◯ **LOW**[1] due to imprecision	60/162 (37%)	40/156 (25.6%)	**RR 0.69** (0.5 to 0.97)	**Study population**	
										370 per 1000	**115 fewer per 1000** (from 11 fewer to 185 fewer)
										Moderate	
										370 per 1000	**115 fewer per 1000** (from 11 fewer to 185 fewer)
Positive mother–infant interaction mean scores post-treatment – available case analysis (at-risk populations) (measured with: infant and Caregiver Engagement Phases (ICEP): maternal positive engagement (% of time during behavioural observation) or Synchrony Scale (Milgrom & Meitz, 1988): Reciprocity/Synchrony; better indicated by lower values)											
175 (2 studies) 15-26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⨁⨁◯◯ **LOW**[3] due to imprecision	86	89	-		The mean positive mother–infant interaction mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was

										0.46 standard deviations higher (0.16 to 0.76 higher)
Maternal sensitivity mean scores post-treatment – available case analysis (at-risk populations) (measured with: maternal Sensitivity and Responsivity Scales (MSRS): maternal sensitivity or Synchrony Scale (Milgrom & Meitz, 1988): maternal Respond ; better indicated by lower values)										
172 (2 studies) 15-26 weeks	no serious risk of bias	very serious[4]	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3,4] due to inconsistency, imprecision	87	85	-	The mean maternal sensitivity mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.62 standard deviations higher** (0.11 lower to 1.35 higher)
Maternal intrusiveness mean scores post-treatment – available case analysis (at-risk populations) (measured with: maternal Sensitivity and Responsivity Scales (MSRS): maternal intrusiveness; better indicated by lower values)										
109 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊕⊖⊖ **LOW**[2,3] due to imprecision	56	53	-	The mean maternal intrusiveness mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.32 standard deviations lower** (0.7 lower to 0.06 higher)
Maternal negative engagement mean scores post-treatment – available case analysis (at-risk populations) (measured with: infant and Caregiver Engagement Phases (ICEP): maternal negative engagement (angry/hostile/stern/sad/sober/expressionless; % of time during behavioural observation); better indicated by lower values)										
112 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3] due to imprecision	55	57	-	See comment	See comment
Infant involvement mean scores post-treatment – available case analysis (at-risk populations) (measured with: infant and Caregiver Engagement Phases (ICEP): infant positive engagement (% of time during behavioural observation); better indicated by lower values)										
112 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊕⊖⊖ **LOW**[2,3] due to imprecision	55	57	-		The mean infant involvement mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was

											0.31 standard deviations lower (0.69 lower to 0.06 higher)
Infant responsivity mean scores post-treatment – available case analysis (at-risk populations) (measured with: infant and Caregiver Engagement Phases (ICEP): infant responsivity (mother–focused attention; % of time during behavioural observation) or Synchrony Scale (Milgrom & Meitz, 1988): Attending to mother ; better indicated by lower values)											
175 (2 studies) 15-26 weeks	no serious risk of bias	very serious[4]	no serious indirectness	very serious[2,3]	undetected	⨁⊖⊖⊖ **VERY LOW**[2,3,4] due to inconsistency, imprecision	86	89	-		The mean infant responsivity mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.52 standard deviations higher** (0.63 lower to 1.68 higher)
Infant negative engagement/behaviour problems mean score post-treatment – available case analysis (at-risk populations) (measured with: infant and Caregiver Engagement Phases (ICEP): infant negative engagement (behaviour problems; % of time during behavioural observation); better indicated by lower values)											
112 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⨁⊖⊖ **LOW**[2,3] due to imprecision	55	57	-		The mean infant negative engagement/behaviour problems mean score post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.16 standard deviations higher** (0.21 lower to 0.53 higher)
Discontinued breastfeeding <6 months – ITT analysis (at-risk populations) (assessed with: infant feeding-breast feeding stopped by 26 weeks)											
106 (1 study) 27 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[6]	⨁⊖⊖⊖ **VERY LOW**[1,2,5,6] due to risk of bias, imprecision, publication bias	22/50 (44%)	22/56 (39.3%)	**RR 0.89** (0.57 to 1.4)	**Study population**	
										440 per 1000	**48 fewer per 1000** (from 189 fewer to 176 more)
										Moderate	
										440 per 1000	**48 fewer per 1000** (from 189 fewer to 176 more)
Discontinued breastfeeding <6 months – available case analysis (at-risk populations) (assessed with: infant feeding-breast feeding stopped by 26 weeks)											

88 (1 study) 27 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[6]	⊕⊖⊖⊖ **VERY LOW**[1,2,5,6] due to risk of bias, imprecision, publication bias	16/44 (36.4%)	10/44 (22.7%)	**RR 0.62** (0.32 to 1.22)	**Study population**	
										364 per 1000	**138 fewer per 1000** (from 247 fewer to 80 more)
										Moderate	
										364 per 1000	**138 fewer per 1000** (from 248 fewer to 80 more)
Discontinued breastfeeding <9 months – ITT analysis (at-risk populations) (assessed with: infant feeding-breast feeding stopped by 39 weeks)											
106 (1 study) 40 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[6]	⊕⊖⊖⊖ **VERY LOW**[1,2,5,6] due to risk of bias, imprecision, publication bias	34/50 (68%)	29/56 (51.8%)	**RR 0.76** (0.56 to 1.04)	**Study population**	
										680 per 1000	**163 fewer per 1000** (from 299 fewer to 27 more)
										Moderate	
										680 per 1000	**163 fewer per 1000** (from 299 fewer to 27 more)
Discontinued breastfeeding <9 months – available case analysis (at-risk populations) (assessed with: infant feeding-breast feeding stopped by 39 weeks)											
81 (1 study) 40 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[6]	⊕⊖⊖⊖ **VERY LOW**[1,5,6] due to risk of bias, imprecision, publication bias	24/40 (60%)	14/41 (34.1%)	**RR 0.57** (0.35 to 0.93)	**Study population**	
										600 per 1000	**258 fewer per 1000** (from 42 fewer to 390 fewer)
										Moderate	
										600 per 1000	**258 fewer per 1000** (from 42 fewer to 390 fewer)
Discontinued breastfeeding <12 months – ITT analysis (at-risk populations) (assessed with: infant feeding-breast feeding stopped by 52 weeks)											
106 (1 study) 53 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[6]	⊕⊖⊖⊖ **VERY LOW**[1,2,5,6] due to risk of bias,	42/50 (84%)	40/56 (71.4%)	**RR 0.85** (0.69 to 1.04)	**Study population**	
										840 per 1000	**126 fewer per 1000** (from 260 fewer to 34 more)

						imprecision, publication bias			**Moderate**	
									840 per 1000	**126 fewer per 1000** (from 260 fewer to 34 more)
Discontinued breastfeeding <12 months – available case analysis (at-risk populations) (assessed with: infant feeding-breast feeding stopped by 52 weeks)										
82 (1 study) 53 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[6]	⊕⊖⊖⊖ **VERY LOW**[1,2,5,6] due to risk of bias, imprecision, publication bias	32/40 (80%)	26/42 (61.9%)	**RR 0.77** (0.58 to 1.03)	**Study population**
										800 per 1000 **184 fewer per 1000** (from 336 fewer to 24 more)
										Moderate
										800 per 1000 **184 fewer per 1000** (from 336 fewer to 24 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Total population size is less than 400 (a threshold rule-of-thumb)
[4] There is evidence of considerable heterogeneity of study effect sizes
[5] High risk of selection bias due to statistically significant baseline difference with the intervention group having more mothers with earlier preterm birth and non-Norwegian origin
[6] Paper omits data

1.1.21 Mother–infant attachment: case management and individualised treatment

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Mother–infant attachment: case management and individualised treatment versus TAU**		**Risk with control**	**Risk difference with Mother–infant attachment: case management and individualised treatment versus TAU** (95% CI)
Maternal sensitivity post-treatment – ITT analysis (at-risk populations) (assessed with: Behavioural observation: maternal sensitivity)											
30 (1 study) 5 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of	10/15 (66.7%)	14/15 (93.3%)	**RR 1.4** (0.95 to 2.05)	**Study population**	
										667 per 1000	**267 more per 1000** (from 33 fewer to 700 more)

						bias, imprecision			**Moderate**	
									667 per 1000	**267 more per 1000** (from 33 fewer to 700 more)
Maternal sensitivity post-treatment – available case analysis (at-risk populations) (assessed with: Behavioural observation: maternal sensitivity)										
30 (1 study) 5 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	10/15 (66.7%) 14/15 (93.3%)	**RR 1.4** (0.95 to 2.05)	**Study population**	
									667 per 1000	**267 more per 1000** (from 33 fewer to 700 more)
									Moderate	
									667 per 1000	**267 more per 1000** (from 33 fewer to 700 more)

[1] High risk of selection bias due to unclear allocation concealment and statistically significant baseline difference in maternal age (29.7 in intervention group and 25.9 in control group)
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.22 Quality of life: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Quality of life: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU**		**Risk with control**	**Risk difference with Quality of life: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU** (95% CI)
Poor social support post-treatment – ITT analysis (at-risk populations) (assessed with: poor social support (interview))											
209 (1 study) 27 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	20/106 (18.9%)	21/103 (20.4%)	**RR 1.08** (0.62 to 1.87)	**Study population**	
										189 per 1000	**15 more per 1000** (from 72 fewer to 164 more)

									Moderate	
									189 per 1000	**15 more per 1000** (from 72 fewer to 164 more)
Poor social support post-treatment – available case (at-risk populations) (assessed with: poor social support (interview))										
190 (1 study) 27 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	10/96 (10.4%) 12/94 (12.8%)	**RR 1.23** (0.56 to 2.7)	**Study population**	
									104 per 1000	**24 more per 1000** (from 46 fewer to 177 more)
									Moderate	
									104 per 1000	**24 more per 1000** (from 46 fewer to 177 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.23 Quality of life: non-mental health-focused education and support versus TAU or enhanced TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Quality of life: non-mental health-focused education and support versus TAU or enhanced TAU**		**Risk with control**	**Risk difference with Quality of life: non-mental health-focused education and support versus TAU or enhanced TAU** (95% CI)
Parental stress mean scores post-treatment – available case analysis (at-risk populations) (measured with: parental Stressor Scale-Neonatal Intensive Care (PSS-NICU) or Parenting Stress Index (PSI); better indicated by lower values)											
369 (2 studies) 0.4-24 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias strongly suspected[2]	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision, publication bias	168	201	-		The mean parental stress mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was

										0.44 standard deviations lower (0.72 to 0.16 lower)
Social support mean scores post-treatment – available case analysis (at-risk populations) (measured with: satisfaction with Motherhood scale: social support; better indicated by lower values)										
133 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,3]	reporting bias strongly suspected[2]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	64	69	-	The mean social support mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.22 standard deviations higher** (0.12 lower to 0.57 higher)
Social support mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) (measured with: satisfaction with Motherhood scale: social support; better indicated by lower values)										
127 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[2]	⊕⊖⊖⊖ **VERY LOW**[1,2] due to imprecision, publication bias	63	64	-	The mean social support mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.39 standard deviations higher** (0.04 to 0.74 higher)
Social support mean scores intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) (measured with: satisfaction with Motherhood scale: social support; better indicated by lower values)										
129 (1 study) 24 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[2]	⊕⊖⊖⊖ **VERY LOW**[1,2] due to imprecision, publication bias	63	66	-	The mean social support mean scores intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.52 standard deviations higher** (0.17 to 0.87 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] Papers omit data
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.24 Quality of life: home visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With Quality of life: home visits versus TAU		Risk with control	Risk difference with Quality of life: home visits versus TAU (95% CI)
Social support mean scores post-treatment – available case analysis (at-risk populations) (measured with: social Support Questionnaire (SSQ); better indicated by lower values)											
29 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	12	17	-		The mean social support mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.58 standard deviations higher** (0.17 lower to 1.34 higher)
Self-esteem mean scores post-treatment – available case analysis (at-risk populations) (measured with: Rosenberg Self-Esteem Scale (SES); better indicated by lower values)											
114 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,3] due to imprecision, publication bias	55	59	-		The mean self-esteem mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.04 standard deviations lower** (0.41 lower to 0.33 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data

1.1.25 Quality of life: mother–infant relationship interventions versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With Quality of life: mother–infant relationship interventions versus TAU		Risk with control	Risk difference with Quality of life: mother–infant relationship interventions versus TAU (95% CI)
Parental stress mean scores post-treatment – available case analysis (at-risk populations) (measured with: Nijmeegse Ouderlijke Stress Index (NOSIK) or Parenting Stress Index (PSI); better indicated by lower values)											
244 (3 studies) 15-52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	112	132	-		The mean parental stress mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.16 standard deviations higher** (0.09 lower to 0.41 higher)
Parental stress mean scores long follow-up (25-104 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Nijmeegse Ouderlijke Stress Index (NOSI) or Parenting Stress Index (PSI); better indicated by lower values)											
183 (2 studies) 53-104 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	82	101	-		The mean parental stress mean scores long follow-up (25-104 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.02 standard deviations lower** (0.33 lower to 0.29 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)

1.1.26 Quality of life: case management and individualised treatment versus TAU

Quality assessment							Summary of findings		
		Inconsistency	Indirectness	Imprecision			Study event rates (%)		Anticipated absolute effects

Participants (studies) Follow-up	Risk of bias				Publication bias	Overall quality of evidence	With control	With Quality of life: case management and individualised treatment versus TAU	Relative effect (95% CI)	Risk with control	Risk difference with Quality of life: case management and individualised treatment versus TAU (95% CI)
Parental stress mean scores post-treatment – ITT analysis (at-risk populations) (measured with: parental Stressor Scale-Neonatal Intensive Care (PSS-NICU); better indicated by lower values)											
34 (1 study) 5 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	16	18	-		The mean parental stress mean scores post-treatment – ITT analysis (at-risk populations) in the intervention groups was **0.43 standard deviations lower** (1.11 lower to 0.25 higher)
Parental stress mean scores post-treatment – available case analysis (at-risk populations) – Case management and individualised treatment (measured with: parental Stressor Scale-Neonatal Intensive Care (PSS-NICU); better indicated by lower values)											
34 (1 study) 5 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	16	18	-		The mean parental stress mean scores post-treatment – available case analysis (at-risk populations) – case management and individualised treatment in the intervention groups was **0.43 standard deviations lower** (1.11 lower to 0.25 higher)
Self-esteem mean scores post-treatment – ITT analysis (at-risk populations) (measured with: maternal Self-Report Inventory (MSRI); better indicated by lower values)											
34 (1 study) 5 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	16	18	-		The mean self-esteem mean scores post-treatment – ITT analysis (at-risk populations) in the intervention groups was **0.3 standard deviations lower** (0.97 lower to 0.38 higher)
Self-esteem mean scores post-treatment – available case analysis (at-risk populations) (measured with: maternal Self-Report Inventory (MSRI); better indicated by lower values)											
34 (1 study) 5 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3]	16	18	-		The mean self-esteem mean scores post-treatment – available case analysis (at-risk populations) in

						due to risk of bias, imprecision			the intervention groups was **0.3 standard deviations lower** (0.97 lower to 0.38 higher)

[1] High risk of selection bias due to unclear allocation concealment and statistically significant baseline difference in maternal age (29.7 in intervention group and 25.9 in control group)
[2] Total population size is less than 400 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.27 Service utilisation: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**
							With control	**With service utilisation: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU**		**Risk with control**	**Risk difference with service utilisation: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU** (95% CI)
Contact with primary and/or secondary care post-treatment – ITT analysis (at-risk populations) (assessed with: primary and secondary health service contact since randomisation)											
209 (1 study) 27 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁◯◯ **LOW**[1,2] due to imprecision	11/106 (10.4%)	13/103 (12.6%)	**RR 1.22** (0.57 to 2.59)	**Study population**	
										104 per 1000	**23 more per 1000** (from 45 fewer to 165 more)
										Moderate	
										104 per 1000	**23 more per 1000** (from 45 fewer to 165 more)
Contact with primary and/or secondary care post-treatment – available case analysis (at-risk populations) (assessed with: primary and secondary health service contact since randomisation)											
190 (1 study) 27 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁◯◯ **LOW**[1,2] due to imprecision	11/96 (11.5%)	13/94 (13.8%)	**RR 1.21** (0.57 to 2.56)	**Study population**	
										115 per 1000	**24 more per 1000** (from 49 fewer to 179 more)

								Moderate	
								115 per 1000	**24 more per 1000** (from 49 fewer to 179 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.28 Service utilisation: home visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With service utilisation: home visits versus TAU**		**Risk with control**	**Risk difference with service utilisation: home visits versus TAU** (95% CI)
Maternal contact with primary and/or secondary care post-treatment – ITT analysis (at-risk populations) (assessed with: Linkage with primary care (Has a regular personal doctor at year 2))											
84 (1 study) 117 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	15/40 (37.5%)	19/44 (43.2%)	**RR 1.15** (0.68 to 1.95)	**Study population**	
										375 per 1000	**56 more per 1000** (from 120 fewer to 356 more)
										Moderate	
										375 per 1000	**56 more per 1000** (from 120 fewer to 356 more)
Maternal contact with primary and/or secondary care post-treatment – available case analysis (at-risk populations) (assessed with: Linkage with primary care (Has a regular personal doctor at year 2))											
63 (1 study) 117 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	15/32 (46.9%)	19/31 (61.3%)	**RR 1.31** (0.82 to 2.08)	**Study population**	
										469 per 1000	**145 more per 1000** (from 84 fewer to 506 more)
										Moderate	

									469 per 1000	**145 more per 1000** (from 84 fewer to 507 more)
Infant admissions to hospital mid-treatment (at 6 months) – ITT analysis (at-risk populations) (assessed with: infant service use: Admissions to hospital since birth)										
131 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[2,3,4] due to imprecision, publication bias	8/63 (12.7%)	5/68 (7.4%)	**RR 0.58** (0.2 to 1.68)	**Study population**	
										127 per 1000	**53 fewer per 1000** (from 102 fewer to 86 more)
										Moderate	
										127 per 1000	**53 fewer per 1000** (from 102 fewer to 86 more)
Infant length of stay in hospital mid-treatment (at 6 months) – ITT analysis (at-risk populations) (measured with: infant service use: median days stayed in hospital; better indicated by lower values)											
131 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3,5]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[3,4,5] due to imprecision, publication bias	63	68	-		The mean infant length of stay in hospital mid-treatment (at 6 months) – ITT analysis (at-risk populations) in the intervention groups was **0.16 standard deviations lower** (0.5 lower to 0.19 higher)

[1] High risk of selection bias due to unclear randomisation method and allocation concealment and statistically significant group difference at baseline (intervention group scored higher on measure of parenting attitudes and beliefs)
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] Paper omits data
[5] Total population size is less than 400 (a threshold rule-of-thumb)

1.1.29 Experience of care: non-mental health-focused education and support versus TAU or enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With experience of care: non-mental health-focused education and support versus TAU or enhanced TAU		Risk with control	Risk difference with experience of care: non-mental health-focused education and support versus TAU or enhanced TAU (95% CI)
Maternal dissatisfaction with care post-treatment – ITT analysis (at-risk populations) (assessed with: self-report)											
162 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	52/82 (63.4%)	40/80 (50%)	**RR 0.79** (0.6 to 1.04)	**Study population**	
										634 per 1000	**133 fewer per 1000** (from 254 fewer to 25 more)
										Moderate	
										634 per 1000	**133 fewer per 1000** (from 254 fewer to 25 more)
Maternal dissatisfaction with care post-treatment – available case analysis (at-risk populations) (assessed with: self-report)											
141 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	39/69 (56.5%)	32/72 (44.4%)	**RR 0.79** (0.56 to 1.09)	**Study population**	
										565 per 1000	**119 fewer per 1000** (from 249 fewer to 51 more)
										Moderate	
										565 per 1000	**119 fewer per 1000** (from 249 fewer to 51 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data

1.1.30 Attrition: post-miscarriage self-help versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With attrition: post-miscarriage self-help versus TAU		Risk with control	Risk difference with attrition: post-miscarriage self-help versus TAU (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
228 (1 study) 5 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	13/113 (11.5%)	16/115 (13.9%)	**RR 1.21** (0.61 to 2.4)	**Study population**	
										115 per 1000	**24 more per 1000** (from 45 fewer to 161 more)
										Moderate	
										115 per 1000	**24 more per 1000** (from 45 fewer to 161 more)

[1] High risk of selection bias due to unclear allocation concealment and statistically significant group differences at baseline
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.31 Attrition: social support versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With attrition: social support versus TAU		Risk with control	Risk difference with attrition: social support versus TAU (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
					undetected					**Study population**	

117 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]		⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	21/56 (37.5%)	31/61 (50.8%)	**RR 1.36** (0.89 to 2.06)	**375 per 1000**	**135 more per 1000** (from 41 fewer to 397 more)
										Moderate		
										375 per 1000	**135 more per 1000** (from 41 fewer to 397 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.32 Attrition: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With attrition: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU**		**Risk with control**	**Risk difference with attrition: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU** (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
360 (3 studies) 26-27 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	12/179 (6.7%)	19/181 (10.5%)	**RR 1.63** (0.5 to 5.28)	**Study population**	
										67 per 1000	**42 more per 1000** (from 34 fewer to 287 more)
										Moderate	
										94 per 1000	**59 more per 1000** (from 47 fewer to 402 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.33 Attrition: psychoeducational booklet versus TAU or enhanced TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With attrition: psychoeducational booklet versus TAU or enhanced TAU		Risk with control	Risk difference with attrition: psychoeducational booklet versus TAU or enhanced TAU (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
600 (1 study)	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	122/301 (40.5%)	107/299 (35.8%)	**RR 0.88** (0.72 to 1.08)	**Study population**	
										405 per 1000	**49 fewer per 1000** (from 113 fewer to 32 more)
										Moderate	
										405 per 1000	**49 fewer per 1000** (from 113 fewer to 32 more)

[1] High risk of selection bias due to statistically significant group differences at baseline
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.34 Attrition: non-mental health-focused education and support versus TAU or enhanced TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With attrition: non-mental health-focused education and support versus TAU or enhanced TAU		Risk with control	Risk difference with attrition: non-mental health-focused education and support versus TAU or enhanced TAU (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
	serious[1]				undetected					**Study population**	

584 (3 studies) 6-28 weeks		no serious inconsistency	no serious indirectness	very serious[2,3]		⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	61/292 (20.9%)	44/292 (15.1%)	**RR 0.72** (0.5 to 1.02)	**209 per 1000**	**58 fewer per 1000** (from 104 fewer to 4 more)
										Moderate	
										207 per 1000	**58 fewer per 1000** (from 104 fewer to 4 more)

[1] High risk of selection bias due to a statistically significant group difference at baseline
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.35 Attrition: home visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With attrition: home visits versus TAU**		**Risk with control**	**Risk difference with attrition: home visits versus TAU** (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
215 (2 studies) 78-117 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	13/103 (12.6%)	17/112 (15.2%)	**RR 1.23** (0.64 to 2.37)	**Study population**	
										126 per 1000	**29 more per 1000** (from 45 fewer to 173 more)
										Moderate	
										140 per 1000	**32 more per 1000** (from 50 fewer to 192 more)

[1] High risk of selection bias due to unclear randomisation method and statistically significant group difference at baseline
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.36 Attrition: post-delivery discussion versus enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With attrition: post-delivery discussion versus enhanced TAU		Risk with control	Risk difference with attrition: post-delivery discussion versus enhanced TAU (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
1041 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	71/521 (13.6%)	53/520 (10.2%)	**RR 0.75** (0.54 to 1.04)	**Study population**	
										136 per 1000	**34 fewer per 1000** (from 63 fewer to 5 more)
										Moderate	
										136 per 1000	**34 fewer per 1000** (from 63 fewer to 5 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)

[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.37 Attrition: mother–infant relationship interventions versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With attrition: mother–infant relationship interventions versus TAU		Risk with control	Risk difference with attrition: mother–infant relationship interventions versus TAU (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
772 (4 studies) 15-26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	78/389 (20.1%)	80/383 (20.9%)	**RR 1.04** (0.76 to 1.43)	**Study population**	
										201 per 1000	**8 more per 1000** (from 48 fewer to 86 more)

									Moderate	
								168 per 1000	**7 more per 1000** (from 40 fewer to 72 more)	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.38 Infant physical health: home visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant physical health: home visits versus TAU**		**Risk with control**	**Risk difference with infant physical health: home visits versus TAU** (95% CI)
Congenital malformations (measured at 6 months) – available case analysis (at-risk populations) (assessed with: number of infants with a disability)											
131 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	1/63 (1.6%)	6/68 (8.8%)	**RR 5.56** (0.69 to 44.9)	**Study population**	
										16 per 1000	**72 more per 1000** (from 5 fewer to 697 more)
										Moderate	
										16 per 1000	**73 more per 1000** (from 5 fewer to 702 more)
Normal weight post-treatment – available case analysis (at-risk populations) (assessed with: number of infants of a normal weight)											
79 (1 study)	serious[3]	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to risk of bias, imprecision	17/38 (44.7%)	20/41 (48.8%)	**RR 1.09** (0.68 to 1.75)	**Study population**	
										447 per 1000	**40 more per 1000** (from 143 fewer to 336 more)
										Moderate	

									447 per 1000	**40 more per 1000** (from 143 fewer to 335 more)
Underweight post-treatment – available case analysis (at-risk populations) (assessed with: number of infants who are underweight)										
79 (1 study)	serious[3]	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	6/38 (15.8%)	4/41 (9.8%)	**RR 0.62** (0.19 to 2.02)	**Study population**
									158 per 1000	**60 fewer per 1000** (from 128 fewer to 161 more)
									Moderate	
									158 per 1000	**60 fewer per 1000** (from 128 fewer to 161 more)
Overweight post-treatment – available case analysis (at-risk populations) (assessed with: number of infants who are overweight)										
79 (1 study)	serious[3]	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	15/38 (39.5%)	17/41 (41.5%)	**RR 1.05** (0.61 to 1.8)	**Study population**
									395 per 1000	**20 more per 1000** (from 154 fewer to 316 more)
									Moderate	
									395 per 1000	**20 more per 1000** (from 154 fewer to 316 more)
Incidence of severe diarrhoea post-treatment – available case analysis (at-risk populations) (assessed with: infant illness: severe diarrhoea (without dehydration))										
87 (1 study)	serious[3]	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	4/42 (9.5%)	5/45 (11.1%)	**RR 1.17** (0.34 to 4.05)	**Study population**
									95 per 1000	**16 more per 1000** (from 63 fewer to 290 more)

									Moderate	
									95 per 1000	**16 more per 1000** (from 63 fewer to 290 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Unclear risk of selection bias due to insufficient detail reported with regards to randomisation method and allocation concealment and unclear risk of detection bias as blinding of outcome assessor not reported

1.1.39 Infant regulatory problems: mother–infant relationship interventions versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant regulatory problems: mother–infant relationship interventions versus TAU**		**Risk with control**	**Risk difference with infant regulatory problems: mother–infant relationship interventions versus TAU** (95% CI)
Infant colic mean scores post-treatment – available case analysis (at-risk populations) (measured with: short Temperament Scale for Infants (STSI): Colic; better indicated by lower values)											
63 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[2]	⊕⊖⊖⊖ **VERY LOW**[1,2] due to imprecision, publication bias	31	32	-		The mean infant colic mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **1.08 standard deviations lower** (1.61 to 0.55 lower)
Infant sleep problems mean score post-treatment – available case analysis (at-risk populations) (measured with: short Temperament Scale for Infants (STSI): sleep problems; better indicated by lower values)											
63 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[2]	⊕⊖⊖⊖ **VERY LOW**[1,2] due to imprecision,	31	32	-		The mean infant sleep problems mean score post-treatment – available case analysis (at-risk populations) in the intervention groups was

						publication bias				**5.27 standard deviations lower** (6.34 to 4.2 lower)
Infant excessive crying mean scores post-treatment – available case analysis (at-risk populations) (measured with: short Temperament Scale for Infants (STSI): Excessive crying; better indicated by lower values)										
63 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[2]	⊕⊖⊖⊖ **VERY LOW**[1,2] due to imprecision, publication bias	31	32	-	The mean infant excessive crying mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **1.13 standard deviations lower** (1.67 to 0.6 lower)
Infant colic mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) (measured with: short Temperament Scale for Infants (STSI): Colic; better indicated by lower values)										
63 (1 study) 28 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[2]	⊕⊖⊖⊖ **VERY LOW**[1,2] due to imprecision, publication bias	31	32	-	The mean infant colic mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **1.72 standard deviations lower** (2.31 to 1.14 lower)
Infant sleep problems mean score short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) (measured with: short Temperament Scale for Infants (STSI): sleep problems; better indicated by lower values)										
63 (1 study) 28 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[2]	⊕⊖⊖⊖ **VERY LOW**[1,2] due to imprecision, publication bias	31	32	-	The mean infant sleep problems mean score short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.6 standard deviations lower** (1.1 to 0.09 lower)
Infant excessive crying mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) (measured with: short Temperament Scale for Infants (STSI): Excessive crying; better indicated by lower values)										

63 (1 study) 28 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,3]	reporting bias strongly suspected[2]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	31	32	-		The mean infant excessive crying mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.43 standard deviations lower** (0.93 lower to 0.07 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] Paper omits data
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.40 Infant physical development: home visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant physical development: home visits versus TAU**		**Risk with control**	**Risk difference with infant physical development: home visits versus TAU** (95% CI)
Infant motor development (delayed or impaired) post-treatment – ITT analysis (at-risk populations) (assessed with: Bayley Scales of Infant Development-Motor (scores<70))											
120 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	9/59 (15.3%)	8/61 (13.1%)	**RR 0.86** (0.36 to 2.08)	**Study population**	
										153 per 1000	**21 fewer per 1000** (from 98 fewer to 165 more)
										Moderate	
										153 per 1000	**21 fewer per 1000** (from 98 fewer to 165 more)
Infant motor development (delayed or impaired) post-treatment – available case analysis (at-risk populations) (assessed with: psychomotor Development Scale – General Development (at risk or delayed) or Bayley Scales of Infant Development-Motor (scores<70))											
	serious[1]									**Study population**	

194 (2 studies) 104 weeks		no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	8/95 (8.4%)	6/99 (6.1%)	**RR 0.73** (0.27 to 2)	**84 per 1000**	**23 fewer per 1000** (from 61 fewer to 84 more)
										Moderate	
										75 per 1000	**20 fewer per 1000** (from 55 fewer to 75 more)
Infant motor development mean scores post-treatment – available case analysis (at-risk populations) (measured with: psychomotor Development Scale – General Development or Bayley Scales of Infant Development-Motor; better indicated by lower values)											
194 (2 studies) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,4,5] due to risk of bias, imprecision, publication bias	95	99	-		The mean infant motor development mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.02 standard deviations higher** (0.26 lower to 0.3 higher)
Infant motor development (delayed or impaired) long follow-up (25-103 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: movement Assessment Battery for Children: Total motor problems (scores =<15th percentile))											
120 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	22/59 (37.3%)	24/61 (39.3%)	**RR 1.06** (0.67 to 1.66)	**Study population**	
										373 per 1000	**22 more per 1000** (from 123 fewer to 246 more)
										Moderate	
										373 per 1000	**22 more per 1000** (from 123 fewer to 246 more)
Infant motor development (delayed or impaired) long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: movement Assessment Battery for Children: Total motor problems (scores =<15th percentile))											
96 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	10/47 (21.3%)	12/49 (24.5%)	**RR 1.15** (0.55 to 2.41)	**Study population**	
										213 per 1000	**32 more per 1000** (from 96 fewer to 300 more)
										Moderate	

									213 per 1000	**32 more per 1000** (from 96 fewer to 300 more)
Infant motor development mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (measured with: movement Assessment Battery for Children: Total motor problems; better indicated by lower values)										
96 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected[4]	⨁⊖⊖⊖ **VERY LOW**[1,4,5] due to risk of bias, imprecision, publication bias	47	49	-	The mean infant motor development mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.03 standard deviations lower** (0.43 lower to 0.37 higher)

[1] High risk of selection bias due to statistcially significant baseline difference between groups with twice the number of participants showing depression symptomatology (EPDS ≥13) in the control group (N=10/17%) relative to the intervention group (N=5/8%)
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] Paper omits data
[5] Total population size is less than 400 (a threshold rule-of-thumb)

1.1.41 Infant cognitive development: home visits versus TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant cognitive development: home visits versus TAU**		**Risk with control**	**Risk difference with infant cognitive development: home visits versus TAU** (95% CI)
Infant cognitive development (impairment) post-treatment – ITT analysis (at-risk populations) (assessed with: Bayley Scales of Infant Development – Cognitive (scores<70))											
120 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⨁⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias,	9/59 (15.3%)	9/61 (14.8%)	**RR 0.97** (0.41 to 2.27)	**Study population**	
										153 per 1000	**5 fewer per 1000** (from 90 fewer to 194 more)

						imprecision, publication bias			**Moderate**	
									153 per 1000	**5 fewer per 1000** (from 90 fewer to 194 more)
Infant cognitive development (impairment) post-treatment – available case analysis (at-risk populations) (assessed with: Bayley Scales of Infant Development – Cognitive (scores<70))										
115 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	7/57 (12.3%)	6/58 (10.3%)	**RR 0.84** (0.3 to 2.35)	**Study population**
									123 per 1000	**20 fewer per 1000** (from 86 fewer to 166 more)
									Moderate	
									123 per 1000	**20 fewer per 1000** (from 86 fewer to 166 more)
Infant cognitive development mean scores post-treatment – available case analysis (at-risk populations) (measured with: Bayley Scales of Infant Development – Cognitive; better indicated by lower values)										
115 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3,5]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4,5] due to risk of bias, imprecision, publication bias	57	58	-	The mean infant cognitive development mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.27 standard deviations higher** (0.1 lower to 0.63 higher)
Infant verbal development (impairment) post-treatment – ITT analysis (at-risk populations) (assessed with: Bayley Scales of Infant Development – Language (scores<70))										
120 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	14/59 (23.7%)	15/61 (24.6%)	**RR 1.04** (0.55 to 1.95)	**Study population**
									237 per 1000	**9 more per 1000** (from 107 fewer to 225 more)
									Moderate	
									237 per 1000	**9 more per 1000** (from 107 fewer to 225 more)

Infant verbal development (impairment) post-treatment – available case analysis (at-risk populations) (assessed with: Bayley Scales of Infant Development – Language (scores<70))										
111 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	11/54 (20.4%)	11/57 (19.3%)	**RR 0.95** (0.45 to 2)	**Study population**
										204 per 1000 **10 fewer per 1000** (from 112 fewer to 204 more)
										Moderate
										204 per 1000 **10 fewer per 1000** (from 112 fewer to 204 more)
Infant verbal development mean scores post-treatment – available case analysis (at-risk populations) (measured with: Bayley Scales of Infant Development – Language; better indicated by lower values)										
111 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,4,5] due to risk of bias, imprecision, publication bias	54	57	-	The mean infant verbal development mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.09 standard deviations lower** (0.47 lower to 0.28 higher)
Infant nonverbal development (impairment) post-treatment – ITT analysis (at-risk populations) (assessed with: Differential Abilities Scale: nonverbal Reasoning composite (scores >1 SD below test mean))										
120 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	14/59 (23.7%)	18/61 (29.5%)	**RR 1.24** (0.68 to 2.27)	**Study population**
										237 per 1000 **57 more per 1000** (from 76 fewer to 301 more)
										Moderate
										237 per 1000 **57 more per 1000** (from 76 fewer to 301 more)
Infant nonverbal development (impairment) post-treatment – available case analysis (at-risk populations) (assessed with: Differential Abilities Scale: nonverbal Reasoning composite (scores >1 SD below test mean))										
	serious[1]									**Study population**

101 (1 study) 208 weeks		no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	4/49 (8.2%)	9/52 (17.3%)	**RR 2.12** (0.7 to 6.44)	**82 per 1000**	**91 more per 1000** (from 24 fewer to 444 more)
										Moderate	
										82 per 1000	**92 more per 1000** (from 25 fewer to 446 more)
Infant nonverbal development mean scores post-treatment – available case analysis (at-risk populations) (measured with: Differential Abilities Scale: nonverbal Reasoning composite; better indicated by lower values)											
101 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3,5]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4,5] due to risk of bias, imprecision, publication bias	49	52	-		The mean infant nonverbal development mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.2 standard deviations lower** (0.6 lower to 0.19 higher)
Infant spatial reasoning development (impairment) post-treatment – ITT analysis (at-risk populations) (assessed with: Differential Abilities Scale: spatial Reasoning composite (scores >1 SD below test mean))											
120 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	18/59 (30.5%)	19/61 (31.1%)	**RR 1.02** (0.6 to 1.75)	**Study population**	
										305 per 1000	**6 more per 1000** (from 122 fewer to 229 more)
										Moderate	
										305 per 1000	**6 more per 1000** (from 122 fewer to 229 more)
Infant spatial reasoning development (impairment) post-treatment – available case analysis (at-risk populations) (assessed with: Differential Abilities Scale: spatial Reasoning composite (scores >1 SD below test mean))											
99 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	8/49 (16.3%)	8/50 (16%)	**RR 0.98** (0.4 to 2.4)	**Study population**	
										163 per 1000	**3 fewer per 1000** (from 98 fewer to 229 more)
										Moderate	

									163 per 1000	**3 fewer per 1000** (from 98 fewer to 228 more)
Infant spatial reasoning development mean scores post-treatment – available case analysis (at-risk populations) (measured with: Differential Abilities Scale: spatial Reasoning composite; better indicated by lower values)										
99 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3,5]	reporting bias strongly suspected[4]	⨁⊖⊖⊖ **VERY LOW**[1,3,4,5] due to risk of bias, imprecision, publication bias	49	50	-	The mean infant spatial reasoning development mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.14 standard deviations higher** (0.26 lower to 0.53 higher)
Infant cognitive development (impairment) long follow-up (25-103 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Differential Abilities Scale: General Conceptual Ability (scores >1 SD below test mean))										
120 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⨁⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	16/59 (27.1%)	18/61 (29.5%)	**RR 1.09** (0.62 to 1.92)	**Study population**
										271 per 1000 **24 more per 1000** (from 103 fewer to 249 more)
										Moderate
										271 per 1000 **24 more per 1000** (from 103 fewer to 249 more)
Infant cognitive development (impairment) long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Differential Abilities Scale: General Conceptual Ability (scores >1 SD below test mean))										
103 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⨁⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	8/51 (15.7%)	9/52 (17.3%)	**RR 1.1** (0.46 to 2.64)	**Study population**
										157 per 1000 **16 more per 1000** (from 85 fewer to 257 more)
										Moderate
										157 per 1000 **16 more per 1000** (from 85 fewer to 257 more)

Infant cognitive development mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Differential Abilities Scale: General Conceptual Ability; better indicated by lower values)										
103 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,4,5] due to risk of bias, imprecision, publication bias	51	52	-	The mean infant cognitive development mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.09 standard deviations higher** (0.3 lower to 0.48 higher)
Infant verbal development (impairment) long follow-up (25-103 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Differential Abilities Scale: Verbal composite (scores >1 SD below test mean))										
120 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	16/59 (27.1%)	13/61 (21.3%)	**RR 0.79** (0.42 to 1.49)	**Study population** **271 per 1000** **57 fewer per 1000** (from 157 fewer to 133 more) **Moderate** **271 per 1000** **57 fewer per 1000** (from 157 fewer to 133 more)
Infant verbal development (impairment) long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Differential Abilities Scale: Verbal composite (scores >1 SD below test mean))										
104 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	9/52 (17.3%)	4/52 (7.7%)	**RR 0.44** (0.15 to 1.35)	**Study population** **173 per 1000** **97 fewer per 1000** (from 147 fewer to 61 more) **Moderate** **173 per 1000** **97 fewer per 1000** (from 147 fewer to 61 more)
Infant verbal development mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Differential Abilities Scale: Verbal composite; better indicated by lower values)										

104 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3,5]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4,5] due to risk of bias, imprecision, publication bias	52	52	-		The mean infant verbal development mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.28 standard deviations higher** (0.1 lower to 0.67 higher)

[1] High risk of selection bias due to statistically significant baseline difference between groups with twice the number of participants showing depression symptomatology (EPDS ≥13) in the control group (N=10/17%) relative to the intervention group (N=5/8%)
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] Paper omits data
[5] Total population size is less than 400 (a threshold rule-of-thumb)

1.1.42 Infant emotional development: home visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant emotional development: home visits versus TAU**		**Risk with control**	**Risk difference with infant emotional development: home visits versus TAU** (95% CI)
Infant adaptive behaviour (impairment) post-treatment – ITT analysis (at-risk populations) (assessed with: infant Toddler Social and Emotional Assessment: Competence (mean scores=<10th percentile))											
120 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	23/59 (39%)	19/61 (31.1%)	**RR 0.8** (0.49 to 1.31)	**Study population**	
										390 per 1000	**78 fewer per 1000** (from 199 fewer to 121 more)
										Moderate	
										390 per 1000	**78 fewer per 1000** (from 199 fewer to 121 more)

Infant adaptive behaviour (impairment) post-treatment – available case analysis (at-risk populations) (assessed with: infant Toddler Social and Emotional Assessment: Competence (mean scores=<10th percentile))										
97 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	15/49 (30.6%)	7/48 (14.6%)	**RR 0.48** (0.21 to 1.06)	**Study population**
										306 per 1000 / **159 fewer per 1000** (from 242 fewer to 18 more)
										Moderate
										306 per 1000 / **159 fewer per 1000** (from 242 fewer to 18 more)
Infant adaptive behaviour mean scores post-treatment – available case analysis (at-risk populations) (measured with: infant Toddler Social and Emotional Assessment: Competence; better indicated by lower values)										
99 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,4,5] due to risk of bias, imprecision, publication bias	51	48	-	The mean infant adaptive behaviour mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.49 standard deviations higher** (0.09 to 0.89 higher)
Infant emotional development (impairment) post-treatment – ITT analysis (at-risk populations) (assessed with: infant Toddler Social and Emotional Assessment: Impairment ≥1 domain)										
120 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,4] due to risk of bias, imprecision, publication bias	33/59 (55.9%)	22/61 (36.1%)	**RR 0.64** (0.43 to 0.97)	**Study population**
										559 per 1000 / **201 fewer per 1000** (from 17 fewer to 319 fewer)
										Moderate
										559 per 1000 / **201 fewer per 1000** (from 17 fewer to 319 fewer)
Infant emotional development (impairment) post-treatment – available case analysis (at-risk populations) (assessed with: infant Toddler Social and Emotional Assessment: Impairment ≥1 domain)										
	serious[1]									**Study population**

98 (1 study) 104 weeks		no serious inconsistency	no serious indirectness	very serious[2]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,4] due to risk of bias, imprecision, publication bias	25/50 (50%)	10/48 (20.8%)	**RR 0.42** (0.22 to 0.77)	**500 per 1000** **Moderate** **500 per 1000**	**290 fewer per 1000** (from 115 fewer to 390 fewer) **290 fewer per 1000** (from 115 fewer to 390 fewer)
Infant externalizing (impairment) post-treatment – ITT analysis (at-risk populations) (assessed with: infant Toddler Social and Emotional Assessment: Externalizing (mean scores ≥90th percentile))											
120 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	16/59 (27.1%)	14/61 (23%)	**RR 0.85** (0.45 to 1.58)	**Study population** **271 per 1000** **Moderate** **271 per 1000**	 **41 fewer per 1000** (from 149 fewer to 157 more) **41 fewer per 1000** (from 149 fewer to 157 more)
Infant externalizing (impairment) post-treatment – available case analysis (at-risk populations) (assessed with: infant Toddler Social and Emotional Assessment: Externalizing (mean scores ≥90th percentile))											
100 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	8/51 (15.7%)	2/49 (4.1%)	**RR 0.26** (0.06 to 1.17)	**Study population** **157 per 1000** **Moderate** **157 per 1000**	 **116 fewer per 1000** (from 147 fewer to 27 more) **116 fewer per 1000** (from 148 fewer to 27 more)
Infant externalizing mean scores post-treatment – available case analysis (at-risk populations) (measured with: infant Toddler Social and Emotional Assessment: Externalizing; better indicated by lower values)											
100 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,4,5] due to risk of bias, imprecision, publication bias	51	49	-		The mean infant externalizing mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was

										0.43 standard deviations lower (0.83 to 0.03 lower)
Infant internalizing (impairment) post-treatment – ITT analysis (at-risk populations) (assessed with: infant Toddler Social and Emotional Assessment: Internalizing (mean scores ≥90th percentile))										
120 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	12/59 (20.3%)	14/61 (23%)	**RR 1.13** (0.57 to 2.23)	**Study population**
										203 per 1000 / **26 more per 1000** (from 87 fewer to 250 more)
										Moderate
										203 per 1000 / **26 more per 1000** (from 87 fewer to 250 more)
Infant internalizing (impairment) post-treatment – available case analysis (at-risk populations) (assessed with: infant Toddler Social and Emotional Assessment: Internalizing (mean scores ≥90th percentile))										
100 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	4/51 (7.8%)	2/49 (4.1%)	**RR 0.52** (0.1 to 2.71)	**Study population**
										78 per 1000 / **38 fewer per 1000** (from 71 fewer to 134 more)
										Moderate
										78 per 1000 / **37 fewer per 1000** (from 70 fewer to 133 more)
Infant internalizing mean scores post-treatment – available case analysis (at-risk populations) (measured with: infant Toddler Social and Emotional Assessment: Internalizing; better indicated by lower values)										
100 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3,5]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4,5] due to risk of bias, imprecision, publication bias	51	49	-	The mean infant internalizing mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.15 standard deviations lower** (0.54 lower to 0.24 higher)
Infant dysregulation (impairment) post-treatment – ITT analysis (at-risk populations) (assessed with: infant Toddler Social and Emotional Assessment: Dysregulation (mean scores ≥90th percentile))										

120 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	20/59 (33.9%)	12/61 (19.7%)	**RR 0.58** (0.31 to 1.08)	**Study population**	
										339 per 1000	**142 fewer per 1000** (from 234 fewer to 27 more)
										Moderate	
										339 per 1000	**142 fewer per 1000** (from 234 fewer to 27 more)
Infant dysregulation (impairment) post-treatment – available case analysis (at-risk populations) (assessed with: infant Toddler Social and Emotional Assessment: Dysregulation (mean scores ≥90th percentile))											
100 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,4] due to risk of bias, imprecision, publication bias	12/51 (23.5%)	0/49 (0%)	**RR 0.04** (0 to 0.68)	**Study population**	
										235 per 1000	**226 fewer per 1000** (from 75 fewer to 235 fewer)
										Moderate	
										235 per 1000	**226 fewer per 1000** (from 75 fewer to 235 fewer)
Infant dysregulation mean scores post-treatment – available case analysis (at-risk populations) (measured with: infant Toddler Social and Emotional Assessment: Dysregulation; better indicated by lower values)											
100 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,4,5] due to risk of bias, imprecision, publication bias	51	49	-		The mean infant dysregulation mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.8 standard deviations lower** (1.21 to 0.39 lower)
Infant adaptive behaviour (impairment) long follow-up (25-103 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Behavioral Assessment Screener for Children: Adaptive skills (scores >1 SD below test mean))											
120 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias,	26/59 (44.1%)	22/61 (36.1%)	**RR 0.82** (0.53 to 1.27)	**Study population**	
										441 per 1000	**79 fewer per 1000** (from 207 fewer to 119 more)

						imprecision, publication bias			**Moderate**	
									441 per 1000	**79 fewer per 1000** (from 207 fewer to 119 more)
Infant adaptive behaviour (impairment) long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Behavioral Assessment Screener for Children: Adaptive skills (scores >1 SD below test mean))										
89 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	9/42 (21.4%)	8/47 (17%)	**RR 0.79** (0.34 to 1.87)	**Study population**
									214 per 1000	**45 fewer per 1000** (from 141 fewer to 186 more)
									Moderate	
									214 per 1000	**45 fewer per 1000** (from 141 fewer to 186 more)
Infant adaptive behaviour mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Behavioral Assessment Screener for Children: Adaptive skills; better indicated by lower values)										
89 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3,5]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4,5] due to risk of bias, imprecision, publication bias	42	47	-	The mean infant adaptive behaviour mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.2 standard deviations higher** (0.22 lower to 0.62 higher)
Infant externalizing (impairment) long follow-up (25-103 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Behavioral Assessment Screener for Children: Externalizing (scores >1 SD above test mean))										
120 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	24/59 (40.7%)	25/61 (41%)	**RR 1.01** (0.65 to 1.55)	**Study population**
									407 per 1000	**4 more per 1000** (from 142 fewer to 224 more)
									Moderate	

									407 per 1000	**4 more per 1000** (from 142 fewer to 224 more)
Infant externalizing (impairment) long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Behavioral Assessment Screener for Children: Externalizing (scores >1 SD above test mean))										
89 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⨁⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	7/42 (16.7%)	11/47 (23.4%)	**RR 1.4** (0.6 to 3.29)	**Study population**
									167 per 1000	**67 more per 1000** (from 67 fewer to 382 more)
									Moderate	
									167 per 1000	**67 more per 1000** (from 67 fewer to 382 more)
Infant externalizing mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Behavioral Assessment Screener for Children: Externalizing; better indicated by lower values)										
89 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected[4]	⨁⊖⊖⊖ **VERY LOW**[1,4,5] due to risk of bias, imprecision, publication bias	42	47	-	The mean infant externalizing mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.05 standard deviations lower** (0.47 lower to 0.36 higher)
Infant internalizing (impairment) long follow-up (25-103 weeks post-intervention) – ITT analysis (at-risk populations) (assessed with: Behavioral Assessment Screener for Children: Internalizing (scores >1 SD above test mean))										
120 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⨁⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	24/59 (40.7%)	21/61 (34.4%)	**RR 0.85** (0.53 to 1.35)	**Study population**
									407 per 1000	**61 fewer per 1000** (from 191 fewer to 142 more)
									Moderate	
									407 per 1000	**61 fewer per 1000** (from 191 fewer to 142 more)

Infant internalizing (impairment) long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (assessed with: Behavioral Assessment Screener for Children: Internalizing (scores >1 SD above test mean))										
88 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	7/42 (16.7%) 6/46 (13%)	**RR 0.78** (0.29 to 2.14)	**Study population**	
									167 per 1000	**37 fewer per 1000** (from 118 fewer to 190 more)
									Moderate	
									167 per 1000	**37 fewer per 1000** (from 119 fewer to 190 more)
Infant internalizing mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) (measured with: Behavioral Assessment Screener for Children: Internalizing; better indicated by lower values)										
88 (1 study) 208 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,4,5] due to risk of bias, imprecision, publication bias	42 46	-		The mean infant internalizing mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.5 standard deviations lower** (0.93 to 0.08 lower)

[1] High risk of selection bias due to statistically significant baseline difference between groups with twice the number of participants showing depression symptomatology (EPDS ≥13) in the control group (N=10/17%) relative to the intervention group (N=5/8%)
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] Paper omits data
[5] Total population size is less than 400 (a threshold rule-of-thumb)

1.1.43 Infant emotional development: mother–infant relationship interventions versus TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**
							With control	**With infant emotional development:**		**Risk with control**	**Risk difference with infant emotional development: mother–infant relationship**

								mother–infant relationship interventions versus TAU			interventions versus TAU (95% CI)
Infant social-communication development mean scores post-treatment – available case analysis (at-risk populations) (measured with: pictoral Infant Communication Scales (PICS); better indicated by lower values)											
82 (1 study) 53 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	40	42	-		The mean infant social-communication development mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **0.03 standard deviations higher** (0.4 lower to 0.47 higher)
Infant social withdrawal mean scores post-treatment – available case analysis (at-risk populations) (measured with: short Temperament Scale for Infants (STSI): Approach; better indicated by lower values)											
63 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[2,3] due to imprecision, publication bias	31	32	-		The mean infant social withdrawal mean scores post-treatment – available case analysis (at-risk populations) in the intervention groups was **1.52 standard deviations higher** (0.95 to 2.08 higher)
Infant social withdrawal mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) (measured with: short Temperament Scale for Infants (STSI): Approach; better indicated by lower values)											
63 (1 study) 28 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,4]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	31	32	-		The mean infant social withdrawal mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) in the intervention groups was **0.14 standard deviations higher** (0.36 lower to 0.63 higher)

[1] High risk of selection bias due to statistically significant baseline difference with the intervention group having more mothers with earlier preterm birth and non-Norwegian origin
[2] Total population size is less than 400 (a threshold rule-of-thumb)
[3] Paper omits data
[4] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.1.44 Prevention of neglect or abuse of the infant: home visits versus TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Prevention of neglect or abuse of the infant: home visits versus TAU**		**Risk with control**	**Risk difference with Prevention of neglect or abuse of the infant: home visits versus TAU** (95% CI)
Child protection issues post-treatment – ITT analysis (at-risk populations)											
131 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	9/63 (14.3%)	12/68 (17.6%)	**RR 1.24** (0.56 to 2.73)	**Study population**	
										143 per 1000	**34 more per 1000** (from 63 fewer to 247 more)
										Moderate	
										143 per 1000	**34 more per 1000** (from 63 fewer to 247 more)
Child removed from home post-treatment – ITT analysis (at-risk populations)											
131 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	0/63 (0%)	4/68 (5.9%)	**RR 8.35** (0.46 to 152)	**Study population**	
										0 per 1000	-
										Moderate	
										0 per 1000	-
Infant mortality post-treatment – ITT analysis (at-risk populations)											

131 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	1/63 (1.6%)	0/68 (0%)	**RR 0.31** (0.01 to 7.45)	**Study population**	
										16 per 1000	**11 fewer per 1000** (from 16 fewer to 102 more)
										Moderate	
										16 per 1000	**11 fewer per 1000** (from 16 fewer to 103 more)
Infant abuse or neglect post-treatment – available case analysis (at-risk populations)											
79 (1 study)	serious[3]	no serious inconsistency	no serious indirectness		undetected	See comment	0/38 (0%)	0/41 (0%)	**not pooled**	See comment	See comment

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Unclear risk of selection bias due to insufficient detail reported with regards to randomisation method and allocation concealment and unclear risk of detection bias due to unclear blinding of outcome assessment

1.2 PSYCHOSOCIAL INTERVENTIONS: PREVENTION (NO RISK FACTORS IDENTIFIED)

1.2.1 Depression: structured psychological interventions (CBT or IPT) versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: structured psychological interventions (CBT or IPT) versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		

Depression symptomatology post-treatment – ITT analysis (no-risk populations) (follow-up mean 26 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥12)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[1]	402/1152 (34.9%)	408/1172 (34.8%)	RR 1 (0.9 to 1.12)	0 fewer per 1000 (from 35 fewer to 42 more)	⊕⊕⊕O MODERATE	
								34.8%		0 fewer per 1000 (from 35 fewer to 42 more)		
Depression symptomatology post-treatment – available case analysis (no-risk populations) (follow-up mean 26 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥12)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[1]	98/848 (11.6%)	150/914 (16.4%)	RR 0.7 (0.56 to 0.89)	49 fewer per 1000 (from 18 fewer to 72 fewer)	⊕⊕OO LOW	
								16.4%		49 fewer per 1000 (from 18 fewer to 72 fewer)		
Depression mean scores post-treatment – available case analysis (no-risk populations) (follow-up mean 26 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[1]	848	914	-	SMD 0.22 lower (0.31 to 0.13 lower)	⊕⊕⊕O MODERATE	

[1] Paper omits data
[2] Total number of events is less than 300 (a threshold rule-of-thumb)

1.2.2 Depression: listening visits versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: listening visits versus TAU**		**Relative (95% CI)**	**Absolute**		
Depression symptomatology post-treatment – ITT analysis (no-risk populations) (follow-up mean 26 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥12)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[1]	335/1125 (29.8%)	408/1172 (34.8%)	RR 0.86 (0.76 to 0.96)	49 fewer per 1000 (from 14 fewer to 84 fewer)	⊕⊕⊕O MODERATE	
								34.8%		49 fewer per 1000 (from 14 fewer to 84 fewer)		
Depression symptomatology post-treatment – available case analysis (no-risk populations) (follow-up mean 26 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥12)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[1]	107/897 (11.9%)	150/914 (16.4%)	RR 0.73 (0.58 to 0.92)	44 fewer per 1000 (from 13 fewer to 69 fewer)	⊕⊕OO LOW	
								16.4%		44 fewer per 1000 (from 13 fewer to 69 fewer)		
Depression mean scores post-treatment – available case analysis (no-risk populations) (follow-up mean 26 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[1]	897	914	-	SMD 0.2 lower (0.3 to 0.11 lower)	⊕⊕⊕O MODERATE	

[1] Paper omits data

[2] Total number of events is less than 300 (a threshold rule-of-thumb)

1.2.3 Depression: psychologically (CBT/IPT)-informed psychoeducation versus enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Depression: psychologically (CBT/IPT)-informed psychoeducation versus enhanced TAU	Control	Relative (95% CI)	Absolute		
Depression symptomatology post-treatment – ITT analysis (no-risk populations) (follow-up 4-17 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10 or Leverton Questionnaire (LQ; Elliott et al., 2000) ≥12)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	90/879 (10.2%)	110/1099 (10%)	RR 1 (0.77 to 1.31)	0 fewer per 1000 (from 23 fewer to 31 more)	⊕⊕OO LOW	
								10.8%		0 fewer per 1000 (from 25 fewer to 33 more)		
Depression symptomatology post-treatment – available case analysis (no-risk populations) (follow-up mean 4 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	15/249 (6%)	14/251 (5.6%)	RR 1.08 (0.53 to 2.19)	4 more per 1000 (from 26 fewer to 66 more)	⊕⊕OO LOW	
								5.6%		4 more per 1000 (from 26 fewer to 67 more)		
Depression symptomatology short follow-up (9-16 weeks post-intervention) – ITT analysis (no-risk populations) (follow-up mean 12 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10)												
1	randomised trials	no serious	no serious inconsistency	no serious indirectness	very serious[1,2]	none	47/270 (17.4%)	53/270 (19.6%)		22 fewer per 1000 (from 75	⊕⊕OO LOW	

		risk of bias							RR 0.89 (0.62 to 1.26)	fewer to 51 more)		
								19.6%		22 fewer per 1000 (from 74 fewer to 51 more)		
Depression symptomatology short follow-up (9-16 weeks post-intervention) – available case analysis (no-risk populations) (follow-up mean 12 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	12/235 (5.1%)	15/232 (6.5%)	RR 0.79 (0.38 to 1.65)	14 fewer per 1000 (from 40 fewer to 42 more)	⊕⊕OO LOW	
								6.5%		14 fewer per 1000 (from 40 fewer to 42 more)		
Depression symptomatology intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (no-risk populations) (follow-up mean 25 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	48/270 (17.8%)	43/270 (15.9%)	RR 1.12 (0.77 to 1.62)	19 more per 1000 (from 37 fewer to 99 more)	⊕⊕OO LOW	
								15.9%		19 more per 1000 (from 37 fewer to 99 more)		
Depression symptomatology intermediate follow-up (17-24 weeks post-intervention) – available case analysis (no-risk populations) (follow-up mean 25 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10)												
1	randomised trials	no serious	no serious inconsistency	no serious indirectness	very serious[1,2]	none	8/230 (3.5%)	11/238 (4.6%)		12 fewer per 1000 (from 32	⊕⊕OO LOW	

		risk of bias							RR 0.75 (0.31 to 1.84)	fewer to 39 more)		
								4.6%		12 fewer per 1000 (from 32 fewer to 39 more)		

[1] Total number of events is less than 300 (a threshold rule-of-thumb)

[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.2.4 Depression: home visits versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: home visits versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression mean scores post-treatment – available case analysis (no-risk populations) (follow-up mean 6 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	276	266	-	SMD 0.13 higher (0.04 lower to 0.3 higher)	⊕⊕⊕O MODERATE	
Depression mean scores intermediate follow-up (17-24 weeks post-intervention) – available case analysis (no-risk populations) (follow-up mean 26 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	252	229	-	SMD 0.02 lower (0.2 lower to 0.16 higher)	⊕⊕⊕O MODERATE	

[1] Risk of bias due to statistically significant group differences at baseline

1.2.5 Depression: post-delivery discussion versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Depression: post-delivery discussion versus TAU	Control	Relative (95% CI)	Absolute		
Depression symptomatology post-treatment – available case analysis (no-risk populations) (follow-up mean 3 weeks; assessed with: Hospital Anxiety and Depression Scale – Depression (HADS ≥11))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	5/58 (8.6%)	31/56 (55.4%)	RR 0.16 (0.07 to 0.37)	465 fewer per 1000 (from 349 fewer to 515 fewer)	⊕⊕OO LOW	
								55.4%		465 fewer per 1000 (from 349 fewer to 515 fewer)		

[1] Total number of events is less than 300 (a threshold rule-of-thumb)

1.2.6 Depression: mother–infant relationship interventions versus enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Depression: mother–infant relationship interventions versus enhanced TAU	Control	Relative (95% CI)	Absolute		
Depression mean scores post-treatment – ITT analysis (no-risk populations) (follow-up mean 26 weeks; measured with: Beck Depression Inventory (BDI); better indicated by lower values)												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	26	28	-	SMD 0.27 lower (0.81 lower to 0.26 higher)	⊕⊕OO LOW	
Depression mean scores post-treatment – available case analysis (no-risk populations) (follow-up mean 26 weeks; measured with: Beck Depression Inventory (BDI); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	26	28	-	SMD 0.27 lower (0.81 lower to 0.26 higher)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.2.7 Depression: mindfulness training versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: mindfulness training versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression mean scores post-treatment – available case analysis (no-risk populations) (follow-up mean 11 weeks; measured with: Depression, Anxiety, and Stress Scale (DASS-21): Depression; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	13	8	-	SMD 0.36 lower (1.25 lower to 0.53 higher)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.2.8 Anxiety: structured psychological interventions (CBT or IPT) versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anxiety: structured psychological interventions (CBT or IPT) versus TAU	Control	Relative (95% CI)	Absolute		
Anxiety mean scores post-treatment – available case analysis (no-risk populations) (follow-up mean 26 weeks; measured with: state-Trait Anxiety Inventory (STAI)-State; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[1]	795	858	-	SMD 0.13 lower (0.23 to 0.04 lower)	⊕⊕⊕O MODERATE	
Trait anxiety mean scores post-treatment – available case analysis (no-risk populations) (follow-up mean 26 weeks; measured with: state-Trait Anxiety Inventory (STAI) – Trait; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[1]	779	839	-	SMD 0.12 lower (0.22 to 0.02 lower)	⊕⊕⊕O MODERATE	

[1] Paper omits data

1.2.9 Anxiety: listening visits versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anxiety: listening visits versus TAU	Control	Relative (95% CI)	Absolute		

Anxiety mean scores post-treatment – available case analysis (no-risk populations) (follow-up mean 26 weeks; measured with: state-Trait Anxiety Inventory (STAI)-State; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[1]	839	858	-	SMD 0.1 lower (0.19 lower to 0 higher)	⊕⊕⊕O MODERATE	
Trait anxiety mean scores post-treatment – available case analysis (no-risk populations) (follow-up mean 26 weeks; measured with: state-Trait Anxiety Inventory (STAI) – Trait; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[1]	856	839	-	SMD 0.11 lower (0.2 to 0.01 lower)	⊕⊕⊕O MODERATE	

[1] Paper omits data

1.2.10 Anxiety: post-delivery discussion versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Anxiety: post-delivery discussion versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Anxiety symptomatology post-treatment – available case analysis (no-risk populations) (follow-up mean 3 weeks; assessed with: Hospital Anxiety and Depression Scale – Anxiety (HADS ≥11))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	4/58 (6.9%)	28/56 (50%)	RR 0.14 (0.05 to 0.37)	430 fewer per 1000 (from 315 fewer to 475 fewer)	⊕⊕OO LOW	
								50%		430 fewer per 1000 (from 315 fewer to 475 fewer)		

[1] Total number of events is less than 300 (a threshold rule-of-thumb)

1.2.11 Anxiety: music therapy versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anxiety: music therapy versus TAU	Control	Relative (95% CI)	Absolute		
Anxiety mean scores post-treatment – available case analysis (no-risk populations) (follow-up mean 2 weeks; measured with: state-Trait Anxiety Inventory (STAI)-State; better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	37	40	-	SMD 0.42 higher (0.04 lower to 0.87 higher)	⊕OOO VERY LOW	

[1] Risk of bias due to statistically significant group differences at baseline
[2] Total population size is less than 400 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.2.12 Anxiety: mindfulness training versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anxiety: mindfulness training versus TAU	Control	Relative (95% CI)	Absolute		
Anxiety mean scores post-treatment – available case analysis (no-risk populations) (follow-up mean 11 weeks; measured with: Depression, Anxiety, and Stress Scale (DASS-21): Anxiety; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	13	8	-	SMD 1.21 lower (2.18 to 0.24 lower)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)

1.2.13 General mental health: structured psychological interventions (CBT or IPT) versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With General mental health: structured psychological interventions (CBT or IPT) versus TAU		Risk with control	Risk difference with General mental health: structured psychological interventions (CBT or IPT) versus TAU (95% CI)
General mental health mean scores post-treatment – available case analysis (no-risk populations) (measured with: sF-12 mental component summary (SF-MCS); better indicated by lower values)											
1700 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[1]	⨁⨁⨁⊖ **MODERATE**[1] due to publication bias	885	815	-		The mean general mental health mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0.15 standard deviations higher** (0.06 to 0.25 higher)
Risk of self-harm mean scores post-treatment – available case analysis (no-risk populations) (measured with: Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM): Risk of self-harm; better indicated by lower values)											
1749 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[1]	⨁⨁⨁⊖ **MODERATE**[1] due to publication bias	906	843	-		The mean risk of self-harm mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0.66 standard deviations lower** (0.75 to 0.56 lower)

[1] Paper omits data

1.2.14 General mental health: listening visits versus TAU

Quality assessment	Summary of findings

Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With General mental health: listening visits versus TAU		Risk with control	Risk difference with General mental health: listening visits versus TAU (95% CI)
General mental health mean scores post-treatment – available case analysis (no-risk populations) (measured with: sF-12 mental component summary (SF-MCS); better indicated by lower values)											
1764 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[1]	⊕⊕⊕⊖ **MODERATE**[1] due to publication bias	885	879	-		The mean general mental health mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0.15 standard deviations higher** (0.06 to 0.25 higher)
Risk of self-harm mean scores post-treatment – available case analysis (no-risk populations) (measured with: Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM): Risk of self-harm; better indicated by lower values)											
1799 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[1]	⊕⊕⊕⊖ **MODERATE**[1] due to publication bias	906	893	-		The mean risk of self-harm mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0.57 standard deviations higher** (0.47 to 0.66 higher)

[1] Paper omits data

1.2.15 General mental health: home visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With General mental health: home visits versus TAU		Risk with control	Risk difference with General mental health: home visits versus TAU (95% CI)

General mental health mean scores post-treatment – available case analysis (no-risk populations) (measured with: sF-36 – Mental health; better indicated by lower values)										
550 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to risk of bias	268	282	-	The mean general mental health mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0.04 standard deviations lower** (0.21 lower to 0.13 higher)
General mental health mean scores Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (no-risk populations) (measured with: sF-36 – Mental health; better indicated by lower values)										
481 (1 study) 26 weeks	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to risk of bias	227	254	-	The mean general mental health mean scores intermediate follow-up (17-24 weeks post-intervention) – available case analysis (no-risk populations) in the intervention groups was **0.07 standard deviations lower** (0.25 lower to 0.11 higher)

[1] High risk of selection bias due to statistically significant baseline group differences for incidence of twins, use of TENS during labour, and adults living with the mother

1.2.16 General mental health: mindfulness training versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With General mental health: mindfulness training versus TAU**		**Risk with control**	**Risk difference with General mental health: mindfulness training versus TAU** (95% CI)
Psychological distress mean scores post-treatment – available case analysis (no-risk populations) (measured with: Depression, Anxiety, and Stress Scale (DASS-21): psychological distress; better indicated by lower values)											
21 (1 study) 11 weeks	no serious	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1]	8	13	-		The mean psychological distress mean scores post-treatment – available case analysis (no-risk

	risk of bias					due to imprecision				populations) in the intervention groups was **1.15 standard deviations lower** (2.11 to 0.19 lower)
Life satisfaction mean scores post-treatment – available case analysis (no-risk populations) (measured with: satisfaction With Life Scale (SWLS); better indicated by lower values)										
21 (1 study) 11 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	8	13	-	The mean life satisfaction mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0.43 standard deviations higher** (0.46 lower to 1.32 higher)
Happiness mean scores post-treatment – available case analysis (no-risk populations) (measured with: subjective Happiness Scale; better indicated by lower values)										
21 (1 study) 11 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	8	13	-	The mean happiness mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0.24 standard deviations higher** (0.65 lower to 1.12 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.2.17 Mother–infant attachment: home visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Mother–infant attachment: home visits versus TAU**		**Risk with control**	**Risk difference with Mother–infant attachment: home visits versus TAU** (95% CI)
Discontinued breastfeeding by 6 weeks – available case analysis (no-risk populations)											

548 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to risk of bias	155/268 (57.8%)	154/280 (55%)	**RR 0.95** (0.82 to 1.1)	**Study population**	
										578 per 1000	**29 fewer per 1000** (from 104 fewer to 58 more)
										Moderate	
										578 per 1000	**29 fewer per 1000** (from 104 fewer to 58 more)
Discontinued breastfeeding by 26 weeks – available case analysis (no-risk populations)											
493 (1 study) 26 weeks	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to risk of bias	185/233 (79.4%)	208/260 (80%)	**RR 1.01** (0.92 to 1.1)	**Study population**	
										794 per 1000	**8 more per 1000** (from 64 fewer to 79 more)
										Moderate	
										794 per 1000	**8 more per 1000** (from 64 fewer to 79 more)

[1] High risk of selection bias due to statistically significant baseline group differences for incidence of twins, use of TENS during labour, and adults living with the mother

1.2.18 Mother–infant attachment: mother–infant relationship interventions versus enhanced TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Mother–infant attachment: mother–infant relationship interventions versus enhanced TAU**		**Risk with control**	**Risk difference with Mother–infant attachment: mother–infant relationship interventions versus enhanced TAU** (95% CI)
Maternal sensitivity mean scores post-treatment – ITT analysis (no-risk populations) (measured with: Ainsworth Strange Situation: Total; better indicated by lower values)											

54 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	28	26	-		The mean maternal sensitivity mean scores post-treatment – ITT analysis (no-risk populations) in the intervention groups was **0.77 standard deviations higher** (0.21 to 1.32 higher)
Maternal sensitivity mean scores post-treatment – available case analysis (no-risk populations) (measured with: Ainsworth Strange Situation: Total; better indicated by lower values)											
54 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	28	26	-		The mean maternal sensitivity mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0.77 standard deviations higher** (0.21 to 1.32 higher)
Child attachment security mean scores post-treatment – ITT analysis (no-risk populations) (measured with: Waters' Attachment Q-set; better indicated by lower values)											
54 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	28	26	-		The mean child attachment security mean scores post-treatment – ITT analysis (no-risk populations) in the intervention groups was **0 standard deviations higher** (0.53 lower to 0.53 higher)
Child attachment security mean scores post-treatment – available case analysis (no-risk populations) (measured with: Waters' Attachment Q-set; better indicated by lower values)											
54 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	28	26	-		The mean child attachment security mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0 standard deviations higher** (0.53 lower to 0.53 higher)
Maternal confidence/competence mean scores post-treatment – ITT analysis (no-risk populations) (measured with: parental Efficacy Questionnaire ; better indicated by lower values)											

54 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	28	26	-		The mean maternal confidence/competence mean scores post-treatment – ITT analysis (no-risk populations) in the intervention groups was **0.3 standard deviations higher** (0.24 lower to 0.84 higher)
Maternal confidence/competence mean scores post-treatment – available case analysis (no-risk populations) (measured with: parental Efficacy Questionnaire ; better indicated by lower values)											
54 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	28	26	-		The mean maternal confidence/competence mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0.3 standard deviations higher** (0.24 lower to 0.84 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.2.19 Mother–infant attachment: mindfulness training versus TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Mother–infant attachment: mindfulness training versus TAU**		**Risk with control**	**Risk difference with Mother–infant attachment: mindfulness training versus TAU** (95% CI)
Maternal confidence/competence mean scores post-treatment – available case analysis (no-risk populations) (measured with: parental Evaluation Scale: maternal self-efficacy; better indicated by lower values)											
21 (1 study) 11 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	8	13	-		The mean maternal confidence/competence mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was

										1.59 standard deviations higher (0.56 to 2.62 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)

1.2.20 Quality of life: structured psychological interventions (CBT or IPT) versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Quality of life: structured psychological interventions (CBT or IPT) versus TAU**		**Risk with control**	**Risk difference with Quality of life: structured psychological interventions (CBT or IPT) versus TAU** (95% CI)
Parental stress mean scores post-treatment – available case analysis (no-risk populations) (measured with: parenting Stress Index (PSI); better indicated by lower values)											
1299 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[1]	⊕⊕⊕⊖ **MODERATE**[1] due to publication bias	698	601	-		The mean parental stress mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0.12 standard deviations higher** (0.01 to 0.23 higher)
Impaired functioning mean scores post-treatment – available case analysis (no-risk populations) (measured with: Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM): Life functioning; better indicated by lower values)											
1747 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[1]	⊕⊕⊕⊖ **MODERATE**[1] due to publication bias	905	842	-		The mean impaired functioning mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0.09 standard deviations lower** (0.18 lower to 0.01 higher)

Wellbeing mean scores post-treatment – available case analysis (no-risk populations) (measured with: Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM): Well-being; better indicated by lower values)										
1749 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[1]	⊕⊕⊕⊖ **MODERATE**[1] due to publication bias	907	842	-		The mean wellbeing mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0.15 standard deviations lower** (0.25 to 0.06 lower)

[1] Paper omits data

1.2.21 Quality of life: listening visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Quality of life: listening visits versus TAU**		**Risk with control**	**Risk difference with Quality of life: listening visits versus TAU** (95% CI)
Parental stress mean scores post-treatment – available case analysis (no-risk populations) (measured with: parenting Stress Index (PSI); better indicated by lower values)											
1407 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[1]	⊕⊕⊕⊖ **MODERATE**[1] due to publication bias	698	709	-		The mean parental stress mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0.17 standard deviations higher** (0.06 to 0.27 higher)
Impaired functioning mean scores post-treatment – available case analysis (no-risk populations) (measured with: Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM): Life functioning; better indicated by lower values)											
1798 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[1]	⊕⊕⊕⊖ **MODERATE**[1] due to publication bias	905	893	-		The mean impaired functioning mean scores post-treatment – available case analysis (no-risk populations) in the intervention

										groups was **0.08 standard deviations lower** (0.18 lower to 0.01 higher)
Wellbeing mean scores post-treatment – available case analysis (no-risk populations) (measured with: Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM): Well-being; better indicated by lower values)										
1800 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[1]	⊕⊕⊕⊖ **MODERATE**[1] due to publication bias	907	893	-	The mean wellbeing mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0.15 standard deviations lower** (0.24 to 0.05 lower)

[1] Paper omits data

1.2.22 Quality of life: home visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Quality of life: home visits versus TAU**		**Risk with control**	**Risk difference with Quality of life: home visits versus TAU** (95% CI)
Social support mean scores post-treatment – available case analysis (no-risk populations) (measured with: Duke Functional Social Support; better indicated by lower values)											
513 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to risk of bias	253	260	-		The mean social support mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0.01 standard deviations higher** (0.16 lower to 0.19 higher)
Social support mean scores Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (no-risk populations) (measured with: Duke Functional Social Support; better indicated by lower values)											

465 (1 study) 26 weeks	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to risk of bias	225	240	-	The mean social support mean scores intermediate follow-up (17-24 weeks post-intervention) – available case analysis (no-risk populations) in the intervention groups was **0.06 standard deviations higher** (0.13 lower to 0.24 higher)

[1] High risk of selection bias due to statistically significant baseline group differences for incidence of twins, use of TENS during labour, and adults living with the mother

1.2.23 Quality of life: mother–infant relationship interventions versus enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Quality of life: mother–infant relationship interventions versus enhanced TAU**		**Risk with control**	**Risk difference with Quality of life: mother–infant relationship interventions versus enhanced TAU** (95% CI)
Parental stress mean scores post-treatment – ITT analysis (no-risk populations) (measured with: Daily Hassles Scale: Intensity ; better indicated by lower values)											
54 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	28	26	-		The mean parental stress mean scores post-treatment – ITT analysis (no-risk populations) in the intervention groups was **0.4 standard deviations lower** (0.94 lower to 0.14 higher)
Parental stress mean scores post-treatment – available case analysis (no-risk populations) (measured with: Daily Hassles Scale: Intensity ; better indicated by lower values)											
54 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	28	26	-		The mean parental stress mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was

										0.4 standard deviations lower (0.94 lower to 0.14 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.2.24 Quality of life: music therapy versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Quality of life: music therapy versus TAU**		**Risk with control**	**Risk difference with Quality of life: music therapy versus TAU** (95% CI)
Parental stress mean scores post-treatment – available case analysis (no-risk populations) (measured with: perceived Stress Scale; better indicated by lower values)											
77 (1 study) 2 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	40	37	-		The mean parental stress mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **0.15 standard deviations higher** (0.3 lower to 0.6 higher)

[1] High risk of selection bias due to unclear allocation concealment and statistically significant group difference at baseline in education (intervention group were more highly educated than control group)
[2] Total population size is less than 400 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.2.25 Quality of life: mindfulness training versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Quality of life: mindfulness training versus TAU**		**Risk with control**	**Risk difference with Quality of life: mindfulness training versus TAU** (95% CI)

Parental stress mean scores post-treatment – available case analysis (no-risk populations) (measured with: Depression, Anxiety, and Stress Scale (DASS-21): stress; better indicated by lower values)										
21 (1 study) 11 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	8	13	-		The mean parental stress mean scores post-treatment – available case analysis (no-risk populations) in the intervention groups was **1.14 standard deviations lower** (2.1 to 0.18 lower)

[1] Total population size is less than 400 (a threshold rule-of-thumb)

1.2.26 Attrition: structured psychological interventions (CBT or IPT) versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With attrition: structured psychological interventions (CBT or IPT) versus TAU**		**Risk with control**	**Risk difference with attrition: structured psychological interventions (CBT or IPT) versus TAU** (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
2324 (1 study)	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[1]	⊕⊕⊕⊖ **MODERATE**[1] due to publication bias	177/1172 (15.1%)	227/1152 (19.7%)	**RR 1.3** (1.09 to 1.56)	**Study population**	
										151 per 1000	**45 more per 1000** (from 14 more to 85 more)
										Moderate	
										151 per 1000	**45 more per 1000** (from 14 more to 85 more)

[1] Paper omits data

1.2.27 Attrition: listening visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With attrition: listening visits versus TAU		Risk with control	Risk difference with attrition: listening visits versus TAU (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
2297 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[1]	⊕⊕⊕⊖ **MODERATE**[1] due to publication bias	177/1172 (15.1%)	170/1125 (15.1%)	**RR 1** (0.82 to 1.21)	**Study population**	
										151 per 1000	**0 fewer per 1000** (from 27 fewer to 32 more)
										Moderate	
										151 per 1000	**0 fewer per 1000** (from 27 fewer to 32 more)

[1] Paper omits data

1.2.28 Attrition: psychologically (CBT/IPT)-informed psychoeducation versus enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With attrition: psychologically (CBT/IPT)-informed psychoeducation versus enhanced TAU		Risk with control	Risk difference with attrition: psychologically (CBT/IPT)-informed psychoeducation versus enhanced TAU (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
				serious[1]	undetected					**Study population**	

540 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness			⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	19/270 (7%)	21/270 (7.8%)	**RR 1.11** (0.61 to 2.01)	**70 per 1000**	**8 more per 1000** (from 27 fewer to 71 more)
										Moderate	
										70 per 1000	**8 more per 1000** (from 27 fewer to 71 more)

[1] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.2.29 Attrition: home visits versus TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With attrition: home visits versus TAU**		**Risk with control**	**Risk difference with attrition: home visits versus TAU** (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
623 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	43/312 (13.8%)	29/311 (9.3%)	**RR 0.68** (0.43 to 1.05)	**Study population**	
										138 per 1000	**44 fewer per 1000** (from 79 fewer to 7 more)
										Moderate	
										138 per 1000	**44 fewer per 1000** (from 79 fewer to 7 more)

[1] High risk of selection bias due to statistically significant baseline group differences for incidence of twins, use of TENS during labour, and adults living with the mother
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.2.30 Attrition: mindfulness training versus TAU

Quality assessment							Summary of findings		
		Inconsistency	**Indirectness**	**Imprecision**			**Study event rates (%)**		**Anticipated absolute effects**

Participants (studies) Follow-up	Risk of bias				Publication bias	Overall quality of evidence	With control	With attrition: mindfulness training versus TAU	Relative effect (95% CI)	Risk with control	Risk difference with attrition: mindfulness training versus TAU (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
26 (1 study) 11 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	0/13 (0%)	5/13 (38.5%)	**RR 11** (0.67 to 180.65)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3 PSYCHOSOCIAL INTERVENTIONS: TREATMENT

1.3.1 Depression: structured psychological interventions (CBT or IPT) versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Depression: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU	Control	Relative (95% CI)	Absolute		
Depression diagnosis post-treatment – ITT analysis (follow-up 12-44 weeks; assessed with: structured Clinical Interview (SCID) or Clinical Interview Schedule – Revised (CIS-R))												
6	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	220/663 (33.2%)	420/644 (65.2%)	RR 0.48 (0.39 to 0.6)	339 fewer per 1000 (from 261 fewer to 398 fewer)	⊕⊕⊕⊕ HIGH	

								68.7%		357 fewer per 1000 (from 275 fewer to 419 fewer)		
Depression diagnosis post-treatment – available case analysis (follow-up 12-44 weeks; assessed with: structured Clinical Interview (SCID) or Clinical Interview Schedule – Revised (CIS-R))												
5	randomised trials	no serious risk of bias	very serious[1]	no serious indirectness	no serious imprecision	none	135/543 (24.9%)	315/523 (60.2%)	RR 0.38 (0.24 to 0.58)	373 fewer per 1000 (from 253 fewer to 458 fewer)	⊕⊕OO LOW	
								61.5%		381 fewer per 1000 (from 258 fewer to 467 fewer)		
Depression symptomatology post-treatment – ITT analysis (follow-up 6-44 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10/EPDS ≥12/Treatment non-response (baseline-endpoint decrease<4 points and EPDS >13)/Treatment non-response (<50% improvement) or Beck Depression Inventory (BDI) ≥16 or Beck Depression Inventory-II (BDI-II) ≥14)												
10	randomised trials	no serious risk of bias	serious[2]	no serious indirectness	no serious imprecision	reporting bias[3]	251/512 (49%)	294/457 (64.3%)	RR 0.69 (0.56 to 0.85)	199 fewer per 1000 (from 96 fewer to 283 fewer)	⊕⊕OO LOW	
								62.6%		194 fewer per 1000 (from 94 fewer to 275 fewer)		
Depression symptomatology post-treatment – available case analysis (follow-up 6-16 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10/EPDS ≥12/Treatment non-response (baseline-endpoint decrease<4 points and EPDS >13) or Beck Depression Inventory (BDI) ≥16 or Beck Depression Inventory-II (BDI-II) ≥14)												
9	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	121/357 (33.9%)	193/345 (55.9%)	RR 0.62 (0.53 to 0.73)	213 fewer per 1000 (from 151 fewer to 263 fewer)	⊕⊕⊕⊕ HIGH	

								58.8%		223 fewer per 1000 (from 159 fewer to 276 fewer)		
Depression mean scores post-treatment – ITT analysis (follow-up 6-44 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS) or Beck Depression Inventory (BDI-II); better indicated by lower values)												
5	randomised trials	no serious risk of bias	very serious[1]	no serious indirectness	serious[4]	none	164	142	-	SMD 1.31 lower (2.36 to 0.26 lower)	⊕OOO VERY LOW	
Depression mean scores post-treatment – available case analysis (follow-up 6-16 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS) or Beck Depression Inventory (BDI) or Beck Depression Inventory (BDI-II) or Hamilton Rating Scale for Depression (HRSD); better indicated by lower values)												
10	randomised trials	no serious risk of bias	serious[2]	no serious indirectness	no serious imprecision	none	763	745	-	SMD 0.6 lower (0.8 to 0.4 lower)	⊕⊕⊕O MODERATE	
Depression diagnosis short follow-up (9-16 weeks post-intervention) – ITT analysis (follow-up mean 28 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	none	8/47 (17%)	20/46 (43.5%)	RR 0.39 (0.19 to 0.8)	265 fewer per 1000 (from 87 fewer to 352 fewer)	⊕⊕OO LOW	
								43.5%		265 fewer per 1000 (from 87 fewer to 352 fewer)		
Depression symptomatology short follow-up (9-16 weeks post-intervention) – ITT analysis (follow-up mean 29 weeks; assessed with: Beck Depression Inventory-II (BDI-II) ≥14)												
1	randomised trials	no serious	no serious inconsistency	no serious indirectness	very serious[5,6]	none	15/30 (50%)	14/25 (56%)		62 fewer per 1000 (from 258	⊕⊕OO LOW	

		risk of bias							RR 0.89 (0.54 to 1.47)	fewer to 263 more)		
								56%		62 fewer per 1000 (from 258 fewer to 263 more)		
Depression symptomatology short follow-up (9-16 weeks post-intervention) – available case analysis (follow-up mean 29 weeks; assessed with: Beck Depression Inventory-II (BDI-II) ≥14)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	none	8/21 (38.1%)	14/21 (66.7%)	RR 0.57 (0.31 to 1.07)	287 fewer per 1000 (from 460 fewer to 47 more)	⊕⊕OO LOW	
								66.7%		287 fewer per 1000 (from 460 fewer to 47 more)		
Depression mean scores short follow-up (9-16 weeks post-intervention) – ITT analysis (follow-up 28-29 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS) or Beck Depression Inventory (BDI-II); better indicated by lower values)												
2	randomised trials	no serious risk of bias	very serious[1]	no serious indirectness	very serious[4,6]	none	77	71	-	SMD 1.84 lower (4.31 lower to 0.64 higher)	⊕OOO VERY LOW	
Depression mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (follow-up 21-29 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS) or Beck Depression Inventory (BDI-II); better indicated by lower values)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	none	43	46	-	SMD 0.66 lower (1.14 to 0.18 lower)	⊕⊕OO LOW	

Depression diagnosis Intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (follow-up mean 33 weeks; assessed with: Clinical Interview Schedule – Revised (CIS-R) or Structured Clinical Interview (SCID))												
2	randomised trials	no serious risk of bias	serious[7]	no serious indirectness	very serious[5,6]	none	21/68 (30.9%)	33/70 (47.1%)	RR 0.59 (0.24 to 1.41)	193 fewer per 1000 (from 358 fewer to 193 more)	⊕OOO VERY LOW	
								57.2%		235 fewer per 1000 (from 435 fewer to 235 more)		
Depression diagnosis intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 33 weeks; assessed with: Clinical Interview Schedule – Revised (CIS-R) or Structured Clinical Interview (SCID))												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5,6]	none	12/59 (20.3%)	22/59 (37.3%)	RR 0.5 (0.23 to 1.08)	186 fewer per 1000 (from 287 fewer to 30 more)	⊕⊕OO LOW	
								47.4%		237 fewer per 1000 (from 365 fewer to 38 more)		
Depression mean depression scores intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 33 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
2	randomised trials	no serious risk of bias	very serious[1]	no serious indirectness	very serious[4,6]	none	59	59	-	SMD 0.51 lower (1.72 lower to 0.7 higher)	⊕OOO VERY LOW	
Depression diagnosis long follow-up (25-103 weeks post-intervention) – ITT analysis (follow-up mean 78 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious	no serious inconsistency	no serious indirectness	very serious[5,6]	none	21/50 (42%)	13/52 (25%)		170 more per 1000 (from 13	⊕⊕OO LOW	

		risk of bias							RR 1.68 (0.95 to 2.98)	fewer to 495 more)		
								25%		170 more per 1000 (from 13 fewer to 495 more)		
Depression diagnosis long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 78 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5,6]	none	12/41 (29.3%)	9/48 (18.8%)	RR 1.56 (0.73 to 3.33)	105 more per 1000 (from 51 fewer to 437 more)	⊕⊕OO LOW	
								18.8%		105 more per 1000 (from 51 fewer to 438 more)		
Depression symptomatology long follow-up (25-103 weeks post-intervention) – ITT analysis (follow-up mean 32 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10)												
1	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	very serious[5,6]	none	3/17 (17.6%)	5/20 (25%)	RR 0.71 (0.2 to 2.53)	73 fewer per 1000 (from 200 fewer to 382 more)	⊕OOO VERY LOW	
								25%		73 fewer per 1000 (from 200 fewer to 382 more)		
Depression symptomatology long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 32 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10)												
1	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	very serious[5,6]	none	1/15 (6.7%)	3/18 (16.7%)		100 fewer per 1000 (from 158	⊕OOO VERY LOW	

									RR 0.4 (0.05 to 3.46)	fewer to 410 more)		
								16.7%		100 fewer per 1000 (from 159 fewer to 411 more)		
Depression mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up 32-78 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS) or Beck Depression Inventory (BDI); better indicated by lower values)												
3	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4,6]	none	68	74	-	SMD 0.28 lower (0.8 lower to 0.23 higher)	⊕⊕OO LOW	
Depression diagnosis Very long follow-up (>104 weeks post-intervention) – ITT analysis (follow-up mean 260 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	none	24/50 (48%)	13/52 (25%)	RR 1.92 (1.11 to 3.33)	230 more per 1000 (from 28 more to 582 more)	⊕⊕OO LOW	
								25%		230 more per 1000 (from 28 more to 582 more)		
Depression diagnosis Very long follow-up (>104 weeks post-intervention) – available case analysis (follow-up mean 260 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5,6]	none	7/33 (21.2%)	9/37 (24.3%)	RR 0.87 (0.37 to 2.08)	32 fewer per 1000 (from 153 fewer to 263 more)	⊕⊕OO LOW	
								24.3%		32 fewer per 1000 (from 153 fewer to 262 more)		

Depression mean depression scores Very long follow-up (>104 weeks post-intervention) – available case analysis (follow-up mean 260 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4,6]	none	28	34	-	SMD 0.17 lower (0.67 lower to 0.33 higher)	⊕⊕OO LOW	
Negative thoughts/mood mean scores – available case analysis (follow-up mean 4 weeks; measured with: Automatic Thought Questionnaire (ATQ); better indicated by lower values)												
1	randomised trials	serious[8]	no serious inconsistency	no serious indirectness	very serious[4]	none	10	12	-	SMD 0.94 lower (1.83 to 0.04 lower)	⊕OOO VERY LOW	

[1] There was evidence of considerable heterogeneity between effect sizes
[2] There was evidence of moderate to substantial heterogeneity between effect sizes
[3] Papers omit data
[4] Total population size is less than 400 (a threshold rule-of-thumb)
[5] Total number of events is less than 300 (a threshold rule-of-thumb)
[6] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[7] There was evidence of substantial heterogeneity between effect sizes
[8] Risk of bias due to statistically significant group differences at baseline

1.3.2 Depression: CBT versus listening visits

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: CBT versus listening visits**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression mean scores post-treatment – available case analysis (follow-up mean 26 weeks; measured with: Beck Depression Inventory (BDI) or Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												

2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious	reporting bias[1]	157	144	-	SMD 0.06 lower (0.33 lower to 0.22 higher)	⊕⊕OO LOW	

[1] Papers omit data

1.3.3 Depression: CBT versus relational constructivist therapy

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: CBT versus relational constructivist therapy**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression mean scores post-treatment – available case analysis (measured with: Beck Depression Inventory (BDI); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias[2]	32	28	-	SMD 0.53 higher (0.01 to 1.05 higher)	⊕OOO VERY LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] Paper omits data

1.3.4 Depression: IPT versus support group

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: IPT versus support group**	**Control**	**Relative (95% CI)**	**Absolute**		

Depression mean scores post-treatment – available case analysis (follow-up mean 12 weeks; measured with: Center for Epidemiologic Studies Depression Scale (CES-D); better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	22	22	-	SMD 0.49 lower (1.09 lower to 0.11 higher)	⊕OOO VERY LOW	

[1] Risk of bias due to statistically significant group differences at baseline
[2] Total population size is less than 400 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.5 Depression: facilitated self-help versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Depression: facilitated self-help versus TAU	Control	Relative (95% CI)	Absolute		
Depression symptomatology post-treatment – ITT Analysis (follow-up 15-20 weeks; assessed with: Beck Depression Inventory-II (BDI-II) ≥14 or Edinburgh postnatal Depression Scale (EPDS) >12)												
3	randomised trials	no serious risk of bias	very serious[1]	no serious indirectness	no serious imprecision	reporting bias[2]	399/574 (69.5%)	459/562 (81.7%)	RR 0.73 (0.53 to 0.99)	221 fewer per 1000 (from 8 fewer to 384 fewer)	⊕OOO VERY LOW	
								76.2%		206 fewer per 1000 (from 8 fewer to 358 fewer)		
Depression symptomatology post-treatment – available case analysis (follow-up 15-20 weeks; assessed with: Beck Depression Inventory-II (BDI-II) ≥14 or Edinburgh postnatal Depression Scale (EPDS) >12)												
3	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	reporting bias[2]	90/265 (34%)	135/238 (56.7%)	RR 0.58 (0.44 to 0.77)	238 fewer per 1000 (from 130 fewer to 318 fewer)	⊕⊕OO LOW	

								58.6%		246 fewer per 1000 (from 135 fewer to 328 fewer)		
Depression mean scores post-treatment – available case analysis (follow-up 15-17 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	218	196	-	SMD 0.56 lower (0.76 to 0.37 lower)	⊕⊕⊕⊕ HIGH	

[1] There was evidence of considerable heterogeneity between effect sizes
[2] Papers omit data
[3] Total number of events is less than 300 (a threshold rule-of-thumb)

1.3.6 Depression: post-miscarriage self-help versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: post-miscarriage self-help versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression symptomatology post-treatment – ITT analysis (follow-up mean 5 weeks; assessed with: Brief Symptom Inventory (BSI): Depression (Treatment non-response: reliable change index))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	22/45 (48.9%)	25/33 (75.8%)	RR 0.65 (0.45 to 0.92)	265 fewer per 1000 (from 61 fewer to 417 fewer)	⊕⊕OO LOW	
								75.8%		265 fewer per 1000 (from 61 fewer to 417 fewer)		
Depression symptomatology post-treatment – available case analysis (follow-up mean 5 weeks; assessed with: Brief Symptom Inventory (BSI): Depression (Treatment non-response: reliable change index))												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	10/33 (30.3%)	18/26 (69.2%)	RR 0.44 (0.25 to 0.78)	388 fewer per 1000 (from 152 fewer to 519 fewer)	⊕⊕OO LOW	
								69.2%		388 fewer per 1000 (from 152 fewer to 519 fewer)		
Depression mean scores post-treatment – ITT analysis (follow-up 5-12 weeks; measured with: Brief Symptom Inventory (BSI): Depression or Center for Epidemiological Studies Depression Scale (CES-D); better indicated by lower values)												
2	randomised trials	no serious risk of bias	very serious[2]	no serious indirectness	serious[3]	none	131	119	-	SMD 0.3 lower (1.19 lower to 0.6 higher)	⊕OOO VERY LOW	
Depression mean scores long follow-up (25-103 weeks post-intervention) – ITT analysis (follow-up mean 46 weeks; measured with: Center for Epidemiological Studies Depression Scale (CES-D); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	none	86	86	-	SMD 0.15 lower (0.45 lower to 0.15 higher)	⊕⊕OO LOW	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] There was evidence of considerable heterogeneity between effect sizes
[3] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.7 Depression: post-miscarriage facilitated self-help versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: post-miscarriage facilitated self-help versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		

Depression mean scores post-treatment – ITT analysis (follow-up mean 12 weeks; measured with: Center for Epidemiological Studies Depression Scale (CES-D); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	85	86	-	SMD 0.13 higher (0.17 lower to 0.43 higher)	⊕⊕OO LOW	
Depression mean scores long follow-up (25-103 weeks podt-intervention) – ITT analysis (follow-up mean 46 weeks; measured with: Center for Epidemiological Studies Depression Scale (CES-D); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	85	86	-	SMD 0.1 lower (0.4 lower to 0.2 higher)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.8 Depression: listening visits versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: listening visits versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression diagnosis post-treatment – ITT analysis (follow-up mean 20 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	22/48 (45.8%)	32/52 (61.5%)	RR 0.74 (0.51 to 1.08)	160 fewer per 1000 (from 302 fewer to 49 more)	⊕⊕OO LOW	
								61.5%		160 fewer per 1000 (from 301 fewer to 49 more)		
Depression diagnosis post-treatment – available case analysis (follow-up 7-20 weeks; assessed with: structured Clinical Interview (SCID) or Goldberg's standardised psychiatric interview: Research diagnostic criteria or psychiatric interview using Montgomery–Åsberg Depression Rating Scale (MADRS))												

3	randomised trials	no serious risk of bias	serious[3]	no serious indirectness	very serious[1]	reporting bias[4]	33/89 (37.1%)	57/90 (63.3%)	See comment	317 fewer per 1000 (from 82 fewer to 551 fewer)	⊕OOO VERY LOW	
								62.5%		312 fewer per 1000 (from 81 fewer to 544 fewer)		
Depression symptomatology post-treatment – ITT analysis (follow-up 26-52 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥12)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[4]	176/372 (47.3%)	334/739 (45.2%)	RR 0.96 (0.84 to 1.09)	18 fewer per 1000 (from 72 fewer to 41 more)	⊕⊕⊕O MODERATE	
								49.4%		20 fewer per 1000 (from 79 fewer to 44 more)		
Depression symptomatology post-treatment – available case analysis (follow-up 26-52 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥12)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1,2]	reporting bias[4]	84/280 (30%)	200/605 (33.1%)	RR 0.82 (0.66 to 1.01)	60 fewer per 1000 (from 112 fewer to 3 more)	⊕⊕OO LOW	
								37.3%		67 fewer per 1000 (from 127 fewer to 4 more)		
Depression mean scores post-treatment – available case analysis (follow-up 20-26 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[4]	178	197	-	SMD 0.34 lower (0.55 to 0.14 lower)	⊕⊕⊕O MODERATE	
Depression diagnosis intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (follow-up mean 20 weeks; assessed with: structured Clinical Interview (SCID))												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	17/48 (35.4%)	19/52 (36.5%)	RR 0.97 (0.57 to 1.64)	11 fewer per 1000 (from 157 fewer to 234 more)	⊕⊕OO LOW	
								36.5%		11 fewer per 1000 (from 157 fewer to 234 more)		
Depression diagnosis intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 20 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	16/47 (34%)	15/48 (31.3%)	RR 1.09 (0.61 to 1.94)	28 more per 1000 (from 122 fewer to 294 more)	⊕⊕OO LOW	
								31.3%		28 more per 1000 (from 122 fewer to 294 more)		
Depression mean scores intermediate follow-up (17-24 weeks post-intervention) – by intervention (follow-up 4-12 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS) or Center for Epidemiological Studies Depression Scale (CES-D); better indicated by lower values)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[5]	none	99	98	-	SMD 0.07 lower (0.35 lower to 0.21 higher)	⊕⊕⊕O MODERATE	
Depression mean scores intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 20 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	none	46	48	-	SMD 0.07 higher (0.33 lower to 0.48 higher)	⊕⊕OO LOW	
Depression diagnosis long follow-up (25-103 weeks post-intervention) – ITT analysis (follow-up mean 20 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	17/48 (35.4%)	13/52 (25%)	RR 1.42 (0.77 to 2.6)	105 more per 1000 (from 58 fewer to 400 more)	⊕⊕OO LOW	

								25%		105 more per 1000 (from 58 fewer to 400 more)		
Depression diagnosis long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 20 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	14/45 (31.1%)	9/48 (18.8%)	RR 1.66 (0.8 to 3.45)	124 more per 1000 (from 37 fewer to 459 more)	⊕⊕OO LOW	
								18.8%		124 more per 1000 (from 38 fewer to 461 more)		
Depression symptomatology long follow-up (25-103 weeks post-intervention) – ITT analysis (follow-up mean 78 weeks; assessed with: General Health Questionnaire (GHQ) ≥12)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[4]	117/183 (63.9%)	357/548 (65.1%)	RR 0.98 (0.87 to 1.11)	13 fewer per 1000 (from 85 fewer to 72 more)	⊕⊕⊕O MODERATE	
								65.2%		13 fewer per 1000 (from 85 fewer to 72 more)		
Depression symptomatology long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 78 weeks; assessed with: General Health Questionnaire (GHQ) ≥12)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias[4]	70/136 (51.5%)	222/413 (53.8%)	RR 0.96 (0.79 to 1.15)	22 fewer per 1000 (from 113 fewer to 81 more)	⊕⊕OO LOW	
								53.8%		22 fewer per 1000 (from 113 fewer to 81 more)		
Depression mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 78 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,5]	none	44	48	-	SMD 0.14 higher (0.26 lower to 0.55 higher)	⊕⊕OO LOW	
Depression diagnosis Very long follow-up (>104 weeks post-intervention) – ITT analysis (follow-up mean 260 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	22/48 (45.8%)	13/52 (25%)	RR 1.83 (1.04 to 3.22)	208 more per 1000 (from 10 more to 555 more)	⊕⊕OO LOW	
								25%		208 more per 1000 (from 10 more to 555 more)		
Depression diagnosis Very long follow-up (>104 weeks post-intervention) – available case analysis (follow-up mean 260 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	7/33 (21.2%)	9/37 (24.3%)	RR 0.87 (0.37 to 2.08)	32 fewer per 1000 (from 153 fewer to 263 more)	⊕⊕OO LOW	
								24.3%		32 fewer per 1000 (from 153 fewer to 262 more)		
Depression mean scores Very long follow-up (>104 weeks post-intervention) – available case analysis (follow-up mean 260 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,5]	none	33	34	-	SMD 0.19 lower (0.67 lower to 0.29 higher)	⊕⊕OO LOW	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] There was evidence of moderate to substantial heterogeneity between effect sizes
[4] Papers omit data
[5] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.9 Depression: directive counselling versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Depression: directive counselling versus TAU	Control	Relative (95% CI)	Absolute		
Depression symptomatology post-treatment – ITT analysis (follow-up mean 12 weeks; assessed with: Beck Depression Inventory (BDI) ≥16)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	69/113 (61.1%)	28/33 (84.8%)	RR 0.72 (0.59 to 0.88)	238 fewer per 1000 (from 102 fewer to 348 fewer)	⊕⊕OO LOW	
								84.9%		238 fewer per 1000 (from 102 fewer to 348 fewer)		
Depression symptomatology post-treatment – available case analysis (follow-up mean 12 weeks; assessed with: Beck Depression Inventory (BDI) ≥16)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	28/72 (38.9%)	13/18 (72.2%)	RR 0.54 (0.36 to 0.81)	332 fewer per 1000 (from 137 fewer to 462 fewer)	⊕⊕OO LOW	
								72.2%		332 fewer per 1000 (from 137 fewer to 462 fewer)		
Depression mean scores post-treatment – available case analysis (follow-up mean 12 weeks; measured with: Beck Depression Inventory (BDI); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	none	72	18	-	SMD 0.42 lower (0.95 lower to 0.1 higher)	⊕⊕OO LOW	
Depression mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 52 weeks; measured with: Beck Depression Inventory (BDI); better indicated by lower values)												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	none	37	8	-	SMD 1.46 lower (2.29 to 0.63 lower)	⊕⊕OO LOW	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] Total population size is less than 400 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.10 Depression: post-miscarriage counselling versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: post-miscarriage counselling versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression mean scores post-treatment – ITT analysis (follow-up 7-12 weeks; measured with: Center for Epidemiological Studies Depression Scale (CES-D) or Hamilton Rating Scale for Depression (HRSD); better indicated by lower values)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	94	95	-	SMD 0.17 higher (0.12 lower to 0.46 higher)	⊕⊕OO LOW	
Depression mean scores post-treatment – available case analysis (follow-up 2-7 weeks; measured with: hamilton Rating Scale for Depression (HRSD) or Hospital Anxiety and Depression Scale – Depression; better indicated by lower values)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	41	40	-	SMD 0.14 higher (0.29 lower to 0.58 higher)	⊕⊕OO LOW	
Depression mean scores Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 17 weeks; measured with: Hospital Anxiety and Depression Scale – Depression; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	33	33	-	SMD 0.23 lower (0.71 lower to 0.26 higher)	⊕⊕OO LOW	

Depression mean scores long follow-up (25-103 weeks post-intervention) – ITT analysis (follow-up mean 46 weeks; measured with: Center for Epidemiological Studies Depression Scale (CES-D); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	84	86	-	SMD 0.08 lower (0.38 lower to 0.22 higher)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.11 Depression: post-traumatic birth counselling versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Depression: post-traumatic birth counselling versus TAU	Control	Relative (95% CI)	Absolute		
Depression symptomatology post-treatment – ITT analysis (follow-up mean 13 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥12)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	4/50 (8%)	17/53 (32.1%)	RR 0.25 (0.09 to 0.69)	241 fewer per 1000 (from 99 fewer to 292 fewer)	⊕⊕OO LOW	
								32.1%		241 fewer per 1000 (from 100 fewer to 292 fewer)		
Depression symptomatology post-treatment – available case analysis (follow-up mean 13 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥12)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	4/50 (8%)	17/53 (32.1%)	RR 0.25 (0.09 to 0.69)	241 fewer per 1000 (from 99 fewer to 292 fewer)	⊕⊕OO LOW	

								32.1%		241 fewer per 1000 (from 100 fewer to 292 fewer)		

[1] Total number of events is less than 300 (a threshold rule-of-thumb)

1.3.12 Depression: social support versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: social support versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression diagnosis post-treatment – ITT analysis (follow-up mean 12 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	66/349 (18.9%)	60/352 (17%)	RR 1.11 (0.81 to 1.52)	19 more per 1000 (from 32 fewer to 89 more)	⊕OOO VERY LOW	
								17.1%		19 more per 1000 (from 32 fewer to 89 more)		
Depression diagnosis post-treatment – available case analysis (follow-up mean 12 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias[3]	14/297 (4.7%)	23/315 (7.3%)	See comment	26 fewer per 1000 (from 60 fewer to 10 more)	⊕OOO VERY LOW	
								7.3%		26 fewer per 1000 (from 60 fewer to 10 more)		

Depression symptomatology post-treatment – ITT analysis (follow-up 8-14 weeks; assessed with: Beck Depression Inventory (BDI) ≥10 or Edinburgh postnatal Depression Scale (EPDS) ≥12)												
3	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	109/403 (27%)	145/404 (35.9%)	RR 0.69 (0.47 to 1.01)	111 fewer per 1000 (from 190 fewer to 4 more)	⊕⊕OO LOW	
								54.6%		169 fewer per 1000 (from 289 fewer to 5 more)		
Depression symptomatology post-treatment – available case analysis (follow-up 8-14 weeks; assessed with: Beck Depression Inventory (BDI) ≥10 or Edinburgh postnatal Depression Scale (EPDS) ≥12)												
3	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	53/347 (15.3%)	107/366 (29.2%)	RR 0.52 (0.39 to 0.7)	140 fewer per 1000 (from 88 fewer to 178 fewer)	⊕⊕⊕O MODERATE	
								52.4%		252 fewer per 1000 (from 157 fewer to 320 fewer)		
Depression mean scores post-treatment – available case analysis (follow-up 12-14 weeks; measured with: Beck Depression Inventory (BDI) or Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
3	randomised trials	no serious risk of bias	very serious[4]	no serious indirectness	serious[2]	none	350	373	-	SMD 0.12 lower (0.68 lower to 0.45 higher)	⊕OOO VERY LOW	
Depression symptomatology short follow-up (9-16 weeks post-intervention) – ITT analysis (follow-up mean 24 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥12)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	93/349 (26.6%)	84/352 (23.9%)	RR 1.12 (0.87 to 1.44)	29 more per 1000 (from 31 fewer to 105 more)	⊕⊕OO LOW	
								23.9%		29 more per 1000 (from 31 fewer to 105 more)		

Depression symptomatology short follow-up (9-16 weeks post-intervention) – available case analysis (follow-up mean 24 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥12)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	33/289 (11.4%)	43/311 (13.8%)	RR 0.83 (0.54 to 1.26)	24 fewer per 1000 (from 64 fewer to 36 more)	⊕⊕OO LOW	
								13.8%		23 fewer per 1000 (from 63 fewer to 36 more)		
Depression mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (follow-up mean 24 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	289	311	-	SMD 0.13 lower (0.29 lower to 0.03 higher)	⊕⊕⊕⊕ HIGH	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Papers omit data
[4] There was evidence of considerable heterogeneity between effect sizes

1.3.13 Depression: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: psychologically (CBT/IPT)-informed psychoeducation versus TAU/enhanced TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression diagnosis post-treatment – ITT analysis (follow-up 4-52 weeks; assessed with: mini International Neuropsychiatric Interview (MINI) or Schedule for Affective Disorders and Schizophrenia (SADS)or Maternal Mood Screener (MMS) or Structured Clinical Interview (SCID) or Longitudinal Interval Follow-up Examination (LIFE))												

8	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	69/556 (12.4%)	70/429 (16.3%)	RR 0.67 (0.41 to 1.08)	54 fewer per 1000 (from 96 fewer to 13 more)	⊕OOO VERY LOW	
								23.9%		79 fewer per 1000 (from 141 fewer to 19 more)		
Depression diagnosis post-treatment – available case analysis (follow-up 4-52 weeks; assessed with: schedule for Affective Disorders and Schizophrenia (SADS) or Maternal Mood Screener (MMS) or Structured Clinical Interview (SCID) or Longitudinal Interval Follow-up Examination (LIFE))												
6	randomised trials	no serious risk of bias	very serious[4]	no serious indirectness	very serious[1,2]	reporting bias[3]	22/240 (9.2%)	38/224 (17%)	See comment	98 fewer per 1000 (from 200 fewer to 10 more)	⊕OOO VERY LOW	
								21.9%		127 fewer per 1000 (from 258 fewer to 13 more)		
Depression symptomatology post-treatment – ITT analysis (follow-up 4-26 weeks; assessed with: hopkins Symptom Checklist: sum/20 >0.75 depression or Edinburgh postnatal Depression Scale (EPDS) ≥13 or Leverton Questionnaire (LQ; Elliott et al., 2000) ≥12 or Quick Inventory of Depressive Symptoms (QIDS) ≥11 or Beck Depression Inventory (BDI): Treatment non-response)												
5	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	181/710 (25.5%)	284/808 (35.1%)	RR 0.74 (0.62 to 0.88)	91 fewer per 1000 (from 42 fewer to 134 fewer)	⊕⊕⊕⊕ HIGH	
								48%		125 fewer per 1000 (from 58 fewer to 182 fewer)		
Depression symptomatology post-treatment – available case analysis (follow-up 4-26 weeks; assessed with: hopkins Symptom Checklist: sum/20 >0.75 depression or Quick Inventory of Depressive Symptoms (QIDS) ≥11 or Beck Depression Inventory (BDI): Treatment non-response)												

3	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	132/494 (26.7%)	161/503 (32%)	RR 0.82 (0.68 to 0.98)	58 fewer per 1000 (from 6 fewer to 102 fewer)	⊕⊕⊕O MODERATE	
								45.8%		82 fewer per 1000 (from 9 fewer to 147 fewer)		
Depression mean scores post-treatment – ITT analysis (follow-up 4-31 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS) or Center for Epidemiological Studies Depression Scale (CES-D); better indicated by lower values)												
4	randomised trials	no serious risk of bias	serious[4]	no serious indirectness	no serious imprecision	none	218	218	-	SMD 0.25 lower (0.58 lower to 0.08 higher)	⊕⊕⊕O MODERATE	
Depression mean scores post-treatment – available case analysis (follow-up 4-31 weeks; measured with: Beck Depression Inventory (BDI-II) or Beck Depression Inventory (BDI) or Edinburgh postnatal Depression Scale (EPDS) or Center for Epidemiological Studies Depression Scale (CES-D); better indicated by lower values)												
7	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[5]	none	185	166	-	SMD 0.26 lower (0.48 to 0.05 lower)	⊕⊕⊕O MODERATE	
Depression mean scores short follow-up (9-16 weeks post-intervention) – ITT analysis (follow-up 13-27 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[5]	none	117	118	-	SMD 0.37 lower (0.63 to 0.11 lower)	⊕⊕⊕O MODERATE	
Depression mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (follow-up 19-27 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS) or Beck Depression Inventory (BDI-II); better indicated by lower values)												

2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	53	47	-	SMD 0.42 lower (0.82 to 0.02 lower)	⊕OOO VERY LOW	
Depression diagnosis intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (follow-up 6-36 weeks; assessed with: mini International Neuropsychiatric Interview (MINI) or Schedule for Affective Disorders and Schizophrenia (SADS) or Maternal Mood Screener (MMS))												
4	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	62/425 (14.6%)	35/309 (11.3%)	RR 1.1 (0.75 to 1.6)	11 more per 1000 (from 28 fewer to 68 more)	⊕OOO VERY LOW	
								8.6%		9 more per 1000 (from 22 fewer to 52 more)		
Depression diagnosis intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up 26-36 weeks; assessed with: schedule for Affective Disorders and Schizophrenia (SADS) or Maternal Mood Screener (MMS))												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	17/116 (14.7%)	15/117 (12.8%)	RR 1.1 (0.58 to 2.09)	13 more per 1000 (from 54 fewer to 140 more)	⊕OOO VERY LOW	
								7.7%		8 more per 1000 (from 32 fewer to 84 more)		
Depression mean scores intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (follow-up 26-36 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	none	99	98	-	SMD 0.07 lower (0.35 lower to 0.21 higher)	⊕⊕OO LOW	

Depression mean scores intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 36 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[2,5]	reporting bias[3]	21	20	-	SMD 0.28 lower (0.89 lower to 0.34 higher)	⊕OOO VERY LOW	
Depression diagnosis long follow-up (25-103 weeks post-intervention) – ITT analysis (follow-up 32-75 weeks; assessed with: mini International Neuropsychiatric Interview (MINI) or Schedule for Affective Disorders and Schizophrenia (SADS) or Maternal Mood Screener (MMS) or Structured Clinical Interview (SCID))												
5	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	83/466 (17.8%)	75/346 (21.7%)	RR 0.8 (0.56 to 1.13)	43 fewer per 1000 (from 95 fewer to 28 more)	⊕OOO VERY LOW	
								25%		50 fewer per 1000 (from 110 fewer to 32 more)		
Depression diagnosis long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up 32-75 weeks; assessed with: schedule for Affective Disorders and Schizophrenia (SADS) or Maternal Mood Screener (MMS) or Structured Clinical Interview (SCID))												
3	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	19/138 (13.8%)	29/128 (22.7%)	RR 0.6 (0.36 to 1.03)	91 fewer per 1000 (from 145 fewer to 7 more)	⊕OOO VERY LOW	
								25%		100 fewer per 1000 (from 160 fewer to 7 more)		
Depression mean scores long follow-up (25-103 weeks post-intervention) – ITT analysis (follow-up 57-75 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												

2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	none	44	42	-	SMD 0.43 lower (0.86 lower to 0 higher)	⊕⊕OO LOW	
Depression mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up 32-75 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS) or Beck Depression Inventory (BDI-II); better indicated by lower values)												
3	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	85	76	-	SMD 0.44 lower (0.75 to 0.12 lower)	⊕OOO VERY LOW	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Papers omit data
[4] There was evidence of substantial heterogeneity between effect sizes
[5] Total population size is less than 400 (a threshold rule-of-thumb)
[6] Risk of bias due to statistically significant group differences at baseline

1.3.14 Depression: IPT-informed psychoeducation versus non-mental health-focused education and support

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: IPT-informed psychoeducation versus non-mental health-focused education and support**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression symptomatology post-treatment – ITT Analysis (follow-up mean 16 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS))												
1	randomised trials	no serious	no serious inconsistency	no serious indirectness	very serious[1,2]	none	14/21 (66.7%)	15/17 (88.2%)		212 fewer per 1000 (from 415	⊕⊕OO LOW	

		risk of bias							RR 0.76 (0.53 to 1.07)	fewer to 62 more)			
								88.2%		212 fewer per 1000 (from 415 fewer to 62 more)			

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.15 Depression: non-mental health-focused education and support versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: non-mental health-focused education and support versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression symptomatology post-treatment – ITT analysis (follow-up mean 12 weeks; assessed with: hopkins Symptom Checklist-25 (HSCL-25): >1.06)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	129/168 (76.8%)	138/163 (84.7%)	RR 0.91 (0.82 to 1.01)	76 fewer per 1000 (from 152 fewer to 8 more)	⊕⊕⊕O MODERATE	
								84.7%		76 fewer per 1000 (from 152 fewer to 8 more)		
Depression symptomatology post-treatment – available case analysis (follow-up mean 12 weeks; assessed with: hopkins Symptom Checklist-25 (HSCL-25): >1.06)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	58/97 (59.8%)	66/91 (72.5%)	RR 0.82 (0.67 to 1.01)	131 fewer per 1000 (from 239 fewer to 7 more)	⊕⊕OO LOW	

								72.5%		131 fewer per 1000 (from 239 fewer to 7 more)		

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.16 Depression: home visits versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: home visits versus TAU/enhanced TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression diagnosis post-treatment – ITT analysis (follow-up mean 6 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	4/9 (44.4%)	6/9 (66.7%)	RR 0.67 (0.28 to 1.58)	220 fewer per 1000 (from 480 fewer to 387 more)	⊕OOO VERY LOW	
								66.7%		220 fewer per 1000 (from 480 fewer to 387 more)		
Depression diagnosis post-treatment – available case analysis (follow-up mean 6 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	2/7 (28.6%)	6/9 (66.7%)	See comment	380 fewer per 1000 (from 840 fewer to 73 more)	⊕OOO VERY LOW	

								66.7%		380 fewer per 1000 (from 840 fewer to 73 more)		
Depression symptomatology post-treatment – ITT analysis (follow-up 22-104 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10/12 or Center for Epidemiological Studies Depression Scale (CES-D) ≥24)												
3	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[4]	203/491 (41.3%)	223/494 (45.1%)	RR 0.92 (0.8 to 1.06)	36 fewer per 1000 (from 90 fewer to 27 more)	⊕⊕⊕O MODERATE	
								47.7%		38 fewer per 1000 (from 95 fewer to 29 more)		
Depression symptomatology post-treatment – available case analysis (follow-up 22-104 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥10/12 or Center for Epidemiological Studies Depression Scale (CES-D) ≥24)												
3	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias[4]	90/378 (23.8%)	105/376 (27.9%)	RR 0.87 (0.69 to 1.1)	36 fewer per 1000 (from 87 fewer to 28 more)	⊕OOO VERY LOW	
								22%		29 fewer per 1000 (from 68 fewer to 22 more)		
Depression mean scores post-treatment – available case analysis (follow-up 22-52 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS) or Center for Epidemiological Studies Depression (CES-D); better indicated by lower values)												
3	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	479	481	-	SMD 0.17 lower (0.3 to 0.05 lower)	⊕⊕⊕⊕ HIGH	

[1] Risk of bias due to unclear blinding of outcome assessment
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] Papers omit data

1.3.17 Depression: mother–infant relationship interventions versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Depression: mother–infant relationship interventions versus TAU/enhanced TAU	Control	Relative (95% CI)	Absolute		
Depression diagnosis post-treatment – ITT analysis (follow-up mean 20 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	19/43 (44.2%)	32/52 (61.5%)	RR 0.72 (0.48 to 1.07)	172 fewer per 1000 (from 320 fewer to 43 more)	⊕⊕OO LOW	
								61.5%		172 fewer per 1000 (from 320 fewer to 43 more)		
Depression diagnosis post-treatment – available case analysis (follow-up mean 20 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	18/42 (42.9%)	30/50 (60%)	See comment	174 fewer per 1000 (from 372 fewer to 30 more)	⊕⊕OO LOW	
								60%		174 fewer per 1000 (from 372 fewer to 30 more)		
Depression symptomatology post-treatment – ITT analysis (follow-up 5-26 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS): Treatment non-response (reliable change index-no improvement)/EPDS ≥12 or Center for Epidemiologic Studies Depression Scale (CES-D) ≥16)												

3	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	98/196 (50%)	113/200 (56.5%)	RR 0.87 (0.69 to 1.1)	73 fewer per 1000 (from 175 fewer to 57 more)	⊕⊕OO LOW	
								71.7%		93 fewer per 1000 (from 222 fewer to 72 more)		
Depression symptomatology post-treatment – available case analysis (follow-up 5-26 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS): Treatment non-response (reliable change index-no improvement)/EPDS ≥12 or Center for Epidemiologic Studies Depression Scale (CES-D) ≥16)													
3	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	50/148 (33.8%)	53/140 (37.9%)	RR 0.85 (0.58 to 1.25)	57 fewer per 1000 (from 159 fewer to 95 more)	⊕⊕OO LOW	
								47.2%		71 fewer per 1000 (from 198 fewer to 118 more)		
Depression mean scores post-treatment – available case (follow-up 5-28 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS) or Beck Depression Inventory (BDI) or Beck Depression Inventory (BDI-II) or Center for Epidemiologic Studies Depression Scale (CES-D); better indicated by lower values)													
6	randomised trials	no serious risk of bias	very serious[3]	no serious indirectness	no serious imprecision	none	283	283	-	SMD 0.02 higher (0.38 lower to 0.41 higher)	⊕⊕OO LOW	
Depression diagnosis intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (follow-up mean 39 weeks; assessed with: structured Clinical Interview (SCID))													
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	13/43 (30.2%)	19/52 (36.5%)	RR 0.83 (0.46 to 1.48)	62 fewer per 1000 (from 197 fewer to 175 more)	⊕⊕OO LOW	
								36.5%		62 fewer per 1000 (from 197 fewer to 175 more)		
Depression diagnosis intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 39 weeks; assessed with: structured Clinical Interview (SCID))													

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	10/40 (25%)	15/48 (31.3%)	RR 0.8 (0.4 to 1.58)	62 fewer per 1000 (from 188 fewer to 181 more)	⊕⊕OO LOW	
								31.3%		63 fewer per 1000 (from 188 fewer to 182 more)		
Depression symptomatology intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (follow-up mean 25 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥12)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	20/60 (33.3%)	16/61 (26.2%)	RR 1.27 (0.73 to 2.21)	71 more per 1000 (from 71 fewer to 317 more)	⊕⊕OO LOW	
								26.2%		71 more per 1000 (from 71 fewer to 317 more)		
Depression symptomatology intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 25 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥12)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	6/46 (13%)	4/50 (8%)	RR 1.63 (0.49 to 5.41)	50 more per 1000 (from 41 fewer to 353 more)	⊕⊕OO LOW	
								8%		50 more per 1000 (from 41 fewer to 353 more)		
Depression mean scores intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 39 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,4]	none	40	48	-	SMD 0.11 lower (0.53 lower to 0.31 higher)	⊕⊕OO LOW	
Depression diagnosis long follow-up (25-103 weeks post-intervention) – ITT analysis (follow-up mean 78 weeks; assessed with: structured Clinical Interview (SCID))												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	13/43 (30.2%)	13/52 (25%)	RR 1.21 (0.63 to 2.33)	53 more per 1000 (from 93 fewer to 332 more)	⊕⊕OO LOW	
								25%		53 more per 1000 (from 93 fewer to 332 more)		
Depression diagnosis long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 78 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	12/42 (28.6%)	9/48 (18.8%)	RR 1.52 (0.71 to 3.25)	97 more per 1000 (from 54 fewer to 422 more)	⊕⊕OO LOW	
								18.8%		98 more per 1000 (from 55 fewer to 423 more)		
Depression mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up 57-78 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS) or Beck Depression Inventory (BDI); better indicated by lower values)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	none	77	84	-	SMD 0.08 higher (0.23 lower to 0.39 higher)	⊕⊕OO LOW	
Depression diagnosis Very long follow-up (≥104 weeks post-intervention) – ITT analysis (follow-up mean 260 weeks; assessed with: structured Clinical Interview (SCID))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	13/43 (30.2%)	13/52 (25%)	RR 1.21 (0.63 to 2.33)	53 more per 1000 (from 93 fewer to 332 more)	⊕⊕OO LOW	
								25%		53 more per 1000 (from 93 fewer to 332 more)		
Depression diagnosis Very long follow-up (≥104 weeks post-intervention) – available case analysis (follow-up mean 260 weeks; assessed with: structured Clinical Interview (SCID))												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	6/36 (16.7%)	9/37 (24.3%)	RR 0.69 (0.27 to 1.73)	75 fewer per 1000 (from 178 fewer to 178 more)	⊕⊕OO LOW	
								24.3%		75 fewer per 1000 (from 177 fewer to 177 more)		
Depression mean scores Very long follow-up (≥104 weeks post-intervention) – available case analysis (follow-up mean 260 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,4]	none	31	34	-	SMD 0.17 lower (0.66 lower to 0.32 higher)	⊕⊕OO LOW	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] There was evidence of considerable heterogeneity between effect sizes
[4] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.18 Depression: mother–infant relationship intervention with video feedback versus mother–infant relationship intervention with verbal feedback

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: mother–infant relationship intervention with video feedback versus mother–infant relationship intervention with verbal feedback**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression mean scores post-treatment – available case analysis (follow-up mean 3 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
1	randomised trials	no serious	no serious inconsistency	no serious indirectness	very serious[1,2]	none	17	20	-	SMD 0.29 higher (0.36	⊕⊕OO LOW	

		risk of bias								lower to 0.94 higher)		

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.19 Depression: co-parenting intervention versus enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: co-parenting intervention versus enhanced TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression diagnosis post-treatment – ITT analysis (follow-up mean 6 weeks; assessed with: mini International Neuropsychiatric Interview (MINI))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	5/16 (31.3%)	8/13 (61.5%)	RR 0.51 (0.22 to 1.18)	302 fewer per 1000 (from 480 fewer to 111 more)	⊕OOO VERY LOW	
								61.5%		301 fewer per 1000 (from 480 fewer to 111 more)		
Depression diagnosis post-treatment – available case analysis (follow-up mean 6 weeks; assessed with: mini International Neuropsychiatric Interview (MINI))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	5/16 (31.3%)	8/13 (61.5%)	See comment	302 fewer per 1000 (from 652 fewer to 49 more)	⊕OOO VERY LOW	
								61.5%		301 fewer per 1000 (from 652 fewer to 49 more)		

Depression mean scores post-treatment – available case analysis (follow-up mean 6 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3,4]	none	15	13	-	SMD 0.47 lower (1.22 lower to 0.29 higher)	⊕OOO VERY LOW	

[1] Risk of bias as blinding of outcome assessment was unclear
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.20 Depression: infant sleep training (controlled crying) versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: infant sleep training (controlled crying) versus TAU/enhanced TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression symptomatology post-treatment – available case analysis (follow-up mean 74 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) >9)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	22/143 (15.4%)	34/129 (26.4%)	RR 0.58 (0.36 to 0.94)	111 fewer per 1000 (from 16 fewer to 169 fewer)	⊕⊕OO LOW	
								26.4%		111 fewer per 1000 (from 16 fewer to 169 fewer)		
Depression mean scores post-treatment – available case analysis (follow-up 9-13 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS) change score or score at endpoint; better indicated by lower values)												

2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	none	96	93	-	SMD 0.47 lower (0.76 to 0.18 lower)	⊕⊕OO LOW	
Depression mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (follow-up 17-22 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS) change score or score at endpoint; better indicated by lower values)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	none	96	88	-	SMD 0.4 lower (0.7 to 0.11 lower)	⊕⊕OO LOW	
Depression mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 74 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	143	129	-	SMD 0.26 lower (0.5 to 0.02 lower)	⊕⊕⊕O MODERATE	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.21 Depression: music therapy during birth versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: music therapy during birth versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Depression symptomatology post-treatment – ITT analysis (follow-up mean 3 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥13)												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	13/80 (16.3%)	23/81 (28.4%)	RR 0.57 (0.31 to 1.05)	122 fewer per 1000 (from 196 fewer to 14 more)	⊕⊕OO LOW	
								28.4%		122 fewer per 1000 (from 196 fewer to 14 more)		
Depression symptomatology post-treatment – available case analysis (follow-up mean 3 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥13)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	4/71 (5.6%)	12/70 (17.1%)	RR 0.33 (0.11 to 0.97)	115 fewer per 1000 (from 5 fewer to 153 fewer)	⊕⊕OO LOW	
								17.1%		115 fewer per 1000 (from 5 fewer to 152 fewer)		
Depression mean scores post-treatment – available case analysis (follow-up mean 3 weeks; measured with: Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	none	71	70	-	SMD 0.37 lower (0.71 to 0.04 lower)	⊕⊕OO LOW	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.22 Depression: psychosomatic intervention versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Depression: psychosomatic intervention versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		

Depression symptomatology post-treatment – ITT analysis (follow-up mean 34 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥12)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	47/92 (51.1%)	61/92 (66.3%)	RR 0.77 (0.6 to 0.99)	152 fewer per 1000 (from 7 fewer to 265 fewer)	⊕OOO VERY LOW	
								66.3%		152 fewer per 1000 (from 7 fewer to 265 fewer)		
Depression symptomatology post-treatment – available case analysis (follow-up mean 34 weeks; assessed with: Edinburgh postnatal Depression Scale (EPDS) ≥12)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	24/69 (34.8%)	27/58 (46.6%)	RR 0.75 (0.49 to 1.14)	116 fewer per 1000 (from 237 fewer to 65 more)	⊕OOO VERY LOW	
								46.6%		116 fewer per 1000 (from 238 fewer to 65 more)		
Depression mean scores post-treatment – available case analysis (follow-up 34-52 weeks; measured with: Hospital Anxiety and Depression Scale – Depression or Edinburgh postnatal Depression Scale (EPDS); better indicated by lower values)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3,4]	none	90	81	-	SMD 0.21 lower (0.54 lower to 0.13 higher)	⊕OOO VERY LOW	

[1] Risk of attrition bias due to statistically significant higher drop-out in the control group
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.23 Depression: mindfulness training versus enhanced TAU

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Depression: mindfulness training versus enhanced TAU	Control	Relative (95% CI)	Absolute		
Depression mean scores post-treatment – available case analysis (follow-up mean 10 weeks; measured with: Center for Epidemiological Studies Depression Scale (CES-D); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	13	18	-	SMD 0.13 lower (0.85 lower to 0.58 higher)	⊕OOO VERY LOW	
Negative affect mean scores post-treatment – available case analysis (follow-up mean 10 weeks; measured with: positive and Negative Affect Schedule-Extended (PANAS-X): negative affect; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	13	18	-	SMD 0.32 lower (1.04 lower to 0.4 higher)	⊕OOO VERY LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data

1.3.24 Anxiety: structured psychological interventions versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anxiety: structured psychological interventions versus TAU/enhanced TAU	Control	Relative (95% CI)	Absolute		
Anxiety mean scores post-treatment – ITT analysis (follow-up mean 44 weeks; measured with: Beck Anxiety Inventory (BAI); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	25	28	-	SMD 1.34 lower (1.94 to 0.74 lower)	⊕⊕OO LOW	

Anxiety mean scores post-treatment – available case analysis (follow-up 12-26 weeks; measured with: Beck Anxiety Inventory (BAI) or State-Trait Anxiety Inventory (STAI)-State; better indicated by lower values)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias[2]	161	154	-	SMD 0.35 lower (0.58 to 0.13 lower)	⊕⊕OO LOW	
Trait anxiety mean scores post-treatment – available case analysis (follow-up mean 26 weeks; measured with: state-Trait Anxiety Inventory (STAI) – Trait; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias[2]	133	130	-	SMD 0.38 lower (0.62 to 0.13 lower)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] Papers omit data

1.3.25 Anxiety: CBT versus relational constructivist therapy

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Anxiety: CBT versus relational constructivist therapy**	**Control**	**Relative (95% CI)**	**Absolute**		
Anxiety mean scores post-treatment – available case analysis (measured with: Beck Anxiety Inventory (BAI); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	32	28	-	SMD 0.26 higher (0.25 lower to 0.77 higher)	⊕OOO VERY LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Papers omit data

1.3.26 Anxiety: IPT versus support group

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anxiety: IPT versus support group	Control	Relative (95% CI)	Absolute		
Anxiety mean scores post-treatment – available case (follow-up mean 12 weeks; measured with: state-Trait Anxiety Inventory (STAI)-State; better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	22	22	-	SMD 0.48 lower (1.09 lower to 0.12 higher)	⊕OOO VERY LOW	

[1] Risk of bias as statistically significant group differences at baseline
[2] Total population size is less than 400 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.27 Anxiety: facilitated self-help versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anxiety: facilitated self-help versus TAU	Control	Relative (95% CI)	Absolute		
Anxiety symptomatology post-treatment – ITT analysis (follow-up mean 20 weeks; assessed with: Depression Anxiety Stress Scale (DASS): Anxiety ≥8)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias[2]	27/71 (38%)	41/72 (56.9%)	RR 0.67 (0.47 to 0.96)	188 fewer per 1000 (from 23 fewer to 302 fewer)	⊕OOO VERY LOW	

								56.9%		188 fewer per 1000 (from 23 fewer to 302 fewer)		
Anxiety symptomatology post-treatment – available case analysis (follow-up mean 20 weeks; assessed with: Depression Anxiety Stress Scale (DASS): Anxiety ≥8)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias[2]	3/47 (6.4%)	11/42 (26.2%)	RR 0.24 (0.07 to 0.81)	199 fewer per 1000 (from 50 fewer to 244 fewer)	⊕OOO VERY LOW	
								26.2%		199 fewer per 1000 (from 50 fewer to 244 fewer)		
Anxiety mean scores post-treatment – available case analysis (follow-up mean 17 weeks; measured with: Generalised Anxiety Disorder Assessment (GAD-7); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3,4]	reporting bias[2]	31	28	-	SMD 0.5 lower (1.02 lower to 0.02 higher)	⊕OOO VERY LOW	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] Paper omits data
[3] Total population size is less than 400 (a threshold rule-of-thumb)
[4] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.28 Anxiety: post-miscarriage self-help versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Anxiety: post-miscarriage self-help versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		

Anxiety symptomatology post-treatment – ITT analysis (follow-up mean 5 weeks; assessed with: Brief Symptom Inventory (BSI): Anxiety (Treatment non-response: reliable change index))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	31/45 (68.9%)	24/33 (72.7%)	RR 0.95 (0.71 to 1.26)	36 fewer per 1000 (from 211 fewer to 189 more)	⊕⊕OO LOW	
								72.7%		36 fewer per 1000 (from 211 fewer to 189 more)		
Anxiety symptomatology post-treatment – available case analysis (follow-up mean 5 weeks; assessed with: Brief Symptom Inventory (BSI): Anxiety (Treatment non-response: reliable change index))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	19/33 (57.6%)	18/26 (69.2%)	RR 0.83 (0.56 to 1.23)	118 fewer per 1000 (from 305 fewer to 159 more)	⊕⊕OO LOW	
								69.2%		118 fewer per 1000 (from 304 fewer to 159 more)		
Anxiety mean scores post-treatment – ITT analysis (follow-up mean 5 weeks; measured with: Brief Symptom Inventory (BSI): Anxiety; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	none	45	33	-	SMD 0.23 lower (0.68 lower to 0.23 higher)	⊕⊕OO LOW	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.29 Anxiety: listening visits versus TAU

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anxiety: listening visits versus TAU	Control	Relative (95% CI)	Absolute		
Anxiety mean scores post-treatment – available case analysis (follow-up mean 26 weeks; measured with: state-Trait Anxiety Inventory (STAI)-State; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias[2]	124	136	-	SMD 0.29 lower (0.53 to 0.04 lower)	⊕⊕OO LOW	
Trait anxiety mean scores post-treatment – available case analysis (follow-up mean 26 weeks; measured with: state-Trait Anxiety Inventory (STAI) – Trait; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias[2]	124	130	-	SMD 0.26 lower (0.51 to 0.02 lower)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] Paper omits data

1.3.30 Anxiety: directive counselling versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anxiety: directive counselling versus TAU	Control	Relative (95% CI)	Absolute		
Anxiety mean scores post-treatment – available case analysis (follow-up mean 12 weeks; measured with: Beck Anxiety Inventory (BAI); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	72	18	-	SMD 0.56 lower (1.09 to 0.04 lower)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.31 Anxiety: post-miscarriage counselling versus enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anxiety: post-miscarriage counselling versus enhanced TAU	Control	Relative (95% CI)	Absolute		
Anxiety mean scores post-treatment – available case analysis (follow-up mean 2 weeks; measured with: Hospital Anxiety and Depression Scale – Anxiety; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	33	33	-	SMD 0.11 higher (0.38 lower to 0.59 higher)	⊕⊕OO LOW	
Anxiety mean scores Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 17 weeks; measured with: Hospital Anxiety and Depression Scale – Anxiety; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	33	33	-	SMD 0.31 lower (0.8 lower to 0.17 higher)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.32 Anxiety: post-traumatic birth counselling versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anxiety: post-traumatic birth counselling versus TAU	Control	Relative (95% CI)	Absolute		

Anxiety symptomatology post-treatment – ITT analysis (follow-up mean 13 weeks; assessed with: Depression Anxiety Stress Scale (DASS): Anxiety >9)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	1/50 (2%)	6/53 (11.3%)	RR 0.18 (0.02 to 1.42)	93 fewer per 1000 (from 111 fewer to 48 more)	⊕⊕OO LOW	
								11.3%		93 fewer per 1000 (from 111 fewer to 47 more)		
Anxiety symptomatology post-treatment – available case analysis (follow-up mean 13 weeks; assessed with: Depression Anxiety Stress Scale (DASS): Anxiety >9)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	1/50 (2%)	6/53 (11.3%)	RR 0.18 (0.02 to 1.42)	93 fewer per 1000 (from 111 fewer to 48 more)	⊕⊕OO LOW	
								11.3%		93 fewer per 1000 (from 111 fewer to 47 more)		

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.33 Anxiety: social support versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Anxiety: social support versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Anxiety symptomatology post-treatment – ITT analysis (follow-up mean 12 weeks; assessed with: state-Trait Anxiety Inventory (STAI)-State >44)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias[2]	113/349 (32.4%)	123/352 (34.9%)	RR 0.93 (0.75 to 1.14)	24 fewer per 1000 (from 87 fewer to 49 more)	⊕⊕OO LOW	

								34.9%		24 fewer per 1000 (from 87 fewer to 49 more)		
Anxiety symptomatology post-treatment – available case analysis (follow-up mean 12 weeks; assessed with: state-Trait Anxiety Inventory (STAI)-State >44)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias[2]	61/297 (20.5%)	86/315 (27.3%)	RR 0.75 (0.56 to 1)	68 fewer per 1000 (from 120 fewer to 0 more)	⊕OOO VERY LOW	
								27.3%		68 fewer per 1000 (from 120 fewer to 0 more)		
Anxiety mean scores post-treatment – available case analysis (follow-up mean 12 weeks; measured with: state-Trait Anxiety Inventory (STAI)-State; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[2]	297	315	-	SMD 0.14 lower (0.3 lower to 0.02 higher)	⊕⊕⊕O MODERATE	
Anxiety mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (follow-up mean 24 weeks; measured with: state-Trait Anxiety Inventory (STAI)-State; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias[2]	289	311	-	SMD 0.07 lower (0.23 lower to 0.09 higher)	⊕⊕⊕O MODERATE	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] Paper omits data

1.3.34 Anxiety: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anxiety: psychologically (CBT/IPT)-informed psychoeducation versus TAU/enhanced TAU	Control	Relative (95% CI)	Absolute		
Anxiety diagnosis post-treatment – ITT analysis (follow-up 9-52 weeks; assessed with: mini International Neuropsychiatric Interview (MINI) or Schedule for Affective Disorders and Schizophrenia (SADS))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias[4]	43/292 (14.7%)	25/184 (13.6%)	RR 0.97 (0.61 to 1.54)	4 fewer per 1000 (from 53 fewer to 73 more)	⊕OOO VERY LOW	
								13.8%		4 fewer per 1000 (from 54 fewer to 75 more)		
Anxiety diagnosis post-treatment – available case analysis (follow-up mean 9 weeks; assessed with: schedule for Affective Disorders and Schizophrenia (SADS))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias[4]	8/101 (7.9%)	10/98 (10.2%)	RR 0.78 (0.32 to 1.88)	22 fewer per 1000 (from 69 fewer to 90 more)	⊕OOO VERY LOW	
								10.2%		22 fewer per 1000 (from 69 fewer to 90 more)		
Anxiety diagnosis long follow-up (25-103 weeks post-intervention) – ITT analysis (assessed with: mini International Neuropsychiatric Interview (MINI))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias[4]	31/191 (16.2%)	14/86 (16.3%)	RR 1 (0.56 to 1.78)	0 fewer per 1000 (from 72 fewer to 127 more)	⊕OOO VERY LOW	
								16.3%		0 fewer per 1000 (from 72 fewer to 127 more)		

[1] Risk of bias as statistically significant group differences at baseline
[2] Total number of events is less than 300 (a threshold rule-of-thumb)

[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] Papers omit data

1.3.35 Anxiety: mother–infant relationship interventions versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anxiety: mother–infant relationship interventions versus TAU/enhanced TAU	Control	Relative (95% CI)	Absolute		
Anxiety symptomatology post-treatment – ITT analysis (follow-up mean 7 weeks; assessed with: state-Trait Anxiety Inventory (STAI)-State >40)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	12/60 (20%)	13/61 (21.3%)	RR 0.94 (0.47 to 1.89)	13 fewer per 1000 (from 113 fewer to 190 more)	⊕⊕OO LOW	
								21.3%		13 fewer per 1000 (from 113 fewer to 190 more)		
Anxiety symptomatology post-treatment – available case analysis (follow-up mean 7 weeks; assessed with: state-Trait Anxiety Inventory (STAI)-State >40)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	0/48 (0%)	2/50 (4%)	RR 0.21 (0.01 to 4.23)	32 fewer per 1000 (from 40 fewer to 129 more)	⊕⊕OO LOW	
								4%		32 fewer per 1000 (from 40 fewer to 129 more)		
Anxiety mean scores post-treatment – available case analysis (follow-up mean 7 weeks; measured with: state-Trait Anxiety Inventory (STAI)-State; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	none	48	50	-	SMD 0.16 lower (0.55 lower to 0.24 higher)	⊕⊕OO LOW	

Anxiety mean scores Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 25 weeks; measured with: state-Trait Anxiety Inventory (STAI)-State; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	none	46	50	-	SMD 0.3 lower (0.7 lower to 0.11 higher)	⊕⊕OO LOW	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.36 Anxiety: music therapy during birth versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anxiety: music therapy during birth versus TAU	Control	Relative (95% CI)	Absolute		
Anxiety mean scores post-treatment – available case analysis (follow-up mean 3 weeks; measured with: Visual Analogue Scale (VAS) Anxiety; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	71	70	-	SMD 2.16 lower (2.58 to 1.74 lower)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.37 Anxiety: psychosomatic intervention versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anxiety: psychosomatic intervention versus TAU	Control	Relative (95% CI)	Absolute		

Anxiety mean scores post-treatment – available case analysis (follow-up mean 52 weeks; measured with: Hospital Anxiety and Depression Scale – Anxiety; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	21	23	-	SMD 0.17 lower (0.76 lower to 0.42 higher)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.38 Anxiety: mindfulness training versus enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anxiety: mindfulness training versus enhanced TAU	Control	Relative (95% CI)	Absolute		
Anxiety mean scores post-treatment – ITT analysis (follow-up mean 6 weeks; measured with: state-Trait Anxiety Inventory (STAI)-State; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	24	23	-	SMD 0.23 higher (0.35 lower to 0.8 higher)	⊕⊕OO LOW	
Anxiety mean scores post-treatment – available case analysis (follow-up mean 10 weeks; measured with: state-Trait Anxiety Inventory (STAI)-State; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	13	18	-	SMD 0.02 lower (0.74 lower to 0.69 higher)	⊕OOO VERY LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data

1.3.39 Adjustment disorder: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Adjustment disorder: psychologically (CBT/IPT)-informed psychoeducation versus TAU/enhanced TAU	Control	Relative (95% CI)	Absolute		
Adjustment disorders diagnosis post-treatment – ITT analysis (follow-up mean 52 weeks; assessed with: schedule for Affective Disorders and Schizophrenia (SADS))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	13/101 (12.9%)	14/98 (14.3%)	RR 0.9 (0.45 to 1.82)	14 fewer per 1000 (from 79 fewer to 117 more)	⊕⊕OO LOW	
								14.3%		14 fewer per 1000 (from 79 fewer to 117 more)		
Adjustment disorders diagnosis post-treatment – available case analysis (follow-up mean 52 weeks; assessed with: schedule for Affective Disorders and Schizophrenia (SADS))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	13/101 (12.9%)	14/98 (14.3%)	RR 0.9 (0.45 to 1.82)	14 fewer per 1000 (from 79 fewer to 117 more)	⊕⊕OO LOW	
								14.3%		14 fewer per 1000 (from 79 fewer to 117 more)		

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.40 PTSD: post-miscarriage self-help versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PTSD: post-miscarriage self-help versus TAU	Control	Relative (95% CI)	Absolute		
PTSD symptomatology post-treatment – ITT analysis (follow-up mean 5 weeks; assessed with: Impact of Events Scale (IES): Treatment non-response (reliable change index))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	17/45 (37.8%)	21/33 (63.6%)	RR 0.59 (0.38 to 0.94)	261 fewer per 1000 (from 38 fewer to 395 fewer)	⊕⊕OO LOW	
								63.6%		261 fewer per 1000 (from 38 fewer to 394 fewer)		
PTSD symptomatology post-treatment – available case analysis (follow-up mean 5 weeks; assessed with: Impact of Events Scale (IES): Treatment non-response (reliable change index))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	6/33 (18.2%)	15/26 (57.7%)	RR 0.32 (0.14 to 0.7)	392 fewer per 1000 (from 173 fewer to 496 fewer)	⊕⊕OO LOW	
								57.7%		392 fewer per 1000 (from 173 fewer to 496 fewer)		
PTSD mean scores post-treatment – ITT analysis (follow-up mean 5 weeks; measured with: Impact of Events Scale (IES): Traumatic stress; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	none	45	33	-	SMD 0.84 lower (1.31 to 0.37 lower)	⊕⊕OO LOW	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.41 PTSD: post-traumatic birth counselling versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PTSD: post-traumatic birth counselling versus TAU	Control	Relative (95% CI)	Absolute		
PTSD diagnosis post-treatment – ITT analysis (follow-up mean 13 weeks; assessed with: mini-PTSD Diagnosis Interview)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	3/50 (6%)	9/53 (17%)	RR 0.35 (0.1 to 1.23)	110 fewer per 1000 (from 153 fewer to 39 more)	⊕⊕OO LOW	
								17%		111 fewer per 1000 (from 153 fewer to 39 more)		
PTSD diagnosis post-treatment – available case analysis (follow-up mean 13 weeks; assessed with: mini-PTSD Diagnosis Interview)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	3/50 (6%)	9/53 (17%)	RR 0.35 (0.1 to 1.23)	110 fewer per 1000 (from 153 fewer to 39 more)	⊕⊕OO LOW	
								17%		111 fewer per 1000 (from 153 fewer to 39 more)		
PTSD mean scores post-treatment – ITT analysis (follow-up mean 13 weeks; measured with: mini-PTSD Diagnosis Interview: 'Trauma symptoms', rating scale unclear ; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	none	50	53	-	SMD 0.41 lower (0.81 to 0.02 lower)	⊕⊕OO LOW	
PTSD mean scores post-treatment – available case analysis (follow-up mean 13 weeks; measured with: mini-PTSD Diagnosis Interview: 'Trauma symptoms', rating scale unclear ; better indicated by lower values)												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	none	50	53	-	SMD 0.41 lower (0.81 to 0.02 lower)	⊕⊕OO LOW	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.42 PTSD: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PTSD: psychologically (CBT/IPT)-informed psychoeducation versus TAU/enhanced TAU	Control	Relative (95% CI)	Absolute		
PTSD diagnosis post-treatment – ITT analysis (follow-up mean 13 weeks; assessed with: Longitudinal Interval Follow-up Examination (LIFE))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias[4]	4/28 (14.3%)	5/26 (19.2%)	RR 0.74 (0.22 to 2.47)	50 fewer per 1000 (from 150 fewer to 283 more)	⊕OOO VERY LOW	
								19.2%		50 fewer per 1000 (from 150 fewer to 282 more)		
PTSD diagnosis post-treatment – available case analysis (follow-up mean 13 weeks; assessed with: Longitudinal Interval Follow-up Examination (LIFE))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias[4]	1/25 (4%)	0/21 (0%)	RR 2.54 (0.11 to 59.23)	-	⊕OOO VERY LOW	
								0%		-		

PTSD mean scores post-treatment – available case analysis (follow-up 6-13 weeks; measured with: Davidson Trauma Scale or Longitudinal Interval Follow-up Examination (LIFE): psychiatric Status Ratings (PSRs) mean PTSD score; better indicated by lower values)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[4]	50	46	-	SMD 0.4 lower (0.81 lower to 0 higher)	⊕OOO VERY LOW	

[1] Risk of bias due to unclear blinding of outcome assessment
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] Papers omit data
[5] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.43 PTSD: mother–infant relationship interventions versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PTSD: mother–infant relationship interventions versus TAU/enhanced TAU	Control	Relative (95% CI)	Absolute		
PTSD symptomatology post-treatment – ITT analysis (follow-up mean 7 weeks; assessed with: perinatal PTSD Questionnaire (PPQ): scores in clinical range (no further detail))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	22/60 (36.7%)	19/61 (31.1%)	RR 1.18 (0.71 to 1.94)	56 more per 1000 (from 90 fewer to 293 more)	⊕⊕OO LOW	
								31.2%		56 more per 1000 (from 90 fewer to 293 more)		
PTSD symptomatology post-treatment – available case analysis (follow-up mean 7 weeks; assessed with: perinatal PTSD Questionnaire (PPQ): scores in clinical range (no further detail))												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	10/48 (20.8%)	8/50 (16%)	RR 1.3 (0.56 to 3.02)	48 more per 1000 (from 70 fewer to 323 more)	⊕⊕OO LOW	
								16%		48 more per 1000 (from 70 fewer to 323 more)		
PTSD mean scores post-treatment – available case analysis (follow-up mean 7 weeks; measured with: perinatal PTSD Questionnaire (PPQ); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	none	48	50	-	SMD 0.1 lower (0.5 lower to 0.29 higher)	⊕⊕OO LOW	
PTSD symptomatology Intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (follow-up mean 25 weeks; assessed with: perinatal PTSD Questionnaire (PPQ): scores in clinical range (no further detail))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	22/60 (36.7%)	22/61 (36.1%)	RR 1.02 (0.63 to 1.63)	7 more per 1000 (from 133 fewer to 227 more)	⊕⊕OO LOW	
								36.1%		7 more per 1000 (from 134 fewer to 227 more)		
PTSD symptomatology Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 25 weeks; assessed with: perinatal PTSD Questionnaire (PPQ): scores in clinical range (no further detail))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	8/46 (17.4%)	11/50 (22%)	RR 0.79 (0.35 to 1.79)	46 fewer per 1000 (from 143 fewer to 174 more)	⊕⊕OO LOW	
								22%		46 fewer per 1000 (from 143 fewer to 174 more)		
PTSD mean scores Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 25 weeks; measured with: perinatal PTSD Questionnaire (PPQ); better indicated by lower values)												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	none	46	50	-	SMD 0.25 lower (0.66 lower to 0.15 higher)	⊕⊕OO LOW	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.44 OCD: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**OCD: psychologically (CBT/IPT)-informed psychoeducation versus TAU/enhanced TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
OCD mean scores post-treatment – available case analysis (follow-up mean 4 weeks; measured with: Yale–Brown Obsessive Compulsive Scale (YBOCS); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	33	25	-	SMD 0.41 lower (0.94 lower to 0.11 higher)	⊕OOO VERY LOW	
Obsessions mean scores post-treatment – available case analysis (follow-up mean 4 weeks; measured with: Yale–Brown Obsessive Compulsive Scale (YBOCS): Obsessions; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	33	25	-	SMD 0.39 lower (0.92 lower to 0.13 higher)	⊕OOO VERY LOW	
Compulsions mean scores post-treatment – available case analysis (follow-up mean 4 weeks; measured with: Yale–Brown Obsessive Compulsive Scale (YBOCS): Compulsions; better indicated by lower values)												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	33	25	-	SMD 0.31 lower (0.83 lower to 0.21 higher)	⊕OOO VERY LOW	
OCD mean scores Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 19 weeks; measured with: Yale–Brown Obsessive Compulsive Scale (YBOCS); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias[3]	31	19	-	SMD 0.71 lower (1.29 to 0.12 lower)	⊕OOO VERY LOW	
Obsessions mean scores Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 19 weeks; measured with: Yale–Brown Obsessive Compulsive Scale (YBOCS): Obsessions; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias[3]	31	19	-	SMD 0.65 lower (1.24 to 0.07 lower)	⊕OOO VERY LOW	
Compulsions mean scores Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 19 weeks; measured with: Yale–Brown Obsessive Compulsive Scale (YBOCS): Compulsions; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias[3]	31	19	-	SMD 0.7 lower (1.29 to 0.11 lower)	⊕OOO VERY LOW	
OCD mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 32 weeks; measured with: Yale–Brown Obsessive Compulsive Scale (YBOCS); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias[3]	29	20	-	SMD 0.76 lower (1.35 to 0.17 lower)	⊕OOO VERY LOW	
Obsessions mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 32 weeks; measured with: Yale–Brown Obsessive Compulsive Scale (YBOCS): Obsessions; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias[3]	29	20	-	SMD 0.73 lower (1.32 to 0.14 lower)	⊕OOO VERY LOW	

Compulsions mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 32 weeks; measured with: Yale–Brown Obsessive Compulsive Scale (YBOCS): Compulsions; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	29	20	-	SMD 0.72 lower (1.31 to 0.13 lower)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data

1.3.45 Fear of childbirth: pre-delivery discussion/psychoeducation versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Fear of childbirth: pre-delivery discussion/psychoeducation versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Elective caesarean post-treatment – ITT analysis (follow-up 0-16 weeks; assessed with: mode of delivery: number of women delivering via elective caesarean or caesarean for psychosocial reasons)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	23/175 (13.1%)	39/286 (13.6%)	RR 0.93 (0.57 to 1.51)	10 fewer per 1000 (from 59 fewer to 70 more)	⊕⊕OO LOW	
								15.2%		11 fewer per 1000 (from 65 fewer to 78 more)		
Choosing vaginal delivery post-treatment – ITT analysis (follow-up mean 16 weeks; assessed with: Delivery preference: number of women choosing vaginal delivery)												
1	randomised trials	no serious	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	35/44 (79.5%)	35/46 (76.1%)		38 more per 1000 (from 122	⊕OOO VERY LOW	

		risk of bias							RR 1.05 (0.84 to 1.3)	fewer to 228 more)		
								76.1%		38 more per 1000 (from 122 fewer to 228 more)		
Vaginal delivery post-treatment – ITT analysis (follow-up 0-16 weeks; assessed with: mode of delivery: spontaneous vaginal delivery/vaginal delivery)												
2	randomised trials	no serious risk of bias	serious[4]	no serious indirectness	very serious[1,2]	none	108/175 (61.7%)	141/287 (49.1%)	RR 1.2 (0.9 to 1.59)	98 more per 1000 (from 49 fewer to 290 more)	⊕OOO VERY LOW	
								52.5%		105 more per 1000 (from 53 fewer to 310 more)		
Fear of pain in labour mean score Mid-treatment (36 weeks gestation) – ITT analysis (follow-up mean 12 weeks; measured with: pregnancy Anxiety Scale: Fear of pain in labour; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	85	91	-	SMD 0.09 lower (0.39 lower to 0.2 higher)	⊕OOO VERY LOW	
Fear of obstetrician's unfriendly behaviour mean scores Mid-treatment (36 weeks gestation) – ITT analysis (follow-up mean 12 weeks; measured with: pregnancy Anxiety Scale: Fear of obstretrician's unfriendly behaviour; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,5]	reporting bias[3]	85	91	-	SMD 0.23 lower (0.53 lower to 0.07 higher)	⊕OOO VERY LOW	
Preparedness for childbirth mean scores Mid-treatment (36 weeks gestation) – available case analysis (follow-up mean 8 weeks; measured with: preparedness for childbirth (study-specific scale); better indicated by lower values)												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[5]	none	96	158	-	SMD 0.19 higher (0.07 lower to 0.44 higher)	⊕⊕⊕O MODERATE	
Satisfaction with childbirth mean scores post-treatment – ITT analysis (follow-up mean 29 weeks; measured with: study-specific scale: satisfaction with childbirth; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,5]	reporting bias[3]	85	91	-	SMD 0.22 lower (0.52 lower to 0.08 higher)	⊕OOO VERY LOW	
Feeling safe during childbirth mean scores post-treatment – ITT analysis (follow-up mean 29 weeks; measured with: satisfaction with childbirth: Feeling safe (study-specific scale); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias[3]	85	91	-	SMD 0.39 lower (0.69 to 0.09 lower)	⊕OOO VERY LOW	
Experience of fear during childbirth mean scores post-treatment – ITT analysis (follow-up mean 13 weeks; measured with: Wijma Delivery Experience Questionnaire (W-DEQ-B); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[5]	none	131	240	-	SMD 0.35 lower (0.57 to 0.14 lower)	⊕⊕⊕O MODERATE	
Maternal attitude to motherhood mean scores post-treatment – available case analysis (follow-up mean 25 weeks; measured with: motherhood and parenting (based on Kumar, Robson & Smith, 1984); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[5]	none	92	160	-	SMD 0.3 higher (0.04 to 0.56 higher)	⊕⊕⊕O MODERATE	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

[3] Paper omits data
[4] There was evidence of moderate heterogeneity between effect sizes
[5] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.46 Eating disorder: mother–infant relationship interventions (and guided self-help) versus listening visits (and guided self-help)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Eating disorder: mother–infant relationship interventions (and guided self-help) versus listening visits (and guided self-help)	Control	Relative (95% CI)	Absolute		
Eating disorder diagnosis post-treatment – ITT analysis (follow-up mean 35 weeks; assessed with: psychiatric interview: DSM-IV Eating Disorder)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	14/40 (35%)	13/40 (32.5%)	RR 1.08 (0.58 to 1.99)	26 more per 1000 (from 137 fewer to 322 more)	⊕OOO VERY LOW	
								32.5%		26 more per 1000 (from 137 fewer to 322 more)		
Eating disorder diagnosis post-treatment – available case analysis (follow-up mean 35 weeks; assessed with: psychiatric interview: DSM-IV Eating Disorder)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	11/37 (29.7%)	12/39 (30.8%)	RR 0.97 (0.49 to 1.91)	9 fewer per 1000 (from 157 fewer to 280 more)	⊕OOO VERY LOW	
								30.8%		9 fewer per 1000 (from 157		

										fewer to 280 more)		

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data

1.3.47 General mental health: structured psychological interventions (CBT or IPT) versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**General mental health: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
General mental health mean scores post-treatment – ITT analysis (follow-up mean 15 weeks; measured with: Brief Symptom Inventory (BSI): Global severity index (Mental health); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	47	46	-	SMD 0.76 lower (1.19 to 0.34 lower)	⊕⊕OO LOW	
General mental health (higher better) mean scores post-treatment – available case analysis (follow-up 15-26 weeks; measured with: sF-12 Mental Component Summary (SF-MCS); better indicated by lower values)												
2	randomised trials	no serious risk of bias	very serious[2]	no serious indirectness	very serious[1,3]	reporting bias[4]	150	155	-	SMD 0.68 higher (0.08 lower to 1.44 higher)	⊕OOO VERY LOW	
Risk of self-harm mean scores post-treatment – available case analysis (follow-up mean 26 weeks; measured with: Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM): Risk of self-harm; better indicated by lower values)												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias[4]	138	145	-	SMD 0.31 lower (0.55 to 0.08 lower)	⊕⊕OO LOW	
General mental health mean scores short follow-up (9-16 weeks post-intervention) – ITT analysis (follow-up mean 28 weeks; measured with: Brief Symptom Inventory (BSI): Global severity index (Mental health); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	47	46	-	SMD 0.73 lower (1.15 to 0.31 lower)	⊕⊕OO LOW	
General mental health mean scores Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 33 weeks; measured with: sF-12 Mental Component Summary (SF-MCS); better indicated by lower values)												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	very serious[1,3]	none	15	11	-	SMD 0.78 higher (0.03 lower to 1.59 higher)	⊕OOO VERY LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] There was evidence of substantial heterogeneity between effect sizes
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] Papers omit data
[5] Risk of bias due to statistically significant group differences at baseline

1.3.48 General mental health: IPT versus support group

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**General mental health: IPT versus support group**	**Control**	**Relative (95% CI)**	**Absolute**		
Anger mean scores post-treatment – available case analysis (follow-up mean 12 weeks; measured with: state Anger Inventory (STAXI); better indicated by lower values)												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	22	22	-	SMD 0.09 lower (0.68 lower to 0.5 higher)	⊕OOO VERY LOW	

[1] Risk of bias due to statistically significant group differences at baseline
[2] Total population size is less than 400 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.49 General mental health: post-miscarriage self-help versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**General mental health: post-miscarriage self-help versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
General mental health symptomatology/treatment non-response post-treatment – ITT analysis (follow-up mean 5 weeks; assessed with: Brief Symptom Inventory (BSI): Global severity index (Treatment non-response: reliable change index))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	22/45 (48.9%)	23/33 (69.7%)	RR 0.7 (0.48 to 1.02)	209 fewer per 1000 (from 362 fewer to 14 more)	⊕⊕OO LOW	
								69.7%		209 fewer per 1000 (from 362 fewer to 14 more)		
General mental health treatment non-response post-treatment – available case analysis (follow-up mean 5 weeks; assessed with: Brief Symptom Inventory (BSI): Global severity index (Treatment non-response: reliable change index))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	10/33 (30.3%)	16/26 (61.5%)	RR 0.49 (0.27 to 0.9)	314 fewer per 1000 (from 62 fewer to 449 fewer)	⊕⊕OO LOW	

								61.5%		314 fewer per 1000 (from 62 fewer to 449 fewer)		
General mental health mean scores post-treatment – ITT analysis (follow-up mean 5 weeks; measured with: Brief Symptom Inventory (BSI): Global severity index (Mental health); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	none	45	33	-	SMD 0.67 lower (1.13 to 0.21 lower)	⊕⊕OO LOW	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.50 General mental health: listening visits versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**General mental health: listening visits versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
General mental health mean scores (higher better) post-treatment – available case analysis (follow-up mean 26 weeks; measured with: sF-12 Mental Component Summary (SF-MCS); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias[2]	129	142	-	SMD 0.42 higher (0.18 to 0.66 higher)	⊕⊕OO LOW	
Risk of self-harm mean score post-treatment – available case analysis (follow-up mean 26 weeks; measured with: Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM): Risk of self-harm; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias[2]	131	145	-	SMD 0.31 lower (0.55 to 0.07 lower)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] Paper omits data

1.3.51 General mental health: post-miscarriage counselling versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	General mental health: post-miscarriage counselling versus TAU	Control	Relative (95% CI)	Absolute		
Self-blame mean score post-treatment – available case analysis (follow-up mean 2 weeks; measured with: study-specific measure: self-blame; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	33	33	-	SMD 0.15 higher (0.34 lower to 0.63 higher)	⊕⊕OO LOW	
Self-blame mean score Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 17 weeks; measured with: study-specific measure: self-blame; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	33	33	-	SMD 0.03 higher (0.45 lower to 0.51 higher)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.52 General mental health: post-traumatic birth counselling versus TAU

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	General mental health: post-traumatic birth counselling versus TAU	Control	Relative (95% CI)	Absolute		
Self-blame mean scores post-treatment – ITT analysis (follow-up mean 13 weeks; measured with: study-specific measure: self-blame; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	50	53	-	SMD 2.37 higher (1.86 to 2.88 higher)	⊕⊕OO LOW	
Self-blame mean scores post-treatment – available case analysis (follow-up mean 13 weeks; measured with: study-specific measure: self-blame; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	50	53	-	SMD 2.37 higher (1.86 to 2.88 higher)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.53 General mental health: psychologically (CBT/IPT)-informed psychoeducation versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	General mental health: psychologically (CBT/IPT)-informed psychoeducation versus TAU/enhanced TAU	Control	Relative (95% CI)	Absolute		
Any psychopathology diagnosis post-treatment – ITT analysis (follow-up mean 52 weeks; assessed with: schedule for Affective Disorders and Schizophrenia (SADS): any psychopathology)												
1	randomised trials	no serious	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	38/101 (37.6%)	36/98 (36.7%)	RR 1.02 (0.71 to 1.47)	7 more per 1000 (from 107 fewer to 173 more)	⊕OOO VERY LOW	

		risk of bias						36.7%		7 more per 1000 (from 106 fewer to 172 more)		
Any psychopathology diagnosis post-treatment – available case analysis (follow-up mean 52 weeks; assessed with: schedule for Affective Disorders and Schizophrenia (SADS): any psychopathology)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	38/101 (37.6%)	36/98 (36.7%)	RR 1.02 (0.71 to 1.47)	7 more per 1000 (from 107 fewer to 173 more)	⊕OOO VERY LOW	
								36.7%		7 more per 1000 (from 106 fewer to 172 more)		
General mental health mean scores post-treatment – ITT analysis (follow-up mean 6 weeks; measured with: General Health Questionnaire (GHQ); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	none	96	98	-	SMD 0.48 lower (0.76 to 0.19 lower)	⊕⊕OO LOW	
General mental health mean scores short follow-up (9-16 weeks post-intervention) – ITT analysis (follow-up mean 13 weeks; measured with: General Health Questionnaire (GHQ); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	none	96	98	-	SMD 0.16 lower (0.44 lower to 0.12 higher)	⊕⊕OO LOW	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data
[4] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.54 General mental health: home visits versus TAU/enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	General mental health: home visits versus TAU/enhanced TAU	Control	Relative (95% CI)	Absolute		
General mental health symptomatology/treatment non-response post-treatment – ITT analysis (follow-up mean 104 weeks; assessed with: mental Health Index (MHI-5)<67)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	91/179 (50.8%)	101/185 (54.6%)	RR 0.93 (0.77 to 1.13)	38 fewer per 1000 (from 126 fewer to 71 more)	⊕⊕OO LOW	
								54.6%		38 fewer per 1000 (from 126 fewer to 71 more)		
General mental health symptomatology/treatment non-response post-treatment – available case analysis (follow-up mean 104 weeks; assessed with: mental Health Index (MHI-5)<67)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	38/126 (30.2%)	39/123 (31.7%)	RR 0.95 (0.66 to 1.38)	16 fewer per 1000 (from 108 fewer to 120 more)	⊕OOO VERY LOW	
								31.7%		16 fewer per 1000 (from 108 fewer to 120 more)		
Alcohol or drug use symptomatology post-treatment – ITT analysis (follow-up mean 104 weeks; assessed with: CAGE Questionnaire: Alcohol or drug use)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	88/179 (49.2%)	103/185 (55.7%)	RR 0.88 (0.73 to 1.08)	67 fewer per 1000 (from 150 fewer to 45 more)	⊕OOO VERY LOW	
								55.7%		67 fewer per 1000 (from 150 fewer to 45 more)		
Alcohol or drug use symptomatology post-treatment – available case analysis (follow-up mean 104 weeks; assessed with: CAGE Questionnaire: Alcohol or drug use)												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	35/126 (27.8%)	41/123 (33.3%)	RR 0.83 (0.57 to 1.21)	57 fewer per 1000 (from 143 fewer to 70 more)	⊕OOO VERY LOW	
								33.3%		57 fewer per 1000 (from 143 fewer to 70 more)		

[1] Risk of bias due to statistically significant group differences at baseline
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.55 General mental health: mother–infant relationship interventions versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**General mental health: mother–infant relationship interventions versus TAU/enhanced TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
General mental health treatment non-response post-treatment – ITT analysis (follow-up mean 26 weeks; assessed with: symptom Checklist-90 (SCL-90): Global Severity Index (GSI): Treatment non-response (no improvement-reliable change index))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	23/40 (57.5%)	20/40 (50%)	RR 1.15 (0.76 to 1.73)	75 more per 1000 (from 120 fewer to 365 more)	⊕OOO VERY LOW	
								50%		75 more per 1000 (from 120 fewer to 365 more)		
General mental health treatment non-response post-treatment – available case analysis (follow-up mean 26 weeks; assessed with: symptom Checklist-90 (SCL-90): Global Severity Index (GSI): Treatment non-response (no improvement-reliable change index))												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	21/38 (55.3%)	17/37 (45.9%)	RR 1.2 (0.77 to 1.89)	92 more per 1000 (from 106 fewer to 409 more)	⊕OOO VERY LOW	
								46%		92 more per 1000 (from 106 fewer to 409 more)		
General mental health mean scores (lower better) post-treatment – available case analysis (follow-up mean 26 weeks; measured with: symptom Checklist-90 (SCL-90): Global Severity Index (GSI); better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3,4]	none	38	37	-	SMD 0.24 lower (0.7 lower to 0.21 higher)	⊕OOO VERY LOW	

[1] Risk of bias due to statistically significant group differences at baseline
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.56 General mental health: co-parenting intervention versus enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**General mental health: co-parenting intervention versus enhanced TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Psychological distress mean scores post-treatment – available case analysis (follow-up mean 6 weeks; measured with: Kellner Symptom Questionnaire: psychological distress; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	15	13	-	SMD 0.65 lower (1.42 lower to 0.11 higher)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.57 Mother–infant attachment: structured psychological interventions (CBT or IPT) versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mother–infant attachment: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU	Control	Relative (95% CI)	Absolute		
Mother–infant attachment problems post-treatment – ITT analysis (follow-up mean 20 weeks; assessed with: maternal report: mother–infant relationship problems)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	27/50 (54%)	43/52 (82.7%)	RR 0.65 (0.49 to 0.87)	289 fewer per 1000 (from 107 fewer to 422 fewer)	⊕⊕OO LOW	
								82.7%		289 fewer per 1000 (from 108 fewer to 422 fewer)		
Mother–infant attachment problems post-treatment – available case analysis (follow-up mean 20 weeks; assessed with: maternal report: mother–infant relationship problems)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	20/43 (46.5%)	26/35 (74.3%)	RR 0.63 (0.43 to 0.91)	275 fewer per 1000 (from 67 fewer to 423 fewer)	⊕⊕OO LOW	
								74.3%		275 fewer per 1000 (from 67 fewer to 424 fewer)		

Mother–infant attachment mean score post-treatment – available case analysis (follow-up 8-15 weeks; measured with: prenatal Attachment Inventory or Maternal Attachment Inventory (MAI); better indicated by lower values)												
2	randomised trials	no serious risk of bias	very serious[2]	no serious indirectness	very serious[3,4]	none	39	37	-	SMD 2.28 higher (1.17 lower to 5.73 higher)	⊕OOO VERY LOW	
Mother–infant play frequency post-treatment – ITT analysis (follow-up mean 52 weeks; assessed with: mother–infant interaction: play frequency (Events were mother played with infant once or more every day))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	247/463 (53.3%)	149/440 (33.9%)	RR 1.58 (1.35 to 1.84)	196 more per 1000 (from 119 more to 284 more)	⊕⊕⊕⊕ HIGH	
								33.9%		197 more per 1000 (from 119 more to 285 more)		
Mother–infant play frequency post-treatment – available case analysis (follow-up mean 52 weeks; assessed with: mother–infant interaction: play frequency (Events were mother played with infant once or more every day))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	247/360 (68.6%)	149/345 (43.2%)	RR 1.59 (1.38 to 1.83)	255 more per 1000 (from 164 more to 358 more)	⊕⊕⊕⊕ HIGH	
								43.2%		255 more per 1000 (from 164 more to 359 more)		
Maternal sensitivity mean scores post-treatment – available case analysis (follow-up mean 15 weeks; measured with: study-specific task: Attentional bias for distressed infant faces reaction time paradigm; better indicated by lower values)												

1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	very serious[3,4]	reporting bias[6]	10	7	-	SMD 0.86 higher (0.16 lower to 1.88 higher)	⊕OOO VERY LOW	
Mother–infant behaviour management problems post-treatment – ITT analysis (follow-up mean 20 weeks; assessed with: maternal report: Behaviour management problems)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,4]	none	26/50 (52%)	30/52 (57.7%)	RR 0.9 (0.63 to 1.28)	58 fewer per 1000 (from 213 fewer to 162 more)	⊕⊕OO LOW	
								57.7%		58 fewer per 1000 (from 213 fewer to 162 more)		
Mother–infant behaviour management problems post-treatment – available case analysis (follow-up mean 20 weeks; assessed with: maternal report: Behaviour management problems)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,4]	none	19/43 (44.2%)	13/35 (37.1%)	RR 1.19 (0.69 to 2.05)	71 more per 1000 (from 115 fewer to 390 more)	⊕⊕OO LOW	
								37.1%		70 more per 1000 (from 115 fewer to 390 more)		
Discontinued (exclusive) breastfeeding <6 months – ITT analysis (follow-up mean 52 weeks; assessed with: infant feeding-no longer exclusively breastfeeding by 26 weeks)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	400/463 (86.4%)	400/440 (90.9%)	RR 0.95 (0.91 to 1)	45 fewer per 1000 (from 82 fewer to 0 more)	⊕⊕⊕⊕ HIGH	
								90.9%		45 fewer per 1000 (from 82 fewer to 0 more)		

Discontinued (exclusive) breastfeeding <6 months post-treatment – available case analysis (follow-up mean 52 weeks; assessed with: infant feeding-no longer exclusively breastfeeding by 26 weeks)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	305/368 (82.9%)	319/359 (88.9%)	RR 0.93 (0.88 to 0.99)	62 fewer per 1000 (from 9 fewer to 107 fewer)	⊕⊕⊕⊕ HIGH	
								88.9%		62 fewer per 1000 (from 9 fewer to 107 fewer)		
Mother–infant attachment mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (follow-up mean 21 weeks; measured with: maternal Attachment Inventory (MAI); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3,4]	none	22	23	-	SMD 0.32 higher (0.27 lower to 0.91 higher)	⊕⊕OO LOW	
Mother–infant attachment problems long follow-up (25-103 weeks post-intervention) – ITT analysis (follow-up mean 78 weeks; assessed with: maternal report: mother–infant relationship problems)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,4]	none	31/50 (62%)	25/52 (48.1%)	RR 1.29 (0.9 to 1.84)	139 more per 1000 (from 48 fewer to 404 more)	⊕⊕OO LOW	
								48.1%		139 more per 1000 (from 48 fewer to 404 more)		
Mother–infant attachment problems long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 78 weeks; assessed with: maternal report: mother–infant relationship problems)												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,4]	none	21/40 (52.5%)	20/47 (42.6%)	RR 1.23 (0.79 to 1.92)	98 more per 1000 (from 89 fewer to 391 more)	⊕⊕OO LOW	
								42.6%		98 more per 1000 (from 89 fewer to 392 more)		

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] There is evidence of considerable heterogeneity of study effect sizes
[3] Total population size is less than 400 (a threshold rule-of-thumb)
[4] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[5] Risk of bias due to unclear blinding of outcome assessment
[6] Paper omits data

1.3.58 Mother–infant attachment: facilitated self-help versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Mother–infant attachment: facilitated self-help versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Maternal attitude towards motherhood mean scores post-treatment – available case analysis (follow-up mean 17 weeks; measured with: postnatal Bonding Questionnaire (PBQ); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	31	28	-	SMD 0.41 higher (0.11 lower to 0.92 higher)	⊕OOO VERY LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data

1.3.59 Mother–infant attachment: listening visits versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mother–infant attachment: listening visits versus TAU	Control	Relative (95% CI)	Absolute		
Mother–infant attachment problems post-treatment – ITT analysis (follow-up mean 20 weeks; assessed with: maternal report: mother–infant relationship problems)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	28/48 (58.3%)	43/52 (82.7%)	RR 0.71 (0.54 to 0.92)	240 fewer per 1000 (from 66 fewer to 380 fewer)	⊕⊕OO LOW	
								82.7%		240 fewer per 1000 (from 66 fewer to 380 fewer)		
Mother–infant attachment problems post-treatment – available case analysis (follow-up mean 20 weeks; assessed with: maternal report: mother–infant relationship problems)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	23/43 (53.5%)	26/35 (74.3%)	RR 0.72 (0.51 to 1.01)	208 fewer per 1000 (from 364 fewer to 7 more)	⊕⊕OO LOW	
								74.3%		208 fewer per 1000 (from 364 fewer to 7 more)		
Mother–infant behaviour management problems post-treatment – ITT analysis (follow-up mean 20 weeks; assessed with: maternal report: Behaviour management problems)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	20/48 (41.7%)	30/52 (57.7%)	RR 0.72 (0.48 to 1.09)	162 fewer per 1000 (from 300 fewer to 52 more)	⊕⊕OO LOW	
								57.7%		162 fewer per 1000 (from 300 fewer to 52 more)		

Mother–infant behaviour management problems post-treatment – available case analysis (follow-up mean 20 weeks; assessed with: maternal report: Behaviour management problems)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	15/43 (34.9%)	13/35 (37.1%)	RR 0.94 (0.52 to 1.7)	22 fewer per 1000 (from 178 fewer to 260 more)	⊕⊕OO LOW	
								37.1%		22 fewer per 1000 (from 178 fewer to 260 more)		
Discontinued breastfeeding <6 months – ITT analysis (follow-up mean 52 weeks; assessed with: infant feeding-breast feeding stopped by 26 weeks)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	77/183 (42.1%)	210/548 (38.3%)	RR 1.1 (0.9 to 1.34)	38 more per 1000 (from 38 fewer to 130 more)	⊕⊕OO LOW	
								38.3%		38 more per 1000 (from 38 fewer to 130 more)		
Discontinued breastfeeding <6 months post-treatment – available case analysis (follow-up mean 52 weeks; assessed with: infant feeding-breast feeding stopped by 26 weeks)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	77/140 (55%)	210/417 (50.4%)	RR 1.09 (0.91 to 1.3)	45 more per 1000 (from 45 fewer to 151 more)	⊕⊕OO LOW	
								50.4%		45 more per 1000 (from 45 fewer to 151 more)		
Mother–infant attachment problems long follow-up (25-103 weeks post-intervention) – ITT analysis (follow-up mean 78 weeks; assessed with: maternal report: mother–infant relationship problems)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	25/48 (52.1%)	25/52 (48.1%)	RR 1.08 (0.73 to 1.6)	38 more per 1000 (from 130 fewer to 288 more)	⊕⊕OO LOW	

								48.1%		38 more per 1000 (from 130 fewer to 289 more)		
Mother–infant attachment problems long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 78 weeks; assessed with: maternal report: mother–infant relationship problems)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	16/39 (41%)	20/47 (42.6%)	RR 0.96 (0.58 to 1.59)	17 fewer per 1000 (from 179 fewer to 251 more)	⊕⊕OO LOW	
								42.6%		17 fewer per 1000 (from 179 fewer to 251 more)		

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.60 Mother–infant attachment: social support versus TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Mother–infant attachment: social support versus TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Mother–infant feeding interaction mean scores post-treatment – available case analysis (follow-up mean 12 weeks; measured with: nursing Child Assessment Satellite Training Scale (NCAST): Feeding; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	19	24	-	SMD 0.18 lower (0.79 lower to 0.42 higher)	⊕⊕OO LOW	
Mother–infant teaching interaction mean scores post-treatment – available case analysis (follow-up mean 12 weeks; measured with: nursing Child Assessment Satellite Training Scale (NCAST): Teaching; better indicated by lower values)												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	21	25	-	SMD 0.45 lower (1.04 lower to 0.13 higher)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.61 Mother–infant attachment: psychologically (CBT/IPT)-informed psychoeducation versus enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Mother–infant attachment: psychologically (CBT/IPT)-informed psychoeducation versus enhanced TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Maternal competence/confidence mean scores post-treatment – available case analysis (follow-up mean 6 weeks; measured with: parenting Sense of Competence Scale (PSCS): Efficacy; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	96	98	-	SMD 0.57 higher (0.29 to 0.86 higher)	⊕⊕OO LOW	
Maternal competence/confidence mean scores short follow-up (9-16 weeks post-intervention) – available case analysis (follow-up mean 13 weeks; measured with: parenting Sense of Competence Scale (PSCS): Efficacy; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	96	98	-	SMD 0.35 higher (0.06 to 0.63 higher)	⊕⊕OO LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.62 Mother–infant attachment: home visits versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mother–infant attachment: home visits versus TAU/enhanced TAU	Control	Relative (95% CI)	Absolute		
Mother–infant attachment problems post-treatment – ITT analysis (follow-up mean 104 weeks; assessed with: nursing Child Assessment Satellite Training Scale (NCAST)<=35)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	74/179 (41.3%)	88/185 (47.6%)	RR 0.87 (0.69 to 1.09)	62 fewer per 1000 (from 147 fewer to 43 more)	⊕OOO VERY LOW	
								47.6%		62 fewer per 1000 (from 148 fewer to 43 more)		
Mother–infant attachment problems post-treatment – available case analysis (follow-up mean 104 weeks; assessed with: nursing Child Assessment Satellite Training Scale (NCAST)<=35)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	21/126 (16.7%)	26/123 (21.1%)	RR 0.79 (0.47 to 1.32)	44 fewer per 1000 (from 112 fewer to 68 more)	⊕OOO VERY LOW	
								21.1%		44 fewer per 1000 (from 112 fewer to 68 more)		

[1] Risk of bias due to statistically significant group differences at baseline
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.63 Mother–infant attachment: mother–infant relationship interventions versus TAU or enhanced TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mother–infant attachment: mother–infant relationship interventions versus TAU/enhanced TAU	Control	Relative (95% CI)	Absolute		
Mother–infant attachment problems post-treatment – ITT analysis (follow-up 20-26 weeks; assessed with: maternal report: mother–infant relationship problems or Parent-Infant Relationship Global Assessment Scale (PIR-GAS): Treatment non-response (no improvement-reliable change index))												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	36/83 (43.4%)	73/92 (79.3%)	RR 0.55 (0.42 to 0.72)	357 fewer per 1000 (from 222 fewer to 460 fewer)	⊕OOO VERY LOW	
								78.9%		355 fewer per 1000 (from 221 fewer to 458 fewer)		
Mother–infant attachment problems post-treatment – available case analysis (follow-up 20-26 weeks; assessed with: maternal report: mother–infant relationship problems or Parent-Infant Relationship Global Assessment Scale (PIR-GAS): Treatment non-response (no improvement-reliable change index))												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	32/79 (40.5%)	53/72 (73.6%)	RR 0.55 (0.41 to 0.74)	331 fewer per 1000 (from 191 fewer to 434 fewer)	⊕OOO VERY LOW	
								73.6%		331 fewer per 1000 (from 191 fewer to 434 fewer)		

Mother–infant positive interaction mean scores post-treatment – available case analysis (follow-up 5-26 weeks; measured with: Dyadic Mutuality Code (DMC) or Parent-Infant Relationship Global Assessment Scale (PIR-GAS) or Behavioural observation: positive mother–infant interaction or Global Rating Scales of Mother–Infant Interaction: Overall mother–infant interaction; better indicated by lower values)												
4	randomised trials	no serious risk of bias	very serious[3]	no serious indirectness	very serious[4,5]	none	197	181	-	SMD 0.15 higher (0.26 lower to 0.56 higher)	⊕OOO VERY LOW	
Maternal sensitivity treatment response post-treatment – ITT analysis (follow-up mean 26 weeks; assessed with: Emotional Availability Scales (EAS): maternal sensitivity: Treatment response (improvement-reliable change index))												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[2,5]	reporting bias[7]	5/40 (12.5%)	3/40 (7.5%)	RR 1.67 (0.43 to 6.51)	50 more per 1000 (from 43 fewer to 413 more)	⊕OOO VERY LOW	
								7.5%		50 more per 1000 (from 43 fewer to 413 more)		
Maternal sensitivity treatment response post-treatment – available case analysis (follow-up mean 26 weeks; assessed with: Emotional Availability Scales (EAS): maternal sensitivity: Treatment response (improvement-reliable change index))												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[2,5]	reporting bias[7]	5/38 (13.2%)	3/37 (8.1%)	RR 1.62 (0.42 to 6.31)	50 more per 1000 (from 47 fewer to 431 more)	⊕OOO VERY LOW	
								8.1%		50 more per 1000 (from 47 fewer to 430 more)		
Maternal sensitivity mean scores post-treatment – available case analysis (follow-up 5-28 weeks; measured with: Emotional Availability Scales (EAS): maternal sensitivity or Behavioural observation: maternal sensitivity or Global Rating Scales of Mother–Infant Interaction: maternal sensitive behaviour; better indicated by lower values)												

4	randomised trials	no serious risk of bias	serious[8]	no serious indirectness	very serious[4,5]	none	172	160	-	SMD 0.23 higher (0.08 lower to 0.53 higher)	⊕OOO VERY LOW	
Maternal structuring treatment response post-treatment – ITT analysis (follow-up mean 26 weeks; assessed with: Emotional Availability Scales (EAS): maternal structuring: Treatment response (improvement-reliable change index))												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[2,5]	reporting bias[7]	6/40 (15%)	4/40 (10%)	RR 1.5 (0.46 to 4.91)	50 more per 1000 (from 54 fewer to 391 more)	⊕OOO VERY LOW	
								10%		50 more per 1000 (from 54 fewer to 391 more)		
Maternal structuring treatment response post-treatment – available case analysis (follow-up mean 26 weeks; assessed with: Emotional Availability Scales (EAS): maternal structuring: Treatment response (improvement-reliable change index))												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[2,5]	reporting bias[7]	6/38 (15.8%)	4/37 (10.8%)	RR 1.46 (0.45 to 4.76)	50 more per 1000 (from 59 fewer to 406 more)	⊕OOO VERY LOW	
								10.8%		50 more per 1000 (from 59 fewer to 406 more)		
Maternal structuring mean scores post-treatment – available case analysis (follow-up 26-28 weeks; measured with: Emotional Availability Scales (EAS): maternal structuring; better indicated by lower values)												
2	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[4,5]	reporting bias[7]	73	73	-	SMD 0.25 higher (0.07 lower to 0.58 higher)	⊕OOO VERY LOW	

Maternal nonintrusiveness treatment response post-treatment – ITT analysis (follow-up mean 26 weeks; assessed with: Emotional Availability Scales (EAS): maternal nonintrusiveness: Treatment response (improvement-reliable change index))												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[2,5]	reporting bias[7]	6/40 (15%)	7/40 (17.5%)	RR 0.86 (0.32 to 2.33)	24 fewer per 1000 (from 119 fewer to 233 more)	⊕OOO VERY LOW	
								17.5%		24 fewer per 1000 (from 119 fewer to 233 more)		
Maternal nonintrusiveness treatment response post-treatment – available case analysis (follow-up mean 26 weeks; assessed with: Emotional Availability Scales (EAS): maternal nonintrusiveness: Treatment response (improvement-reliable change index))												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[2,5]	reporting bias[7]	6/38 (15.8%)	7/37 (18.9%)	RR 0.83 (0.31 to 2.25)	32 fewer per 1000 (from 131 fewer to 236 more)	⊕OOO VERY LOW	
								18.9%		32 fewer per 1000 (from 130 fewer to 236 more)		
Maternal nonintrusive behaviour mean scores post-treatment – available case analysis (follow-up 26-28 weeks; measured with: Emotional Availability Scales (EAS): maternal nonintrusiveness; better indicated by lower values)												
2	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[4,5]	reporting bias[7]	73	73	-	SMD 0.24 higher (0.08 lower to 0.57 higher)	⊕OOO VERY LOW	
Maternal intrusive behaviour mean scores post-treatment – available case analysis (follow-up mean 7 weeks; measured with: Global Rating Scales of Mother–Infant Interaction: maternal intrusive behaviour; better indicated by lower values)												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4,5]	none	48	50	-	SMD 0.28 higher (0.11 lower to 0.68 higher)	⊕⊕OO LOW	
Maternal nonhostility mean scores post-treatment – available case analysis (follow-up mean 28 weeks; measured with: Emotional Availability Scales (EAS): maternal nonhostility; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4,5]	reporting bias[9]	35	36	-	SMD 0.1 higher (0.37 lower to 0.57 higher)	⊕OOO VERY LOW	
Child responsiveness treatment response post-treatment – ITT analysis (follow-up mean 26 weeks; assessed with: Emotional Availability Scales (EAS): child responsiveness: Treatment response (improvement-reliable change index))												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[2,5]	reporting bias[7]	3/40 (7.5%)	4/40 (10%)	RR 0.75 (0.18 to 3.14)	25 fewer per 1000 (from 82 fewer to 214 more)	⊕OOO VERY LOW	
								10%		25 fewer per 1000 (from 82 fewer to 214 more)		
Child responsiveness treatment response post-treatment – available case analysis (follow-up mean 26 weeks; assessed with: Emotional Availability Scales (EAS): child responsiveness: Treatment response (improvement-reliable change index))												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[2,5]	reporting bias[7]	3/38 (7.9%)	4/37 (10.8%)	RR 0.73 (0.18 to 3.04)	29 fewer per 1000 (from 89 fewer to 221 more)	⊕OOO VERY LOW	
								10.8%		29 fewer per 1000 (from 89 fewer to 220 more)		

Child responsiveness mean scores post-treatment – available case analysis (follow-up 26-28 weeks; measured with: Emotional Availability Scales (EAS): child responsiveness; better indicated by lower values)												
2	randomised trials	serious[6]	very serious[3]	no serious indirectness	very serious[4,5]	reporting bias[7]	73	73	-	SMD 0.38 higher (0.15 lower to 0.92 higher)	⊕OOO VERY LOW	
Child involvement treatment response post-treatment – ITT analysis (follow-up mean 26 weeks; assessed with: Emotional Availability Scales (EAS): child involvement: Treatment response (improvement-reliable change index))												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[2,5]	reporting bias[7]	7/40 (17.5%)	7/40 (17.5%)	RR 1 (0.39 to 2.59)	0 fewer per 1000 (from 107 fewer to 278 more)	⊕OOO VERY LOW	
								17.5%		0 fewer per 1000 (from 107 fewer to 278 more)		
Child involvement treatment response post-treatment – available case analysis (follow-up mean 26 weeks; assessed with: Emotional Availability Scales (EAS): child involvement: Treatment response (improvement-reliable change index))												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[2,5]	reporting bias[7]	7/38 (18.4%)	7/37 (18.9%)	RR 0.97 (0.38 to 2.5)	6 fewer per 1000 (from 117 fewer to 284 more)	⊕OOO VERY LOW	
								18.9%		6 fewer per 1000 (from 117 fewer to 283 more)		
Child involvement/positive engagement mean scores post-treatment – available case analysis (follow-up 5-28 weeks; measured with: Emotional Availability Scales (EAS): child involvement or Behavioural observation: child involvement or Global Rating Scales of Mother–Infant Interaction: infant positive engagement; better indicated by lower values)												

4	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	none	172	160	-	SMD 0.14 higher (0.09 lower to 0.37 higher)	⊕⊕⊕O MODERATE	
Child attachment security mean scores post-treatment – available case analysis (follow-up mean 57 weeks; measured with: Attachment Q Set (AQS III): child attachment security; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4,5]	none	35	36	-	SMD 0.45 higher (0.02 lower to 0.93 higher)	⊕⊕OO LOW	
Mother–infant behaviour management problems post-treatment – ITT analysis (follow-up mean 20 weeks; assessed with: maternal report: Behaviour management problems)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	none	15/43 (34.9%)	30/52 (57.7%)	RR 0.6 (0.38 to 0.97)	231 fewer per 1000 (from 17 fewer to 358 fewer)	⊕⊕OO LOW	
								57.7%		231 fewer per 1000 (from 17 fewer to 358 fewer)		
Mother–infant behaviour management problems post-treatment – available case analysis (follow-up mean 20 weeks; assessed with: maternal report: Behaviour management problems)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,5]	none	13/41 (31.7%)	13/35 (37.1%)	RR 0.85 (0.46 to 1.59)	56 fewer per 1000 (from 201 fewer to 219 more)	⊕⊕OO LOW	
								37.1%		56 fewer per 1000 (from 200 fewer to 219 more)		

Maternal confidence/competence mean scores post-treatment – available case analysis (follow-up mean 25 weeks; measured with: maternal report: Beliefs about competence; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4,5]	none	46	50	-	SMD 0.12 lower (0.52 lower to 0.28 higher)	⊕⊕OO LOW	
Mother–infant positive interaction mean scores Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 25 weeks; measured with: Global Rating Scales of Mother–Infant Interaction: Overall mother–infant interaction; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	none	46	50	-	SMD 0 higher (0.4 lower to 0.4 higher)	⊕⊕OO LOW	
Maternal sensitivity mean scores Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 25 weeks; measured with: Global Rating Scales of Mother–Infant Interaction: maternal sensitive behaviour; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4,5]	none	46	50	-	SMD 0.15 higher (0.25 lower to 0.55 higher)	⊕⊕OO LOW	
Maternal intrusive behaviour mean scores Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (follow-up mean 25 weeks; measured with: Global Rating Scales of Mother–Infant Interaction: maternal intrusive behaviour; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4,5]	none	46	50	-	SMD 0.13 higher (0.27 lower to 0.53 higher)	⊕⊕OO LOW	
Mother–infant attachment problems long follow-up (25-103 weeks post-intervention) – ITT analysis (follow-up mean 78 weeks; assessed with: maternal report: mother–infant relationship problems)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,5]	none	24/43 (55.8%)	25/52 (48.1%)	RR 1.16 (0.79 to 1.71)	77 more per 1000 (from 101 fewer to 341 more)	⊕⊕OO LOW	

								48.1%		77 more per 1000 (from 101 fewer to 342 more)		
Mother–infant attachment problems long follow-up (25-103 weeks post-intervention) – available case (follow-up mean 78 weeks; assessed with: maternal report: mother–infant relationship problems)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,5]	none	22/41 (53.7%)	20/47 (42.6%)	RR 1.26 (0.81 to 1.95)	111 more per 1000 (from 81 fewer to 404 more)	⊕⊕OO LOW	
								42.6%		111 more per 1000 (from 81 fewer to 405 more)		
Maternal sensitivity mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 57 weeks; measured with: Emotional Availability Scales (EAS): maternal sensitivity; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	none	35	36	-	SMD 0.81 higher (0.33 to 1.3 higher)	⊕⊕OO LOW	
Maternal structuring mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 57 weeks; measured with: Emotional Availability Scales (EAS): maternal structuring; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	none	35	36	-	SMD 0.56 higher (0.09 to 1.03 higher)	⊕⊕OO LOW	
Maternal nonintrusive behaviour mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 57 weeks; measured with: Emotional Availability Scales (EAS): maternal nonintrusiveness; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4,5]	none	35	36	-	SMD 0.34 higher (0.13	⊕⊕OO LOW	

										lower to 0.81 higher)		
Maternal nonhostility mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 57 weeks; measured with: Emotional Availability Scales (EAS): maternal nonhostility; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	none	35	36	-	SMD 0.02 lower (0.48 lower to 0.45 higher)	⊕⊕OO LOW	
Child responsiveness mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 57 weeks; measured with: Emotional Availability Scales (EAS): child responsiveness; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	none	35	36	-	SMD 0.68 higher (0.2 to 1.16 higher)	⊕⊕OO LOW	
Child involvement mean scores long follow-up (25-103 weeks post-intervention) – available case analysis (follow-up mean 57 weeks; measured with: Emotional Availability Scales (EAS): child involvement; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	none	35	36	-	SMD 0.74 higher (0.26 to 1.23 higher)	⊕⊕OO LOW	
Mother–infant positive interaction mean scores Very long follow-up (>104 weeks post-intervention) – available case analysis (follow-up mean 271 weeks; measured with: Behavioural observation: positive mother–infant interaction; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	none	29	29	-	SMD 1.82 lower (2.44 to 1.2 lower)	⊕⊕OO LOW	
Child attachment security mean scores Very long follow-up (>104 weeks post-intervention) – available case analysis (follow-up mean 271 weeks; measured with: Attatchment Story Completion Task; better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4,5]	none	29	29	-	SMD 0.42 higher (0.1	⊕⊕OO LOW	

										lower to 0.95 higher)		

[1] Risk of bias due to statistically significant group differences at baseline and non-blind outcome assessment
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] There is evidence of substantial heterogeneity of study effect sizes
[4] Total population size is less than 400 (a threshold rule-of-thumb)
[5] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[6] Risk of bias due to statistically significant group differences at baseline
[7] Paper omits data
[8] There is evidence of moderate heterogeneity of study effect sizes
[9] Evidence of selective reporting for this outcome measure

1.3.64 Mother–infant attachment: mother–infant relationship intervention with video feedback versus mother–infant relationship intervention with verbal feedback

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Mother–infant attachment: mother–infant relationship intervention with video feedback versus mother–infant relationship intervention with verbal feedback**	**Control**	**Relative (95% CI)**	**Absolute**		
Maternal confidence/competence mean scores post-treatment – available case analysis (follow-up mean 3 weeks; measured with: parenting Sense of Competence Scale (PSCS); better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	20	17	-	SMD 0.48 lower (1.13 lower to 0.18 higher)	⊕OOO VERY LOW	
Maternal perceptions of infant behaviour mean scores post-treatment – available case analysis (follow-up mean 3 weeks; measured with: neonatal Perception Inventory (NPI): maternal perceptions of infant behaviour; better indicated by lower values)												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias[3]	20	20	-	SMD 0.17 higher (0.45 lower to 0.8 higher)	⊕OOO VERY LOW	

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data

1.3.65 Mother–infant attachment: mother–infant relationship intervention (and guided self-help) versus listening visits (and guided self-help)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Mother–infant attachment: mother–infant relationship intervention (and guided self-help) versus listening visits (and guided self-help)**	**Control**	**Relative (95% CI)**	**Absolute**		
Mealtime conflict post-treatment – ITT analysis (follow-up mean 35 weeks; assessed with: Behavioural observation of mealtime: significant mealtime conflict (conflict was judged to have occurred if a conflict was at a severe or marked level of clinical concern [rating of 1 or 2] for any 2-minute observational period))												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias[2]	11/40 (27.5%)	22/40 (55%)	RR 0.5 (0.28 to 0.89)	275 fewer per 1000 (from 61 fewer to 396 fewer)	⊕OOO VERY LOW	
								55%		275 fewer per 1000 (from 61 fewer to 396 fewer)		
Mealtime conflict post-treatment – available case analysis (follow-up mean 35 weeks; assessed with: Behavioural observation of mealtime: significant mealtime conflict (conflict was judged to have occurred if a conflict was at a severe or marked level of clinical concern [rating of 1 or 2] for any 2-minute observational period))												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias[2]	9/38 (23.7%)	21/39 (53.8%)	RR 0.44 (0.23 to 0.83)	302 fewer per 1000 (from 92 fewer to 415 fewer)	⊕OOO VERY LOW	
								53.9%		302 fewer per 1000 (from 92 fewer to 415 fewer)		
Maternal inappropriate verbal responses post-treatment – ITT analysis (follow-up mean 35 weeks; assessed with: Behavioural observation of mealtime: maternal inappropriate verbal responses)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,3]	reporting bias[2]	19/40 (47.5%)	27/40 (67.5%)	RR 0.7 (0.48 to 1.04)	203 fewer per 1000 (from 351 fewer to 27 more)	⊕OOO VERY LOW	
								67.5%		203 fewer per 1000 (from 351 fewer to 27 more)		
Maternal inappropriate verbal responses post-treatment – available case analysis (follow-up mean 35 weeks; assessed with: Behavioural observation of mealtime: maternal inappropriate verbal responses)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,3]	reporting bias[2]	17/38 (44.7%)	26/39 (66.7%)	RR 0.67 (0.44 to 1.02)	220 fewer per 1000 (from 373 fewer to 13 more)	⊕OOO VERY LOW	
								66.7%		220 fewer per 1000 (from 374 fewer to 13 more)		
Maternal intrusions post-treatment – ITT analysis (follow-up mean 35 weeks; assessed with: Behavioural observation of mealtime: maternal intrusions)												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,3]	reporting bias[2]	13/40 (32.5%)	16/40 (40%)	RR 0.81 (0.45 to 1.46)	76 fewer per 1000 (from 220 fewer to 184 more)	⊕OOO VERY LOW	
								40%		76 fewer per 1000 (from 220 fewer to 184 more)		
Maternal intrusions post-treatment – available case analysis (follow-up mean 35 weeks; assessed with: Behavioural observation of mealtime: maternal intrusions)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,3]	reporting bias[2]	11/38 (28.9%)	15/39 (38.5%)	RR 0.75 (0.4 to 1.42)	96 fewer per 1000 (from 231 fewer to 162 more)	⊕OOO VERY LOW	
								38.5%		96 fewer per 1000 (from 231 fewer to 162 more)		
Infant autonomy post-treatment – ITT analysis (follow-up mean 35 weeks; assessed with: Behavioural observation of mealtime: infant autonomy)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias[2]	34/40 (85%)	25/40 (62.5%)	RR 1.36 (1.04 to 1.79)	225 more per 1000 (from 25 more to 494 more)	⊕OOO VERY LOW	
								62.5%		225 more per 1000 (from 25 more to 494 more)		
Infant autonomy post-treatment – available case analysis (follow-up mean 35 weeks; assessed with: Behavioural observation of mealtime: infant autonomy)												
1	randomised trials	no serious	no serious inconsistency	no serious indirectness	very serious[1,3]	reporting bias[2]	34/38 (89.5%)	25/39 (64.1%)		256 more per 1000 (from 51		

		risk of bias							RR 1.4 (1.08 to 1.81)	more to 519 more)	⊕OOO VERY LOW	
								64.1%		256 more per 1000 (from 51 more to 519 more)		

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] Paper omits data
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.66 Quality of life: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Quality of life: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU**		**Risk with control**	**Risk difference with Quality of life: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU** (95% CI)
Social support post-treatment (mean score at endpoint or first measurement) – ITT analysis (measured with: Interpersonal Support Evaluation List (ISEL); better indicated by lower values)											
93 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	46	47	-		The mean social support post-treatment (mean score at endpoint or first measurement) – ITT analysis in the intervention groups was **0.38 standard deviations higher** (0.03 lower to 0.79 higher)
Social support post-treatment (mean score at endpoint or first measurement or change score) – available case analysis (measured with: social Provision Scale (SPS): social support or Interpersonal Support Evaluation List (ISEL) or Multidimensional Scale for Perceived Social Support; better indicated by lower values)											

897 (3 studies) 12-52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	431	466	-		The mean social support post-treatment (mean score at endpoint or first measurement or change score) – available case analysis in the intervention groups was **0.63 standard deviations higher** (0.5 to 0.77 higher)
Life functioning post-treatment (mean score at endpoint or first measurement) – ITT analysis (measured with: Global Assessment of Functioning Scale or Social Adjustment Scale (SAS): social and leisure domain; better indicated by lower values)											
146 (2 studies) 15-44 weeks	no serious risk of bias	very serious[3]	no serious indirectness	very serious[1,2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to inconsistency, imprecision	74	72	-		The mean life functioning post-treatment (mean score at endpoint or first measurement) – ITT analysis in the intervention groups was **0.44 standard deviations lower** (2.65 lower to 1.78 higher)
Life functioning post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: social Adjustment Scale (SAS) or Global Assessment of Functioning Scale; better indicated by lower values)											
897 (2 studies) 12-52 weeks	no serious risk of bias	very serious[3]	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to inconsistency, imprecision	437	460	-		The mean life functioning post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.1 standard deviations lower** (1.92 lower to 1.72 higher)
Functional impairment post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM): Life functioning; better indicated by lower values)											
284 (1 study) 26 weeks	no serious	no serious inconsistency	no serious indirectness	serious[1]	reporting bias strongly suspected[4]	⊕⊕⊖⊖ **LOW**[1,4] due to	146	138	-		The mean functional impairment post-treatment

	risk of bias					imprecision, publication bias				(mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.4 standard deviations lower** (0.63 to 0.16 lower)
Parental stress post-treatment (mean score at endpoint or first measurement or change score) – available case analysis (measured with: parenting Stress Index (PSI); better indicated by lower values)										
212 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias strongly suspected[4]	⊕⊕⊖⊖ **LOW**[1,4] due to imprecision, publication bias	106	106	-	The mean parental stress post-treatment (mean score at endpoint or first measurement or change score) – available case analysis in the intervention groups was **0.53 standard deviations higher** (0.26 to 0.81 higher)
Wellbeing post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM): Well-being; better indicated by lower values)										
284 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias strongly suspected[4]	⊕⊕⊖⊖ **LOW**[1,4] due to imprecision, publication bias	146	138	-	The mean wellbeing post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.42 standard deviations lower** (0.65 to 0.18 lower)
Social support short follow-up (mean score at 9-16-week follow-up) – ITT analysis (measured with: Interpersonal Support Evaluation List (ISEL); better indicated by lower values)										
93 (1 study) 28 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	46	47	-	The mean social support short follow-up (mean score at 9-16-week follow-up) – ITT analysis in the

										intervention groups was **0.64 standard deviations higher** (0.22 to 1.06 higher)
Social support short follow-up (mean score at 9-16-week follow-up) – available case analysis (measured with: Interpersonal Support Evaluation List (ISEL); better indicated by lower values)										
45 (1 study) 21 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	23	22	-	The mean social support short follow-up (mean score at 9-16-week follow-up) – available case analysis in the intervention groups was **0.29 standard deviations higher** (0.3 lower to 0.88 higher)
Life functioning short follow-up (mean score at 9-16-week follow-up) – ITT analysis (measured with: Global Assessment of Functioning Scale; better indicated by lower values)										
93 (1 study) 28 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	46	47	-	The mean life functioning short follow-up (mean score at 9-16-week follow-up) – ITT analysis in the intervention groups was **0.6 standard deviations higher** (0.18 to 1.02 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] There was evidence of considerable heterogeneity between effect sizes
[4] Paper omits data

1.3.67 Quality of life: IPT versus support group

Quality assessment							Summary of findings		
		Inconsistency	Indirectness	Imprecision			Study event rates (%)		Anticipated absolute effects

Participants (studies) Follow-up	Risk of bias				Publication bias	Overall quality of evidence	With control	With Quality of life: IPT versus support group	Relative effect (95% CI)	Risk with control	Risk difference with Quality of life: IPT versus support group (95% CI)
Maternal stress post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: maternal cortisol levels; better indicated by lower values)											
44 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	22	22	-		The mean maternal stress post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.45 standard deviations lower** (1.05 lower to 0.15 higher)

[1] High risk of selection bias due to unclear allocation concealment and statistically significant baseline differences with the control group showing a higher SES score/lower income and higher depression (CES-D) mean score
[2] Total population size is less than 400 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.68 Quality of life: facilitated self-help versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With Quality of life: facilitated self-help versus TAU		Risk with control	Risk difference with Quality of life: facilitated self-help versus TAU (95% CI)
Social support post-treatment (mean score at endpoint or first measurement or change score) – available case analysis (measured with: social Provision Scale (SPS): social support; better indicated by lower values)											
59 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	28	31	-		The mean social support post-treatment (mean score at endpoint or first measurement or change score) – available case analysis in the intervention groups was **0.51 standard deviations higher** (0.01 lower to 1.03 higher)

Functional impairment post-treatment (mean score at endpoint or first measurement) – available analysis (measured with: Work and Social Adjustment Scale (WASAS): Functional impairment; better indicated by lower values)										
59 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,3] due to imprecision, publication bias	28	31	-	The mean functional impairment post-treatment (mean score at endpoint or first measurement) – available analysis in the intervention groups was **0.57 standard deviations lower** (1.1 to 0.05 lower)
Parental stress post-treatment (symptomatology at endpoint or first measurement) – ITT analysis (assessed with: parenting Stress Index (PSI) ≥260)										
143 (1 study) 20 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,4]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to imprecision, publication bias	44/72 (61.1%)	29/71 (40.8%)	**RR 0.67** (0.48 to 0.93)	**Study population**
										611 per 1000 — **202 fewer per 1000** (from 43 fewer to 318 fewer)
										Moderate
										611 per 1000 — **202 fewer per 1000** (from 43 fewer to 318 fewer)
Parental stress post-treatment (symptomatology at endpoint or first measurement) – available case analysis (assessed with: parenting Stress Index (PSI) ≥260)										
84 (1 study) 20 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[3,4] due to imprecision, publication bias	11/39 (28.2%)	3/45 (6.7%)	**RR 0.24** (0.07 to 0.79)	**Study population**
										282 per 1000 — **214 fewer per 1000** (from 59 fewer to 262 fewer)
										Moderate
										282 per 1000 — **214 fewer per 1000** (from 59 fewer to 262 fewer)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data
[4] Total number of events is less than 300 (a threshold rule-of-thumb)

1.3.69 Quality of life: listening visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With Quality of life: listening visits versus TAU		Risk with control	Risk difference with Quality of life: listening visits versus TAU (95% CI)
Functional impairment post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM): Life functioning; better indicated by lower values)											
277 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias strongly suspected[2]	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision, publication bias	146	131	-		The mean functional impairment post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.37 standard deviations lower** (0.61 to 0.14 lower)
Parental stress post-treatment (mean score at endpoint or first measurement or change score) – available case analysis (measured with: parenting Stress Index (PSI); better indicated by lower values)											
211 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias strongly suspected[2]	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision, publication bias	106	105	-		The mean parental stress post-treatment (mean score at endpoint or first measurement or change score) – available case analysis in the intervention groups was **0.45 standard deviations higher** (0.18 to 0.72 higher)
Wellbeing post-treatment (improved wellbeing at endpoint or first measurement) – available case analysis (assessed with: maternal report: Improvements in wellbeing)											
41 (1 study) 7 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3,4]	reporting bias strongly suspected[2]	⊕⊖⊖⊖ **VERY LOW**[2,3,4] due to imprecision, publication bias	12/21 (57.1%)	17/20 (85%)	**RR 1.49** (0.98 to 2.25)	**Study population**	
										571 per 1000	**280 more per 1000** (from 11 fewer to 714 more)
										Moderate	

									571 per 1000	**280 more per 1000** (from 11 fewer to 714 more)
Wellbeing post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM): Well-being; better indicated by lower values)										
277 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias strongly suspected[2]	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision, publication bias	146	131	-	The mean wellbeing post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.42 standard deviations lower** (0.66 to 0.18 lower)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] Paper omits data
[3] Total number of events is less than 300 (a threshold rule-of-thumb)
[4] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.70 Quality of life: directive counselling versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Quality of life: directive counselling versus TAU**		**Risk with control**	**Risk difference with Quality of life: directive counselling versus TAU** (95% CI)
Social support post-treatment (mean score at endpoint or first measurement or change score) – available case analysis (measured with: social Provision Scale (SPS): social support; better indicated by lower values)											
90 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	18	72	-		The mean social support post-treatment (mean score at endpoint or first measurement or change score) – available case analysis in the intervention groups was **0.53 standard deviations higher** (0.01 to 1.06 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.71 Quality of life: post-miscarriage counselling versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Quality of life: post-miscarriage counselling versus TAU**		**Risk with control**	**Risk difference with Quality of life: post-miscarriage counselling versus TAU** (95% CI)
Functional impairment post-treatment (mean score at endpoint or first measurement) – ITT analysis (measured with: short Form (36) Health Survey (SF-36): Role functioning (sum of role limitation-emotional and social functioning subscales); better indicated by lower values)											
19 (1 study) 7 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	9	10	-		The mean functional impairment post-treatment (mean score at endpoint or first measurement) – ITT analysis in the intervention groups was **0.37 standard deviations lower** (1.28 lower to 0.54 higher)
Functional impairment post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: short Form (36) Health Survey (SF-36): Role functioning (sum of role limitation-emotional and social functioning subscales); better indicated by lower values)											
15 (1 study) 7 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	7	8	-		The mean functional impairment post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.68 standard deviations lower** (1.73 lower to 0.37 higher)

[1] High risk of selection bias due to unclear allocation concealment and statistically significant baseline differences between groups in ethnicity (80% Hispanic in intervention group and 44% in TAU) and Hispanic ethnicity was associated with primary outcome with higher depression scores in Hispanic group
[2] Total population size is less than 400 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.72 Quality of life: post-traumatic birth counselling versus TAU

Quality assessment							Summary of findings		
		Inconsistency	**Indirectness**	**Imprecision**			**Study event rates (%)**		**Anticipated absolute effects**

Participants (studies) Follow-up	Risk of bias				Publication bias	Overall quality of evidence	With control	With Quality of life: post-traumatic birth counselling versus TAU	Relative effect (95% CI)	Risk with control	Risk difference with Quality of life: post-traumatic birth counselling versus TAU (95% CI)
Parental stress post-treatment (symptomatology at endpoint or first measurement) – ITT analysis (assessed with: Depression Anxiety Stress Scale (DASS): stress >19)											
103 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⨁⨁⨀⨀ **LOW**[1] due to imprecision	17/53 (32.1%)	7/50 (14%)	**RR 0.44** (0.2 to 0.96)	**Study population**	
										321 per 1000	**180 fewer per 1000** (from 13 fewer to 257 fewer)
										Moderate	
										321 per 1000	**180 fewer per 1000** (from 13 fewer to 257 fewer)
Parental stress post-treatment (symptomatology at endpoint or first measurement) – available case analysis (assessed with: Depression Anxiety Stress Scale (DASS): stress >19)											
103 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⨁⨁⨀⨀ **LOW**[1] due to imprecision	17/53 (32.1%)	7/50 (14%)	**RR 0.44** (0.2 to 0.96)	**Study population**	
										321 per 1000	**180 fewer per 1000** (from 13 fewer to 257 fewer)
										Moderate	
										321 per 1000	**180 fewer per 1000** (from 13 fewer to 257 fewer)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)

1.3.73 Quality of life: social support versus TAU

Quality assessment	Summary of findings

Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With Quality of life: social support versus TAU		Risk with control	Risk difference with Quality of life: social support versus TAU (95% CI)
Social support post-treatment (mean score at endpoint or first measurement or change score) – available case analysis (measured with: Interpersonal Support Evaluation List (ISEL) or Social Provision Scale (SPS): social support; better indicated by lower values)											
111 (2 studies) 12-14 weeks	no serious risk of bias	very serious[1]	no serious indirectness	very serious[2,3]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to inconsistency, imprecision	58	53	-		The mean social support post-treatment (mean score at endpoint or first measurement or change score) – available case analysis in the intervention groups was **0.04 standard deviations higher** (0.87 lower to 0.96 higher)
Parental stress post-treatment (mean score at endpoint or first measurement or change score) – available case analysis (measured with: perceived Stress Scale or Child-Care Stress Checklist; better indicated by lower values)											
101 (2 studies) 8-14 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⨁⨁⊖⊖ **LOW**[2] due to imprecision	51	50	-		The mean parental stress post-treatment (mean score at endpoint or first measurement or change score) – available case analysis in the intervention groups was **0.43 standard deviations lower** (0.83 to 0.04 lower)
Maternal cortisol levels post-treatment (mean score at endpoint or first measurement) – available case analysis (Better indicated by lower values)											
30 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⨁⊖⊖ **LOW**[2,3] due to imprecision	16	14	-		The mean maternal cortisol levels post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.23 standard deviations higher** (0.49 lower to 0.95 higher)
Self-esteem post-treatment (mean score at endpoint or first measurement or change score) – available case analysis (measured with: Coopersmith's Self-Esteem Inventory (SEI) or Rosenberg Self-Esteem Scale (SES); better indicated by lower values)											

101 (2 studies) 8-14 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊕⊖⊖ **LOW**[2,3] due to imprecision	51	50	-	The mean self-esteem post-treatment (mean score at endpoint or first measurement or change score) – available case analysis in the intervention groups was **0.14 standard deviations higher** (0.25 lower to 0.53 higher)
Loneliness post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: UCLA Loneliness Scale (LS); better indicated by lower values)										
653 (2 studies) 8-12 weeks	no serious risk of bias	serious[4]	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3,4] due to inconsistency, imprecision	336	317	-	The mean loneliness post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.26 standard deviations lower** (0.74 lower to 0.22 higher)
Loneliness short follow-up (mean score at 9-16-week follow-up) – available case analysis (measured with: UCLA Loneliness Scale (LS); better indicated by lower values)										
600 (1 study) 24 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	311	289	-	The mean loneliness short follow-up (mean score at 9-16-week follow-up) – available case analysis in the intervention groups was **0.11 standard deviations lower** (0.27 lower to 0.05 higher)

[1] There was evidence of considerable heterogeneity between effect sizes
[2] Total population size is less than 400 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] There was evidence of moderate heterogeneity between effect sizes

1.3.74 Quality of life: psychologically (CBT/IPT)-informed psychoeducation versus TAU/enhanced TAU

Quality assessment							Summary of findings		
		Inconsistency	**Indirectness**	**Imprecision**			**Study event rates (%)**		**Anticipated absolute effects**

Participants (studies) Follow-up	Risk of bias				Publication bias	Overall quality of evidence	With control	With Quality of life: psychologically (CBT/IPT)-informed psychoeducation versus TAU/enhanced TAU	Relative effect (95% CI)	Risk with control	Risk difference with Quality of life: psychologically (CBT/IPT)-informed psychoeducation versus TAU/enhanced TAU (95% CI)
Social support post-treatment (mean score at endpoint or first measurement) – ITT analysis (measured with: perceived Social Support Scale (PSSS); better indicated by lower values)											
194 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	98	96	-		The mean social support post-treatment (mean score at endpoint or first measurement) – ITT analysis in the intervention groups was **0.74 standard deviations higher** (0.45 to 1.03 higher)
Functional impairment post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: social Adjustment Scale (SAS) or Longitudinal Interval Follow-up Examination: Range of Impaired Functioning Tool (LIFE-RIFT); better indicated by lower values)											
128 (2 studies) 13 weeks	serious[2]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	63	65	-		The mean functional impairment post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.46 standard deviations lower** (0.81 to 0.1 lower)
Parental stress post-treatment (mean score at endpoint or first measurement) – ITT analysis (measured with: perceived Stress Scale ; better indicated by lower values)											
156 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,3]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to imprecision	78	78	-		The mean parental stress post-treatment (mean score at endpoint or first measurement) – ITT analysis in the intervention groups was **0.18 standard deviations**

										lower (0.5 lower to 0.13 higher)
Parental stress post-treatment (mean score at endpoint or first measurement or change score) – available case analysis (measured with: Visual Analogue Scale (VAS): maternal stress or Perceived Stress Scale; better indicated by lower values)										
95 (2 studies) 13-49 weeks	serious[4]	very serious[5]	no serious indirectness	very serious[1,3]	reporting bias strongly suspected[6]	⊕⊖⊖⊖ **VERY LOW**[1,3,4,5,6] due to risk of bias, inconsistency, imprecision, publication bias	52	43	-		The mean parental stress post-treatment (mean score at endpoint or first measurement or change score) – available case analysis in the intervention groups was **0.13 standard deviations lower** (1.33 lower to 1.07 higher)
Maternal cortisol levels post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: Average (morning/evening) cortisol (log scores); better indicated by lower values)										
53 (1 study) 49 weeks	serious[4]	no serious inconsistency	no serious indirectness	very serious[1,3]	reporting bias strongly suspected[6]	⊕⊖⊖⊖ **VERY LOW**[1,3,4,6] due to risk of bias, imprecision, publication bias	29	24	-		The mean maternal cortisol levels post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.37 standard deviations higher** (0.17 lower to 0.92 higher)
Happiness post-treatment (mean score at endpoint or first measurement) – ITT analysis (measured with: subjective Happiness Scale; better indicated by lower values)										
156 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	78	78	-		The mean happiness post-treatment (mean score at endpoint or first measurement) – ITT analysis in the intervention groups was **0.05 standard deviations higher** (0.27 lower to 0.36 higher)

Social support short follow-up (mean score at 9-16-week follow-up) – ITT analysis (measured with: perceived Social Support Scale (PSSS); better indicated by lower values)											
194 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	98	96	-		The mean social support short follow-up (mean score at 9-16-week follow-up) – ITT analysis in the intervention groups was **0.33 standard deviations higher** (0.05 to 0.62 higher)
Functional impairment Intermediate follow-up (mean score at 17-24-week follow-up) – available case analysis (measured with: social Adjustment Scale (SAS); better indicated by lower values)											
42 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,3]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to imprecision	21	21	-		The mean functional impairment intermediate follow-up (mean score at 17-24-week follow-up) – available case analysis in the intervention groups was **0.43 standard deviations lower** (1.05 lower to 0.18 higher)
Parental stress Intermediate follow-up (mean score at 17-24-week follow-up) – ITT analysis (measured with: perceived Stress Scale ; better indicated by lower values)											
156 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	78	78	-		The mean parental stress intermediate follow-up (mean score at 17-24-week follow-up) – ITT analysis in the intervention groups was **0.09 standard deviations lower** (0.4 lower to 0.23 higher)
Parental stress Intermediate follow-up (mean score at 17-24-week follow-up) – available case analysis (measured with: perceived Stress Scale; better indicated by lower values)											
42 (1 study) 26 weeks	no serious	no serious inconsistency	no serious indirectness	very serious[1,3]	undetected	⊕⊕⊖⊖ **LOW**[1,3]	21	21	-		The mean parental stress intermediate follow-up (mean score at 17-24-week

	risk of bias					due to imprecision					follow-up) – available case analysis in the intervention groups was **0.16 standard deviations lower** (0.77 lower to 0.45 higher)
Happiness Intermediate follow-up (mean score at 17-24-week follow-up) – ITT analysis (measured with: subjective Happiness Scale; better indicated by lower values)											
156 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,3]	undetected	⨁⨁⨀⨀ **LOW**[1,3] due to imprecision	78	78	-		The mean happiness intermediate follow-up (mean score at 17-24-week follow-up) – ITT analysis in the intervention groups was **0.18 standard deviations higher** (0.13 lower to 0.5 higher)
Parental stress long follow-up (mean score at >24-week follow-up) – available case analysis (measured with: Visual Analogue Scale (VAS): maternal stress; better indicated by lower values)											
46 (1 study) 101 weeks	serious[4]	no serious inconsistency	no serious indirectness	very serious[1,3]	reporting bias strongly suspected[6]	⨁⨀⨀⨀ **VERY LOW**[1,3,4,6] due to risk of bias, imprecision, publication bias	22	24	-		The mean parental stress long follow-up (mean score at >24-week follow-up) – available case analysis in the intervention groups was **0.12 standard deviations higher** (0.46 lower to 0.7 higher)
Maternal cortisol levels long follow-up (mean score at >24-week follow-up) – available case analysis (measured with: Average (morning/evening) cortisol (log scores); better indicated by lower values)											
46 (1 study) 101 weeks	serious[4]	no serious inconsistency	no serious indirectness	very serious[1,3]	reporting bias strongly suspected[6]	⨁⨀⨀⨀ **VERY LOW**[1,3,4,6] due to risk of bias, imprecision, publication bias	22	24	-		The mean maternal cortisol levels long follow-up (mean score at >24-week follow-up) – available case analysis in the intervention groups was **0.52 standard deviations**

										lower (1.11 lower to 0.07 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] Unclear risk of selection bias as insufficient detail reported with regards to randomisation method and allocation concealment and unclear risk of detection bias as blinding of outcome assessment is not reported
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] High risk of selection bias due to statistically significant baseline/mid-treatment difference in average maternal salivary cortisol levels (0.62 in intervention group and 0.75 in control group)
[5] There was evidence of considerable heterogeneity between effect sizes
[6] Papers omit data

1.3.75 Quality of life: home visits versus TAU/enhanced TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Quality of life: home visits versus TAU/enhanced TAU**		**Risk with control**	**Risk difference with Quality of life: home visits versus TAU/enhanced TAU** (95% CI)
Parental stress post-treatment (symptomatology at endpoint or first measurement) – ITT analysis (assessed with: parenting Stress Index (PSI): severe parenting stress (as defined by Abidin))											
364 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	72/185 (38.9%)	61/179 (34.1%)	**RR 0.88** (0.67 to 1.15)	**Study population**	
										389 per 1000	**47 fewer per 1000** (from 128 fewer to 58 more)
										Moderate	
										389 per 1000	**47 fewer per 1000** (from 128 fewer to 58 more)
Parental stress post-treatment (symptomatology at endpoint or first measurement) – available case analysis (assessed with: parenting Stress Index (PSI): severe parenting stress (as defined by Abidin))											
	serious[1]				undetected					**Study population**	

249 (1 study) 104 weeks		no serious inconsistency	no serious indirectness	very serious[2,3]		⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	10/123 (8.1%)	8/126 (6.3%)	**RR 0.78** (0.32 to 1.91)	**81 per 1000**	**18 fewer per 1000** (from 55 fewer to 74 more)
										Moderate	
										81 per 1000	**18 fewer per 1000** (from 55 fewer to 74 more)
Parental stress post-treatment (mean score at endpoint or first measurement or change score) – available case analysis (measured with: parenting Stress Index (PSI) or Perceived Stress Scale; better indicated by lower values)											
595 (2 studies) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[4]	⊕⊕⊕⊖ **MODERATE**[4] due to publication bias	299	296	-		The mean parental stress post-treatment (mean score at endpoint or first measurement or change score) – available case analysis in the intervention groups was **0.06 standard deviations lower** (0.29 lower to 0.18 higher)

[1] High risk of selection bias due to unclear allocation concealment and statistically significant baseline differences in poor psychological resources (37% intervention group versus 50% control) and in prenatal enrolment (41% intervention group and 53% control)
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] Paper omits data

1.3.76 Quality of life: mother–infant relationship interventions versus TAU/enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Quality of life: mother–infant relationship interventions versus TAU/enhanced TAU**		**Risk with control**	**Risk difference with Quality of life: mother–infant relationship interventions versus TAU/enhanced TAU** (95% CI)
Parental stress post-treatment (symptomatology at endpoint or first measurement) – ITT analysis (assessed with: parenting Stress Index (PSI): Treatment non-response (no improvement-reliable change index))											

80 (1 study) 26 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to risk of bias, imprecision	33/40 (82.5%)	27/40 (67.5%)	**RR 0.82** (0.63 to 1.06)	**Study population**	
										825 per 1000	**149 fewer per 1000** (from 305 fewer to 49 more)
										Moderate	
										825 per 1000	**149 fewer per 1000** (from 305 fewer to 49 more)
Parental stress post-treatment (symptomatology at endpoint or first measurement) – available case analysis (assessed with: parenting Stress Index (PSI): Treatment non-response (no improvement-reliable change index))											
75 (1 study) 26 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to risk of bias, imprecision	30/37 (81.1%)	25/38 (65.8%)	**RR 0.81** (0.62 to 1.07)	**Study population**	
										811 per 1000	**154 fewer per 1000** (from 308 fewer to 57 more)
										Moderate	
										811 per 1000	**154 fewer per 1000** (from 308 fewer to 57 more)
Parental stress post-treatment (mean score at endpoint or first measurement or change score) – available case analysis (measured with: parenting Stress Index (PSI) or Parental Stress Scale-Neonatal Intensive Care (PSS-NICU): parental role restriction; better indicated by lower values)											
173 (2 studies) 4-26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⨁⨁⨀⨀ **LOW**[4] due to imprecision	87	86	-		The mean parental stress post-treatment (mean score at endpoint or first measurement or change score) – available case analysis in the intervention groups was **0.06 standard deviations lower** (0.36 lower to 0.24 higher)

[1] High risk of selection bias due to a statistically significant baseline difference in the age of infants (4.4 months old in intervention group versus 5.9 months old in TAU group)
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.77 Quality of life: psychosomatic intervention versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With Quality of life: psychosomatic intervention versus TAU		Risk with control	Risk difference with Quality of life: psychosomatic intervention versus TAU (95% CI)
Poor social support mean scores post-treatment – available case analysis (measured with: Functional Social Support Questionnaire (FSSQ): Lack of social support; better indicated by lower values)											
127 (1 study) 34 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	58	69	-		The mean poor social support mean scores post-treatment – available case analysis in the intervention groups was **0.18 standard deviations lower** (0.53 lower to 0.17 higher)
Parental stress mean scores post-treatment – available case analysis (measured with: stress Events Scale (Holmes & Rahe, 1967): stress score value; better indicated by lower values)											
127 (1 study) 34 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	58	69	-		The mean parental stress mean scores post-treatment – available case analysis in the intervention groups was **0.11 standard deviations lower** (0.46 lower to 0.24 higher)

[1] Risk of attrition bias due to statistically significant higher drop-out in the control group
[2] Total population size is less than 400 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.78 Quality of life: mindfulness training versus TAU/enhanced TAU

Quality assessment							Summary of findings		
		Inconsistency	Indirectness	Imprecision			Study event rates (%)		Anticipated absolute effects

Participants (studies) Follow-up	Risk of bias				Publication bias	Overall quality of evidence	With control	With Quality of life: mindfulness training versus TAU/enhanced TAU	Relative effect (95% CI)	Risk with control	Risk difference with Quality of life: mindfulness training versus TAU/enhanced TAU (95% CI)
Parental stress post-treatment (mean score at endpoint or first measurement) – ITT analysis (measured with: perceived Stress Scale (PSS); better indicated by lower values)											
47 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	23	24	-		The mean parental stress post-treatment (mean score at endpoint or first measurement) – ITT analysis in the intervention groups was **0.22 standard deviations higher** (0.36 lower to 0.79 higher)
Parental stress post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: perceived Stress Scale (PSS); better indicated by lower values)											
31 (1 study) 10 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to imprecision, publication bias	18	13	-		The mean parental stress post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.19 standard deviations lower** (0.91 lower to 0.52 higher)
Positive affect post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: positive and Negative Affect Schedule-Extended (PANAS-X): positive affect; better indicated by lower values)											
31 (1 study) 10 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to imprecision, publication bias	18	13	-		The mean positive affect post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.44 standard deviations higher** (0.28 lower to 1.16 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data

1.3.79 Service utilisation: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With service utilisation: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU**		**Risk with control**	**Risk difference with service utilisation: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU** (95% CI)
Use of NHS health visitor post-treatment (service utilisation at endpoint or first measurement) – ITT analysis (assessed with: mACH nurse advice)											
57 (1 study) 21 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	15/28 (53.6%)	16/29 (55.2%)	**RR 1.03** (0.64 to 1.66)	**Study population**	
										536 per 1000	**16 more per 1000** (from 193 fewer to 354 more)
										Moderate	
										536 per 1000	**16 more per 1000** (from 193 fewer to 354 more)
Use of NHS health visitor post-treatment (service utilisation at endpoint or first measurement) – available case analysis (assessed with: mACH nurse advice)											
46 (1 study) 21 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	10/23 (43.5%)	10/23 (43.5%)	**RR 1** (0.52 to 1.93)	**Study population**	
										435 per 1000	**0 fewer per 1000** (from 209 fewer to 404 more)
										Moderate	

									435 per 1000	**0 fewer per 1000** (from 209 fewer to 405 more)
Antidepressant medication post-treatment (medication use at endpoint or first measurement) – ITT analysis (assessed with: antidepressant use)										
57 (1 study) 21 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	18/28 (64.3%)	18/29 (62.1%)	**RR 0.97** (0.65 to 1.44)	**Study population**
									643 per 1000	**19 fewer per 1000** (from 225 fewer to 283 more)
									Moderate	
									643 per 1000	**19 fewer per 1000** (from 225 fewer to 283 more)
Antidepressant medication post-treatment (medication use at endpoint or first measurement) – available case analysis (assessed with: antidepressant use)										
46 (1 study) 21 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	13/23 (56.5%)	12/23 (52.2%)	**RR 0.92** (0.54 to 1.57)	**Study population**
									565 per 1000	**45 fewer per 1000** (from 260 fewer to 322 more)
									Moderate	
									565 per 1000	**45 fewer per 1000** (from 260 fewer to 322 more)
Psychotherapy post-treatment (service utilisation at endpoint or first measurement) – ITT analysis										
57 (1 study) 21 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	13/28 (46.4%)	8/29 (27.6%)	**RR 0.59** (0.29 to 1.21)	**Study population**
									464 per 1000	**190 fewer per 1000** (from 330 fewer to 98 more)
									Moderate	
									464 per 1000	**190 fewer per 1000** (from 329 fewer to 97 more)
Psychotherapy post-treatment (service utilisation at endpoint or first measurement) – available case analysis										

46 (1 study) 21 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	8/23 (34.8%)	2/23 (8.7%)	**RR 0.25** (0.06 to 1.05)	**Study population**	
										348 per 1000	**261 fewer per 1000** (from 327 fewer to 17 more)
										Moderate	
										348 per 1000	**261 fewer per 1000** (from 327 fewer to 17 more)
Counselling post-treatment (service utilisation at endpoint or first measurement) – ITT analysis											
57 (1 study) 21 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	17/28 (60.7%)	11/29 (37.9%)	**RR 0.62** (0.36 to 1.09)	**Study population**	
										607 per 1000	**231 fewer per 1000** (from 389 fewer to 55 more)
										Moderate	
										607 per 1000	**231 fewer per 1000** (from 388 fewer to 55 more)
Counselling post-treatment (service utilisation at endpoint or first measurement) – available case analysis											
46 (1 study) 21 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	12/23 (52.2%)	5/23 (21.7%)	**RR 0.42** (0.17 to 0.99)	**Study population**	
										522 per 1000	**303 fewer per 1000** (from 5 fewer to 433 fewer)
										Moderate	
										522 per 1000	**303 fewer per 1000** (from 5 fewer to 433 fewer)
Self help support group post-treatment (service utilisation at endpoint or first measurement) – ITT analysis											
57 (1 study) 21 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	11/28 (39.3%)	11/29 (37.9%)	**RR 0.97** (0.5 to 1.86)	**Study population**	
										393 per 1000	**12 fewer per 1000** (from 196 fewer to 338 more)

									Moderate	
									393 per 1000	**12 fewer per 1000** (from 197 fewer to 338 more)
Self help support group post-treatment (service utilisation at endpoint or first measurement) – available case analysis										
46 (1 study) 21 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	6/23 (26.1%)	5/23 (21.7%)	**RR 0.83** (0.3 to 2.35)	**Study population**
									261 per 1000	**44 fewer per 1000** (from 183 fewer to 352 more)
									Moderate	
									261 per 1000	**44 fewer per 1000** (from 183 fewer to 352 more)
Alternative therapies post-treatment (service utilisation at endpoint or first measurement) – ITT analysis										
57 (1 study) 21 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	8/28 (28.6%)	11/29 (37.9%)	**RR 1.33** (0.63 to 2.81)	**Study population**
									286 per 1000	**94 more per 1000** (from 106 fewer to 517 more)
									Moderate	
									286 per 1000	**94 more per 1000** (from 106 fewer to 518 more)
Alternative therapies post-treatment (service utilisation at endpoint or first measurement) – available case analysis										
46 (1 study) 21 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	3/23 (13%)	5/23 (21.7%)	**RR 1.67** (0.45 to 6.17)	**Study population**
									130 per 1000	**87 more per 1000** (from 72 fewer to 674 more)
									Moderate	
									130 per 1000	**87 more per 1000** (from 71 fewer to 672 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.80 Service utilisation: facilitated self-help versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With service utilisation: facilitated self-help versus TAU**		**Risk with control**	**Risk difference with service utilisation: facilitated self-help versus TAU** (95% CI)
Use of childbirth hospital post-treatment (service utilisation at endpoint) – ITT analysis (assessed with: Adult Service Use Schedule (AD-SUS): childbirth hospital)											
83 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	17/42 (40.5%)	12/41 (29.3%)	**RR 0.72** (0.4 to 1.32)	**Study population**	
										405 per 1000	**113 fewer per 1000** (from 243 fewer to 130 more)
										Moderate	
										405 per 1000	**113 fewer per 1000** (from 243 fewer to 130 more)
Use of childbirth hospital post-treatment (service utilisation at endpoint) – available case analysis (assessed with: Adult Service Use Schedule (AD-SUS): childbirth hospital)											
57 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	2/27 (7.4%)	1/30 (3.3%)	**RR 0.45** (0.04 to 4.69)	**Study population**	
										74 per 1000	**41 fewer per 1000** (from 71 fewer to 273 more)
										Moderate	
										74 per 1000	**41 fewer per 1000** (from 71 fewer to 273 more)
Use of childbirth hospital post-treatment (service utilisation at endpoint) – available case analysis (measured with: Adult Service Use Schedule (AD-SUS): childbirth hospital; better indicated by lower values)											

57 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,4]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[2,3,4] due to imprecision, publication bias	27	30	**-**		The mean use of childbirth hospital post-treatment (service utilisation at endpoint) – available case analysis in the intervention groups was **0.24 standard deviations lower** (0.77 lower to 0.28 higher)
Use of maternal general health hospital post-treatment (service utilisation at endpoint) – ITT analysis (assessed with: Adult Service Use Schedule (AD-SUS): maternal general health hospital)											
83 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	15/42 (35.7%)	11/41 (26.8%)	**RR 0.75** (0.39 to 1.44)	**Study population**	
										357 per 1000	**89 fewer per 1000** (from 218 fewer to 157 more)
										Moderate	
										357 per 1000	**89 fewer per 1000** (from 218 fewer to 157 more)
Use of maternal general health hospital post-treatment (service utilisation at endpoint) – available case analysis (assessed with: Adult Service Use Schedule (AD-SUS): maternal general health hospital)											
57 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness		reporting bias strongly suspected[3]	See comment	0/27 (0%)	0/30 (0%)	**not pooled**	See comment	See comment
Use of maternal general health hospital post-treatment (service utilisation at endpoint) – available case analysis (measured with: Adult Service Use Schedule (AD-SUS): maternal general health hospital; better indicated by lower values)											
57 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness		reporting bias strongly suspected[3]	See comment	27	30	**-**	See comment	See comment
Use of mental health hospital post-treatment (service utilisation at endpoint) – ITT analysis (assessed with: Adult Service Use Schedule (AD-SUS): mental health hospital)											
										Study population	

83 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	16/42 (38.1%)	11/41 (26.8%)	**RR 0.7** (0.37 to 1.33)	**381 per 1000** **Moderate** **381 per 1000**	**114 fewer per 1000** (from 240 fewer to 126 more) **114 fewer per 1000** (from 240 fewer to 126 more)
Use of mental health hospital post-treatment (service utilisation at endpoint) – available case analysis (assessed with: Adult Service Use Schedule (AD-SUS): mental health hospital)											
57 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	1/27 (3.7%)	0/30 (0%)	**RR 0.3** (0.01 to 7.09)	**Study population** **37 per 1000** **Moderate** **37 per 1000**	 **26 fewer per 1000** (from 37 fewer to 226 more) **26 fewer per 1000** (from 37 fewer to 225 more)
Use of mental health hospital post-treatment (service utilisation at endpoint) – available case analysis (measured with: Adult Service Use Schedule (AD-SUS): mental health hospital; better indicated by lower values)											
57 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness		reporting bias strongly suspected[3]	See comment	27	30	-	See comment	See comment
Use of mental health outpatient post-treatment (service utilisation at endpoint) – ITT analysis (assessed with: Adult Service Use Schedule (AD-SUS): mental health out-patient)											
83 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	26/42 (61.9%)	25/41 (61%)	**RR 0.98** (0.7 to 1.39)	**Study population** **619 per 1000** **Moderate** **619 per 1000**	 **12 fewer per 1000** (from 186 fewer to 241 more) **12 fewer per 1000** (from 186 fewer to 241 more)

Use of mental health outpatient post-treatment (service utilisation at endpoint) – available case analysis (assessed with: Adult Service Use Schedule (AD-SUS): mental health out-patient)											
57 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	11/27 (40.7%)	14/30 (46.7%)	**RR 1.15** (0.63 to 2.08)	**Study population**	
										407 per 1000	**61 more per 1000** (from 151 fewer to 440 more)
										Moderate	
										407 per 1000	**61 more per 1000** (from 151 fewer to 440 more)
Use of mental health outpatient post-treatment (service utilisation at endpoint) – available case analysis (measured with: Adult Service Use Schedule (AD-SUS): mental health out-patient; better indicated by lower values)											
57 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,4]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[2,3,4] due to imprecision, publication bias	27	30	-		The mean use of mental health outpatient post-treatment (service utilisation at endpoint) – available case analysis in the intervention groups was **0.47 standard deviations lower** (1 lower to 0.06 higher)
Use of health community service post-treatment (service utilisation at endpoint) – ITT analysis (assessed with: Adult Service Use Schedule (AD-SUS): health community service)											
83 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,3] due to imprecision, publication bias	40/42 (95.2%)	39/41 (95.1%)	**RR 1** (0.91 to 1.1)	**Study population**	
										952 per 1000	**0 fewer per 1000** (from 86 fewer to 95 more)
										Moderate	
										952 per 1000	**0 fewer per 1000** (from 86 fewer to 95 more)
Use of health community service post-treatment (service utilisation at endpoint) – available case analysis (assessed with: Adult Service Use Schedule (AD-SUS): health community service)											
										Study population	

57 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,3] due to imprecision, publication bias	25/27 (92.6%)	28/30 (93.3%)	**RR 1.01** (0.87 to 1.16)	**926 per 1000** **Moderate** **926 per 1000**	**9 more per 1000** (from 120 fewer to 148 more) **9 more per 1000** (from 120 fewer to 148 more)
Use of health community service post-treatment (service utilisation at endpoint) – available case analysis (measured with: Adult Service Use Schedule (AD-SUS): health community service; better indicated by lower values)											
57 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,4]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[2,3,4] due to imprecision, publication bias	27	30	-		The mean use of health community service post-treatment (service utilisation at endpoint) – available case analysis in the intervention groups was **0.1 standard deviations higher** (0.42 lower to 0.62 higher)
Antidepressant medication post-treatment (medication use at endpoint or first measurement) – ITT analysis (assessed with: Adult Service Use Schedule (AD-SUS): antidepressant medication)											
83 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	31/42 (73.8%)	33/41 (80.5%)	**RR 1.09** (0.86 to 1.38)	**Study population** **738 per 1000** **Moderate** **738 per 1000**	 **66 more per 1000** (from 103 fewer to 280 more) **66 more per 1000** (from 103 fewer to 280 more)
Antidepressant medication post-treatment (medication use at endpoint or first measurement) – available case analysis (assessed with: Adult Service Use Schedule (AD-SUS): antidepressant medication)											
57 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	19/30 (63.3%)	19/27 (70.4%)	**RR 1.11** (0.77 to 1.6)	**Study population** **633 per 1000** **Moderate**	 **70 more per 1000** (from 146 fewer to 380 more)

									633 per 1000	**70 more per 1000** (from 146 fewer to 380 more)
Antidepressant medication post-treatment (medication use at endpoint) – available case analysis (measured with: Adult Service Use Schedule (AD-SUS): antidepressant medication; better indicated by lower values)										
57 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,4]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[2,3,4] due to imprecision, publication bias	30	27	-	The mean antidepressant medication post-treatment (medication use at endpoint) – available case analysis in the intervention groups was **0.14 standard deviations lower** (0.66 lower to 0.38 higher)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data
[4] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.81 Service utilisation: listening visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With service utilisation: listening visits versus TAU**		**Risk with control**	**Risk difference with service utilisation: listening visits versus TAU** (95% CI)
Use of maternal general health hospital post-treatment (service utilisation at endpoint) – ITT analysis (assessed with: health service use – use of hospital doctor in last month)											
731 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	120/548 (21.9%)	38/183 (20.8%)	**RR 0.95** (0.69 to 1.31)	**Study population**	
										219 per 1000	**11 fewer per 1000** (from 68 fewer to 68 more)
										Moderate	
										219 per 1000	**11 fewer per 1000** (from 68 fewer to 68 more)
Use of maternal general health hospital post-treatment (service utilisation at endpoint) – available case analysis (assessed with: health service use – use of hospital doctor in last month)											
657 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	64/492 (13%)	20/165 (12.1%)	**RR 0.93** (0.58 to 1.49)	**Study population**	
										130 per 1000	**9 fewer per 1000** (from 55 fewer to 64 more)
										Moderate	
										130 per 1000	**9 fewer per 1000** (from 55 fewer to 64 more)
Use of NHS health visitor post-treatment (service utilisation at endpoint or first measurement) – ITT analysis (assessed with: health service use – maternal use of NHS health visitor in last month)											

731 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	72/548 (13.1%)	31/183 (16.9%)	**RR 1.29** (0.88 to 1.9)	**Study population**	
										131 per 1000	**38 more per 1000** (from 16 fewer to 118 more)
										Moderate	
										131 per 1000	**38 more per 1000** (from 16 fewer to 118 more)
Use of NHS health visitor post-treatment (service utilisation at endpoint or first measurement) – available case analysis (assessed with: health service use – maternal use of NHS health visitor in last month)											
657 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,3] due to imprecision, publication bias	16/492 (3.3%)	13/165 (7.9%)	**RR 2.42** (1.19 to 4.93)	**Study population**	
										33 per 1000	**46 more per 1000** (from 6 more to 128 more)
										Moderate	
										33 per 1000	**47 more per 1000** (from 6 more to 130 more)
Health visitor telephone contact post-treatment (service utilisation [in last month] at endpoint) – ITT analysis (assessed with: health service use – health visitor telephone contact in last month)											
731 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	60/548 (10.9%)	29/183 (15.8%)	**RR 1.45** (0.96 to 2.18)	**Study population**	
										109 per 1000	**49 more per 1000** (from 4 fewer to 129 more)
										Moderate	
										110 per 1000	**50 more per 1000** (from 4 fewer to 130 more)

Health visitor telephone contact post-treatment (service utilisation [in last month] at endpoint) – available case analysis (assessed with: health service use – health visitor telephone contact in last month)											
657 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,3] due to imprecision, publication bias	4/492 (0.8%)	11/165 (6.7%)	**RR 8.2** (2.65 to 25.4)	**Study population**	
										8 per 1000	**59 more per 1000** (from 13 more to 198 more)
										Moderate	
										8 per 1000	**58 more per 1000** (from 13 more to 195 more)
Maternal use of midwife post-treatment (service utilisation [in last month] at endpoint) – ITT analysis (assessed with: health service use – maternal use of midwife in last month)											
731 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	135/548 (24.6%)	44/183 (24%)	**RR 0.98** (0.73 to 1.31)	**Study population**	
										246 per 1000	**5 fewer per 1000** (from 67 fewer to 76 more)
										Moderate	
										246 per 1000	**5 fewer per 1000** (from 66 fewer to 76 more)
Maternal use of midwife post-treatment (service utilisation [in last month] at endpoint) – available case analysis (assessed with: health service use – maternal use of midwife in last month)											
601 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	43/456 (9.4%)	6/145 (4.1%)	**RR 0.44** (0.19 to 1.01)	**Study population**	
										94 per 1000	**53 fewer per 1000** (from 76 fewer to 1 more)
										Moderate	

										94 per 1000	**53 fewer per 1000** (from 76 fewer to 1 more)
Use of GP post-treatment (service utilisation [in last month] at endpoint) – ITT analysis (assessed with: health service use – use of GP in last month)											
731 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[3]	⊕⊕⊕⊖ **MODERATE**[3] due to publication bias	275/548 (50.2%)	89/183 (48.6%)	**RR 0.97** (0.82 to 1.15)	**Study population**	
										502 per 1000	**15 fewer per 1000** (from 90 fewer to 75 more)
										Moderate	
										502 per 1000	**15 fewer per 1000** (from 90 fewer to 75 more)
Use of GP post-treatment (service utilisation [in last month] at endpoint) – available case analysis (assessed with: health service use – use of GP in last month)											
657 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias strongly suspected[3]	⊕⊕⊖⊖ **LOW**[1,3] due to imprecision, publication bias	219/492 (44.5%)	71/165 (43%)	**RR 0.97** (0.79 to 1.18)	**Study population**	
										445 per 1000	**13 fewer per 1000** (from 93 fewer to 80 more)
										Moderate	
										445 per 1000	**13 fewer per 1000** (from 93 fewer to 80 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data

1.3.82 Service utilisation: social support versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With service utilisation: social support versus TAU		Risk with control	Risk difference with service utilisation: social support versus TAU (95% CI)
Health service use post-treatment (service utilisation at endpoint) – available case analysis (measured with: health service utilisation and cost of care questionnaire: health service use; better indicated by lower values)											
612 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	315	297	-		The mean health service use post-treatment (service utilisation at endpoint) – available case analysis in the intervention groups was **0.08 standard deviations higher** (0.08 lower to 0.23 higher)
Antidepressant medication post-treatment (medication use at endpoint or first measurement) – ITT analysis (assessed with: health service utilisation and cost of care questionnaire: current antidepressant use)											
701 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	56/352 (15.9%)	63/349 (18.1%)	**RR 1.13** (0.82 to 1.58)	**Study population**	
										159 per 1000	**21 more per 1000** (from 29 fewer to 92 more)
										Moderate	
										159 per 1000	**21 more per 1000** (from 29 fewer to 92 more)
Antidepressant medication post-treatment (medication use at endpoint or first measurement) – available case analysis (assessed with: health service utilisation and cost of care questionnaire: current antidepressant use)											
612 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	19/315 (6%)	11/297 (3.7%)	**RR 0.61** (0.3 to 1.27)	**Study population**	
										60 per 1000	**24 fewer per 1000** (from 42 fewer to 16 more)

									Moderate	
									60 per 1000	**23 fewer per 1000** (from 42 fewer to 16 more)
Health service use short follow-up (service utilisation at 9-16-week follow-up) – available case analysis (measured with: health service utilisation and cost of care questionnaire: health service use; better indicated by lower values)										
600 (1 study) 24 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	311	289	-	The mean health service use short follow-up (service utilisation at 9-16-week follow-up) – available case analysis in the intervention groups was **0.02 standard deviations lower** (0.18 lower to 0.14 higher)
Antidepressant medication short follow-up (medication use at 9-16-week follow-up) – ITT analysis (assessed with: health service utilisation and cost of care questionnaire: current antidepressant use)										
701 (1 study) 24 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	70/352 (19.9%)	76/349 (21.8%)	**RR 1.1** (0.82 to 1.46)	**Study population**
										199 per 1000 **20 more per 1000** (from 36 fewer to 91 more)
										Moderate
										199 per 1000 **20 more per 1000** (from 36 fewer to 92 more)
Antidepressant medication short follow-up (medication use at 9-16-week follow-up) – available case analysis (assessed with: health service utilisation and cost of care questionnaire: current antidepressant use)										
600 (1 study) 24 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	29/311 (9.3%)	16/289 (5.5%)	**RR 0.59** (0.33 to 1.07)	**Study population**
										93 per 1000 **38 fewer per 1000** (from 62 fewer to 7 more)
										Moderate
										93 per 1000 **38 fewer per 1000** (from 62 fewer to 7 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.83 Experience of care: mother–infant relationship interventions versus TAU/enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With experience of care: mother–infant relationship interventions versus TAU/enhanced TAU		Risk with control	Risk difference with experience of care: mother–infant relationship interventions versus TAU/enhanced TAU (95% CI)
Satisfaction with intervention post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: maternal report; better indicated by lower values)											
98 (1 study) 7 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	50	48	-		The mean satisfaction with intervention post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.25 standard deviations higher** (0.14 lower to 0.65 higher)
Satisfaction with therapeutic alliance (empathetic) post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: Visual Analogue Scale (VAS): Therapeutic alliance (mother felt understood); better indicated by lower values)											
98 (1 study) 7 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⨁⨁⨀⨀ **LOW**[1] due to imprecision	50	48	-		The mean satisfaction with therapeutic alliance (empathetic) post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0 standard deviations higher** (0.4 lower to 0.4 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.84 Attrition: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With attrition: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU		Risk with control	Risk difference with attrition: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
1983 (12 studies) 6-26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⨁⨁⨁⨀ **MODERATE**[1] due to imprecision	148/951 (15.6%)	195/1032 (18.9%)	**RR 1.14** (0.83 to 1.55)	**Study population**	
										156 per 1000	**22 more per 1000** (from 26 fewer to 86 more)
										Moderate	
										155 per 1000	**22 more per 1000** (from 26 fewer to 85 more)

[1] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.85 Attrition: CBT versus relational constructivist therapy

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With attrition: CBT versus relational constructivist therapy		Risk with control	Risk difference with attrition: CBT versus relational constructivist therapy (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
60 (1 study)	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	2/28 (7.1%)	2/32 (6.3%)	**RR 0.88** (0.13 to 5.81)	**Study population**	
										71 per 1000	**9 fewer per 1000** (from 62 fewer to 344 more)

									Moderate	
								71 per 1000	**9 fewer per 1000** (from 62 fewer to 342 more)	

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.86 Attrition: IPT versus support group

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With attrition: IPT versus support group**		**Risk with control**	**Risk difference with attrition: IPT versus support group** (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
48 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	2/24 (8.3%)	2/24 (8.3%)	**RR 1** (0.15 to 6.53)	**Study population**	
										83 per 1000	**0 fewer per 1000** (from 71 fewer to 461 more)
										Moderate	
										83 per 1000	**0 fewer per 1000** (from 71 fewer to 459 more)

[1] High risk of selection bias due to unclear allocation concealment and statistically significant baseline differences with the control group showing a higher SES score/lower income and higher depression (CES-D) mean score
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.87 Attrition: facilitated self-help versus TAU

Quality assessment							Summary of findings		
		Inconsistency	**Indirectness**	**Imprecision**			**Study event rates (%)**		**Anticipated absolute effects**

Participants (studies) Follow-up	Risk of bias				Publication bias	Overall quality of evidence	With control	With attrition: facilitated self-help versus TAU	Relative effect (95% CI)	Risk with control	Risk difference with attrition: facilitated self-help versus TAU (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
1136 (3 studies) 15-20 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	324/562 (57.7%)	309/574 (53.8%)	**RR 0.94** (0.85 to 1.04)	**Study population**	
										577 per 1000	**35 fewer per 1000** (from 86 fewer to 23 more)
										Moderate	
										417 per 1000	**25 fewer per 1000** (from 63 fewer to 17 more)

1.3.88 Attrition: listening visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With attrition: listening visits versus TAU		Risk with control	Risk difference with attrition: listening visits versus TAU (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
1211 (3 studies) 20-52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	104/791 (13.1%)	82/420 (19.5%)	**RR 1.22** (0.93 to 1.6)	**Study population**	
										131 per 1000	**29 more per 1000** (from 9 fewer to 79 more)
										Moderate	
										102 per 1000	**22 more per 1000** (from 7 fewer to 61 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)

[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.89 Attrition: directive counselling versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With attrition: directive counselling versus TAU		Risk with control	Risk difference with attrition: directive counselling versus TAU (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
146 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	15/33 (45.5%)	41/113 (36.3%)	**RR 0.8** (0.51 to 1.25)	**Study population**	
										455 per 1000	**91 fewer per 1000** (from 223 fewer to 114 more)
										Moderate	
										455 per 1000	**91 fewer per 1000** (from 223 fewer to 114 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.90 Attrition: post-miscarriage counselling versus TAU/enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With attrition: post-miscarriage counselling versus TAU/enhanced TAU		Risk with control	Risk difference with attrition: post-miscarriage counselling versus TAU/enhanced TAU (95% CI)
Dropout (assessed with: incomplete data at endpoint)											

99 (2 studies) 2-7 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⊖⊖ **LOW**[1,2] due to imprecision	10/50 (20%)	8/49 (16.3%)	**RR 0.81** (0.35 to 1.89)	**Study population**	
										200 per 1000	**38 fewer per 1000** (from 130 fewer to 178 more)
										Moderate	
										209 per 1000	**40 fewer per 1000** (from 136 fewer to 186 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.91 Attrition: post-traumatic birth counselling versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With attrition: post-traumatic birth counselling versus TAU**		**Risk with control**	**Risk difference with attrition: post-traumatic birth counselling versus TAU** (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
103 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness		undetected	See comment	0/53 (0%)	0/50 (0%)	**not pooled**	See comment	See comment

1.3.92 Attrition: social support versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With attrition: social support versus TAU**		**Risk with control**	**Risk difference with attrition: social support versus TAU** (95% CI)
Dropout (assessed with: incomplete data at endpoint)											

807 (3 studies) 8-14 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	37/404 (9.2%)	56/403 (13.9%)	**RR 1.49** (0.83 to 2.68)	**Study population**	
										92 per 1000	**45 more per 1000** (from 16 fewer to 154 more)
										Moderate	
										46 per 1000	**23 more per 1000** (from 8 fewer to 77 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.93 Attrition: psychologically (CBT/IPT)-informed psychoeducation versus TAU/enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With attrition: psychologically (CBT/IPT)-informed psychoeducation versus TAU/enhanced TAU**		**Risk with control**	**Risk difference with attrition: psychologically (CBT/IPT)-informed psychoeducation versus TAU/enhanced TAU** (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
2375 (13 studies) 4-31 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	155/1125 (13.8%)	222/1250 (17.8%)	**RR 1.17** (0.94 to 1.45)	**Study population**	
										138 per 1000	**23 more per 1000** (from 8 fewer to 62 more)
										Moderate	
										80 per 1000	**14 more per 1000** (from 5 fewer to 36 more)

[1] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.94 Attrition: non-mental health-focused education and support versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With attrition: non-mental health-focused education and support versus TAU**		**Risk with control**	**Risk difference with attrition: non-mental health-focused education and support versus TAU** (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
331 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	72/163 (44.2%)	71/168 (42.3%)	**RR 0.96** (0.75 to 1.22)	**Study population**	
										442 per 1000	**18 fewer per 1000** (from 110 fewer to 97 more)
										Moderate	
										442 per 1000	**18 fewer per 1000** (from 111 fewer to 97 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)

[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.95 Attrition: home visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With attrition: home visits versus TAU**		**Risk with control**	**Risk difference with attrition: home visits versus TAU** (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
					undetected					**Study population**	

1252 (4 studies) 6-52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]		⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	129/624 (20.7%)	139/628 (22.1%)	**RR 1.07** (0.86 to 1.32)	**207 per 1000**	**14 more per 1000** (from 29 fewer to 66 more)
										Moderate	
										196 per 1000	**14 more per 1000** (from 27 fewer to 63 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.96 Attrition: mother–infant relationship interventions versus TAU/enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With attrition: mother–infant relationship interventions versus TAU/enhanced TAU**		**Risk with control**	**Risk difference with attrition: mother–infant relationship interventions versus TAU/enhanced TAU** (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
576 (5 studies) 5-28 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	70/294 (23.8%)	58/282 (20.6%)	**RR 0.84** (0.63 to 1.12)	**Study population**	
										238 per 1000	**38 fewer per 1000** (from 88 fewer to 29 more)
										Moderate	
										143 per 1000	**23 fewer per 1000** (from 53 fewer to 17 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.97 Attrition: mother–infant relationship intervention with video feedback versus mother–infant relationship intervention with verbal feedback

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With attrition: mother–infant relationship intervention with video feedback versus mother–infant relationship intervention with verbal feedback**		**Risk with control**	**Risk difference with attrition: mother–infant relationship intervention with video feedback versus mother–infant relationship intervention with verbal feedback** (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
51 (1 study) 3 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	6/26 (23.1%)	5/25 (20%)	**RR 0.87** (0.3 to 2.48)	**Study population**	
										231 per 1000	**30 fewer per 1000** (from 162 fewer to 342 more)
										Moderate	
										231 per 1000	**30 fewer per 1000** (from 162 fewer to 342 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.98 Attrition: mother–infant relationship intervention (and guided self-help) versus listening visits (and guided self-help)

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**
							With control	**With attrition: mother–infant relationship intervention (and guided self-help) versus listening visits (and guided self-help)**		**Risk with control**	**Risk difference with attrition: mother–infant relationship intervention (and guided self-help) versus listening visits (and guided self-help)** (95% CI)

Dropout (assessed with: incomplete data at endpoint)										
80 (1 study) 35 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⊖⊖ **LOW**[1,2] due to imprecision	1/40 (2.5%) 2/40 (5%)	**RR 2** (0.19 to 21.18)	**Study population**	
									25 per 1000	**25 more per 1000** (from 20 fewer to 505 more)
									Moderate	
									25 per 1000	**25 more per 1000** (from 20 fewer to 505 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.99 Attrition: co-parenting intervention versus enhanced TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With attrition: co-parenting intervention versus enhanced TAU**		**Risk with control**	**Risk difference with attrition: co-parenting intervention versus enhanced TAU** (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
29 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness		undetected	See comment	0/13 (0%)	0/16 (0%)	**not pooled**	See comment	See comment

1.3.100 Attrition: music therapy during birth versus TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With attrition: music therapy**		**Risk with control**	**Risk difference with attrition: music therapy**

							during birth versus TAU			during birth versus TAU (95% CI)	
Dropout (assessed with: incomplete data at endpoint)											
141 (1 study) 3 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	11/70 (15.7%)	9/71 (12.7%)	**RR 0.81** (0.36 to 1.83)	**Study population**	
										157 per 1000	**30 fewer per 1000** (from 101 fewer to 130 more)
										Moderate	
										157 per 1000	**30 fewer per 1000** (from 100 fewer to 130 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.101 Attrition: psychosomatic intervention versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With attrition: psychosomatic intervention versus TAU**		**Risk with control**	**Risk difference with attrition: psychosomatic intervention versus TAU** (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
276 (2 studies) 34-52 weeks	no serious risk of bias	serious[1]	no serious indirectness	very serious[2,3]	undetected	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to inconsistency, imprecision	57/138 (41.3%)	48/138 (34.8%)	**RR 0.87** (0.54 to 1.39)	**Study population**	
										413 per 1000	**54 fewer per 1000** (from 190 fewer to 161 more)
										Moderate	

									435 per 1000	**57 fewer per 1000** (from 200 fewer to 170 more)

[1] There was evidence of moderate to substantial heterogeneity between effect sizes
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.102 Attrition: mindfulness training versus enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With attrition: mindfulness training versus enhanced TAU**		**Risk with control**	**Risk difference with attrition: mindfulness training versus enhanced TAU** (95% CI)
Dropout (assessed with: incomplete data at endpoint)											
47 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⊖⊖ **LOW**[1,2] due to imprecision	3/23 (13%)	4/24 (16.7%)	**RR 1.28** (0.32 to 5.1)	**Study population**	
										130 per 1000	**37 more per 1000** (from 89 fewer to 535 more)
										Moderate	
										130 per 1000	**36 more per 1000** (from 88 fewer to 533 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.103 Infant service use: facilitated self-help versus TAU

Quality assessment							Summary of findings		
		Inconsistency	**Indirectness**	**Imprecision**			**Study event rates (%)**		**Anticipated absolute effects**

Participants (studies) Follow-up	**Risk of bias**				**Publication bias**	**Overall quality of evidence**	**With control**	**With infant service use: facilitated self-help versus TAU**	**Relative effect** (95% CI)	**Risk with control**	**Risk difference with infant service use: facilitated self-help versus TAU** (95% CI)
Infant hospital post-treatment (service utilisation at endpoint) – ITT analysis (assessed with: Adult Service Use Schedule (AD-SUS): infant hospital)											
83 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	21/42 (50%)	15/41 (36.6%)	**RR 0.73** (0.44 to 1.21)	**Study population**	
										500 per 1000	**135 fewer per 1000** (from 280 fewer to 105 more)
										Moderate	
										500 per 1000	**135 fewer per 1000** (from 280 fewer to 105 more)
Infant hospital post-treatment (service utilisation at endpoint) – available case analysis (assessed with: Adult Service Use Schedule (AD-SUS): infant hospital)											
57 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	6/27 (22.2%)	4/30 (13.3%)	**RR 0.6** (0.19 to 1.9)	**Study population**	
										222 per 1000	**89 fewer per 1000** (from 180 fewer to 200 more)
										Moderate	
										222 per 1000	**89 fewer per 1000** (from 180 fewer to 200 more)
Infant hospital post-treatment (service utilisation at endpoint) – available case analysis (measured with: Adult Service Use Schedule (AD-SUS): infant hospital; better indicated by lower values)											
57 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,4]	reporting bias strongly suspected[3]	⨁⊖⊖⊖ **VERY LOW**[2,3,4] due to imprecision, publication bias	27	30	-		The mean infant hospital post-treatment (service utilisation at endpoint) – available case analysis in the intervention groups was **0.12 standard deviations lower** (0.64 lower to 0.4 higher)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data
[4] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.104 Infant service use: listening visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant service use: listening visits versus TAU**		**Risk with control**	**Risk difference with infant service use: listening visits versus TAU** (95% CI)
Infant hospital post-treatment (service utilisation at endpoint) – ITT analysis (assessed with: child health service use – visits to hospital doctors (previous month))											
731 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to imprecision, publication bias	130/548 (23.7%)	40/183 (21.9%)	**RR 0.92** (0.67 to 1.26)	**Study population**	
										237 per 1000	**19 fewer per 1000** (from 78 fewer to 62 more)
										Moderate	
										237 per 1000	**19 fewer per 1000** (from 78 fewer to 62 more)
Infant hospital post-treatment (service utilisation at endpoint) – available case analysis (assessed with: child health service use – visits to hospital doctors (previous month))											
653 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to imprecision, publication bias	70/488 (14.3%)	22/165 (13.3%)	**RR 0.93** (0.6 to 1.45)	**Study population**	
										143 per 1000	**10 fewer per 1000** (from 57 fewer to 65 more)
										Moderate	

									143 per 1000	**10 fewer per 1000** (from 57 fewer to 64 more)
Visit to A&E post-treatment (service utilisation measured at endpoint) – ITT analysis (assessed with: child health service use – visits to A&E (previous month))										
731 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias strongly suspected[3]	⊕⊕⊖⊖ **LOW**[1,3] due to imprecision, publication bias	209/548 (38.1%)	70/183 (38.3%)	**RR 1** (0.81 to 1.24)	**Study population**
									381 per 1000	**0 fewer per 1000** (from 72 fewer to 92 more)
									Moderate	
									381 per 1000	**0 fewer per 1000** (from 72 fewer to 91 more)
Visit to A&E post-treatment (service utilisation measured at endpoint) – available case analysis (assessed with: child health service use – visits to A&E (previous month))										
621 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	123/462 (26.6%)	46/159 (28.9%)	**RR 1.09** (0.82 to 1.45)	**Study population**
									266 per 1000	**24 more per 1000** (from 48 fewer to 120 more)
									Moderate	
									266 per 1000	**24 more per 1000** (from 48 fewer to 120 more)
Visit to NHS health visitor at clinic post-treatment (service utilisation [in past month] at endpoint) – ITT analysis (assessed with: child health service use – visits to NHS health visitor at clinic (previous month))										
731 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias strongly suspected[3]	⊕⊕⊖⊖ **LOW**[1,3] due to imprecision, publication bias	215/548 (39.2%)	70/183 (38.3%)	**RR 0.97** (0.79 to 1.2)	**Study population**
									392 per 1000	**12 fewer per 1000** (from 82 fewer to 78 more)

										Moderate	
										392 per 1000	**12 fewer per 1000** (from 82 fewer to 78 more)
Visit to NHS health visitor at clinic post-treatment (service utilisation [in past month] at endpoint) – available case analysis (assessed with: child health service use – visits to NHS health visitor at clinic (previous month))											
653 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1,2]	reporting bias strongly suspected[3]	⊕⊕⊖⊖ **LOW**[1,2,3] due to imprecision, publication bias	155/488 (31.8%)	52/165 (31.5%)	**RR 0.99** (0.77 to 1.29)	**Study population**	
										318 per 1000	**3 fewer per 1000** (from 73 fewer to 92 more)
										Moderate	
										318 per 1000	**3 fewer per 1000** (from 73 fewer to 92 more)
Visit from NHS health visitor at home post-treatment (service utilisation [in past month] at endpoint) – by intervention (assessed with: child health service use – visits from NHS health visitor at home (previous month))											
731 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	77/548 (14.1%)	29/183 (15.8%)	**RR 1.13** (0.76 to 1.67)	**Study population**	
										141 per 1000	**18 more per 1000** (from 34 fewer to 94 more)
										Moderate	
										141 per 1000	**18 more per 1000** (from 34 fewer to 94 more)
Visit from NHS health visitor at home post-treatment (service utilisation [in past month] at endpoint) – by intervention (assessed with: child health service use – visits from NHS health visitor at home (previous month))											
				very serious[1,2]						**Study population**	

653 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness		reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	17/488 (3.5%)	11/165 (6.7%)	**RR 1.91** (0.92 to 4)	**35 per 1000**	**32 more per 1000** (from 3 fewer to 105 more)
										Moderate	
										35 per 1000	**32 more per 1000** (from 3 fewer to 105 more)
Visit to GP post-treatment (service utilisation [in past month] at endpoint) – ITT analysis (assessed with: child health service use – visit to GP (previous month))											
731 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[3]	⊕⊕⊕⊖ **MODERATE**[3] due to publication bias	299/548 (54.6%)	81/183 (44.3%)	**RR 0.81** (0.68 to 0.97)	**Study population**	
										546 per 1000	**104 fewer per 1000** (from 16 fewer to 175 fewer)
										Moderate	
										546 per 1000	**104 fewer per 1000** (from 16 fewer to 175 fewer)
Visit to GP post-treatment (service utilisation [in past month] at endpoint) – available case analysis (assessed with: child health service use – visit to GP (previous month))											
653 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[3]	⊕⊕⊕⊖ **MODERATE**[3] due to publication bias	239/488 (49%)	63/165 (38.2%)	**RR 0.78** (0.63 to 0.97)	**Study population**	
										490 per 1000	**108 fewer per 1000** (from 15 fewer to 181 fewer)
										Moderate	
										490 per 1000	**108 fewer per 1000** (from 15 fewer to 181 fewer)
Any medication post-treatment (medication use [in past week] at endpoint) – ITT analysis (assessed with: child medication use: any medication (previous week))											

731 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[3]	⊕⊕⊕⊖ **MODERATE**[3] due to publication bias	366/548 (66.8%)	130/183 (71%)	**RR 1.06** (0.95 to 1.19)	**Study population**	
										668 per 1000	**40 more per 1000** (from 33 fewer to 127 more)
										Moderate	
										668 per 1000	**40 more per 1000** (from 33 fewer to 127 more)
Any medication post-treatment (past medication use measured at endpoint) – by intervention (assessed with: child medication use: any medication (previous week))											
657 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[3]	⊕⊕⊕⊖ **MODERATE**[3] due to publication bias	310/492 (63%)	109/165 (66.1%)	**RR 1.05** (0.92 to 1.19)	**Study population**	
										630 per 1000	**32 more per 1000** (from 50 fewer to 120 more)
										Moderate	
										630 per 1000	**31 more per 1000** (from 50 fewer to 120 more)
Antibiotics post-treatment (medication use [in past week] at endpoint) – ITT analysis (assessed with: child medication use: Antibiotics (previous week))											
731 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	106/548 (19.3%)	35/183 (19.1%)	**RR 0.99** (0.7 to 1.39)	**Study population**	
										193 per 1000	**2 fewer per 1000** (from 58 fewer to 75 more)
										Moderate	
										193 per 1000	**2 fewer per 1000** (from 58 fewer to 75 more)

Antibiotics post-treatment (medication use [in past week] at endpoint) – available case analysis (assessed with: child medication use: Antibiotics (previous week))											
657 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	50/492 (10.2%)	17/165 (10.3%)	**RR 1.01** (0.6 to 1.71)	**Study population**	
										102 per 1000	**1 more per 1000** (from 41 fewer to 72 more)
										Moderate	
										102 per 1000	**1 more per 1000** (from 41 fewer to 72 more)
Asthma medication post-treatment (medication use [in past week] at endpoint) – ITT analysis (assessed with: child medication use: Asthma medication (previous week))											
731 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	76/548 (13.9%)	20/183 (10.9%)	**RR 0.79** (0.5 to 1.25)	**Study population**	
										139 per 1000	**29 fewer per 1000** (from 69 fewer to 35 more)
										Moderate	
										139 per 1000	**29 fewer per 1000** (from 69 fewer to 35 more)
Asthma medication post-treatment (medication use [in past week] at endpoint) – available case analysis (assessed with: child medication use: Asthma medication (previous week))											
657 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	20/492 (4.1%)	2/165 (1.2%)	**RR 0.3** (0.07 to 1.26)	**Study population**	
										41 per 1000	**28 fewer per 1000** (from 38 fewer to 11 more)
										Moderate	

									41 per 1000	**29 fewer per 1000** (from 38 fewer to 11 more)
Skin ointment post-treatment (medication use [in past week] at endpoint) – ITT analysis (assessed with: child medication use: skin ointment (previous week))										
731 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias strongly suspected[3]	⊕⊕⊖⊖ **LOW**[1,3] due to imprecision, publication bias	178/548 (32.5%)	41/183 (22.4%)	**RR 0.69** (0.51 to 0.93)	**Study population**
									325 per 1000	**101 fewer per 1000** (from 23 fewer to 159 fewer)
									Moderate	
									325 per 1000	**101 fewer per 1000** (from 23 fewer to 159 fewer)
Skin ointment post-treatment (medication use [in past week] at endpoint) – available case analysis (assessed with: child medication use: skin ointment (previous week))										
657 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	reporting bias strongly suspected[3]	⊕⊕⊖⊖ **LOW**[1,3] due to imprecision, publication bias	122/492 (24.8%)	23/165 (13.9%)	**RR 0.56** (0.37 to 0.85)	**Study population**
									248 per 1000	**109 fewer per 1000** (from 37 fewer to 156 fewer)
									Moderate	
									248 per 1000	**109 fewer per 1000** (from 37 fewer to 156 fewer)
Visit to A&E long follow-up (service utilisation [in past month] at >24-week follow-up) – ITT analysis (assessed with: child health service use – visits to A&E (previous month))										
731 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1,2]	reporting bias strongly suspected[3]	⊕⊕⊖⊖ **LOW**[1,2,3] due to imprecision, publication bias	186/548 (33.9%)	67/183 (36.6%)	**RR 1.08** (0.86 to 1.35)	**Study population**
									339 per 1000	**27 more per 1000** (from 48 fewer to 119 more)

									Moderate	
									339 per 1000	**27 more per 1000** (from 47 fewer to 119 more)
Visit to A&E long follow-up (service utilisation [in past month] at >24-week follow-up) – available case analysis (assessed with: child health service use – visits to A&E (previous month))										
597 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	91/453 (20.1%)	28/144 (19.4%)	**RR 0.97** (0.66 to 1.42)	**Study population**
										201 per 1000 **6 fewer per 1000** (from 68 fewer to 84 more)
									Moderate	
									201 per 1000	**6 fewer per 1000** (from 68 fewer to 84 more)
Visit to NHS health visitor at clinic long follow-up (service utilisation [in past month] at >24-week follow-up) – ITT analysis (assessed with: child health service use – visits to NHS health visitor at clinic (previous month))										
731 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1,2]	reporting bias strongly suspected[3]	⊕⊕⊖⊖ **LOW**[1,2,3] due to imprecision, publication bias	144/548 (26.3%)	61/183 (33.3%)	**RR 1.27** (0.99 to 1.63)	**Study population**
										263 per 1000 **71 more per 1000** (from 3 fewer to 166 more)
									Moderate	
									263 per 1000	**71 more per 1000** (from 3 fewer to 166 more)
Visit to NHS health visitor at clinic long follow-up (service utilisation [in past month] at >24-week follow-up) – available case analysis (assessed with: child health service use – visits to NHS health visitor at clinic (previous month))										
				very serious[1,2]						**Study population**

601 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness		reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	52/456 (11.4%)	23/145 (15.9%)	**RR 1.39** (0.88 to 2.19)	**114 per 1000**	**44 more per 1000** (from 14 fewer to 136 more)
										Moderate	
										114 per 1000	**44 more per 1000** (from 14 fewer to 136 more)
Visit to GP long follow-up (service utilisation [in past month] at >24-week follow-up) – ITT analysis (assessed with: child health service use – visit to GP (previous month))											
731 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly suspected[3]	⊕⊕⊕⊖ **MODERATE**[3] due to publication bias	277/548 (50.5%)	91/183 (49.7%)	**RR 0.98** (0.83 to 1.16)	**Study population**	
										505 per 1000	**10 fewer per 1000** (from 86 fewer to 81 more)
										Moderate	
										506 per 1000	**10 fewer per 1000** (from 86 fewer to 81 more)
Visit to GP long follow-up (service utilisation [in past month] at >24-week follow-up) – available case analysis (assessed with: child health service use – visit to GP (previous month))											
601 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1,2]	reporting bias strongly suspected[3]	⊕⊕⊖⊖ **LOW**[1,2,3] due to imprecision, publication bias	185/456 (40.6%)	53/145 (36.6%)	**RR 0.9** (0.71 to 1.15)	**Study population**	
										406 per 1000	**41 fewer per 1000** (from 118 fewer to 61 more)
										Moderate	
										406 per 1000	**41 fewer per 1000** (from 118 fewer to 61 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)

[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

[3] Paper omits data

1.3.105 Infant service use: home visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With infant service use: home visits versus TAU		Risk with control	Risk difference with infant service use: home visits versus TAU (95% CI)
Infant hospital post-treatment (service utilisation at endpoint) – ITT analysis (assessed with: medical record: child hospitalisations)											
364 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to risk of bias, imprecision	106/185 (57.3%)	83/179 (46.4%)	**RR 0.81** (0.66 to 0.99)	**Study population**	
										573 per 1000	**109 fewer per 1000** (from 6 fewer to 195 fewer)
										Moderate	
										573 per 1000	**109 fewer per 1000** (from 6 fewer to 195 fewer)
Infant hospital post-treatment (service utilisation at endpoint) – available case analysis (assessed with: medical record: child hospitalisations)											
268 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to risk of bias, imprecision	58/137 (42.3%)	35/131 (26.7%)	**RR 0.63** (0.45 to 0.89)	**Study population**	
										423 per 1000	**157 fewer per 1000** (from 47 fewer to 233 fewer)
										Moderate	
										423 per 1000	**157 fewer per 1000** (from 47 fewer to 233 fewer)

Visit to A&E post-treatment (service utilisation measured at endpoint) – ITT analysis (assessed with: medical record: child seen in emergency department)											
364 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to risk of bias	155/185 (83.8%)	154/179 (86%)	**RR 1.03** (0.94 to 1.12)	**Study population**	
										838 per 1000	**25 more per 1000** (from 50 fewer to 101 more)
										Moderate	
										838 per 1000	**25 more per 1000** (from 50 fewer to 101 more)
Visit to A&E post-treatment (service utilisation measured at endpoint) – available case analysis (assessed with: medical record: child seen in emergency department)											
268 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to risk of bias	107/137 (78.1%)	106/131 (80.9%)	**RR 1.04** (0.92 to 1.17)	**Study population**	
										781 per 1000	**31 more per 1000** (from 62 fewer to 133 more)
										Moderate	
										838 per 1000	**34 more per 1000** (from 67 fewer to 142 more)
Any medication post-treatment (past medication use measured at endpoint) – available case analysis (assessed with: study-specific child health questionnaire: administration of medication to child without advice of medical practitioner)											
138 (1 study) 52 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2,4]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3,4] due to risk of bias, imprecision	8/70 (11.4%)	14/68 (20.6%)	**RR 1.8** (0.81 to 4.02)	**Study population**	
										114 per 1000	**91 more per 1000** (from 22 fewer to 345 more)
										Moderate	

									114 per 1000	**91 more per 1000** (from 22 fewer to 344 more)

[1] High risk of selection bias due to unclear allocation concealment and statistically significant baseline differences in poor psychological resources (37% intervention group versus 50% control) and in prenatal enrolment (41% intervention group and 53% control)
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] High risk of selection bias due to statistically significant baseline group differences in: parity (54% of intervention group primiparous versus 33% of control); identification as indigenous Australian (9% of intervention versus 2% of control); mental illness of partner (3% of intervention versus 14% of control); history of postnatal depression (11% of intervention versus 28% of control); physical domestic abuse (2% of intervention versus 10% of control); potential for child abuse (mean CAPI score in intervention was 123 versus 159 in control, and elevated CAPI score for 12% of intervention group versus 30% of control group)
[4] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.106 Infant service use: mother–infant relationship interventions versus TAU/enhanced TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant service use: mother–infant relationship interventions versus TAU/enhanced TAU**		**Risk with control**	**Risk difference with infant service use: mother–infant relationship interventions versus TAU/enhanced TAU** (95% CI)
Infant hospital post-treatment (service utilisation at endpoint) – ITT analysis (assessed with: infant service use: rehospitalised after discharge from NICU)											
121 (1 study) 25 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	26/61 (42.6%)	31/60 (51.7%)	**RR 1.21** (0.83 to 1.77)	**Study population**	
										426 per 1000	**90 more per 1000** (from 72 fewer to 328 more)
										Moderate	
										426 per 1000	**89 more per 1000** (from 72 fewer to 328 more)
Infant hospital post-treatment (service utilisation at endpoint) – available case analysis (assessed with: infant service use: rehospitalised after discharge from NICU)											
					undetected					**Study population**	

95 (1 study) 25 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]		⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	14/49 (28.6%)	17/46 (37%)	**RR 1.29** (0.72 to 2.31)	**286 per 1000**	**83 more per 1000** (from 80 fewer to 374 more)
										Moderate	
										286 per 1000	**83 more per 1000** (from 80 fewer to 375 more)
Contact with specialised healthcare services post-treatment (service utilisation at endpoint) – ITT analysis (assessed with: infant service use: contact with specialised health care services)											
121 (1 study) 25 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	39/61 (63.9%)	46/60 (76.7%)	**RR 1.2** (0.95 to 1.52)	**Study population**	
										639 per 1000	**128 more per 1000** (from 32 fewer to 332 more)
										Moderate	
										639 per 1000	**128 more per 1000** (from 32 fewer to 332 more)
Contact with specialised healthcare services post-treatment (service utilisation at endpoint) – available case analysis (assessed with: infant service use: contact with specialised health care services)											
95 (1 study) 25 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	27/49 (55.1%)	32/46 (69.6%)	**RR 1.26** (0.92 to 1.73)	**Study population**	
										551 per 1000	**143 more per 1000** (from 44 fewer to 402 more)
										Moderate	
										551 per 1000	**143 more per 1000** (from 44 fewer to 402 more)
Contact with developmental/rehabilitation specialist post-treatment (service utilisation at endpoint) – ITT analysis (assessed with: infant service use: contact with developmental/rehabilitation specialist)											
121 (1 study) 25 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	42/61 (68.9%)	44/60 (73.3%)	**RR 1.07** (0.85 to 1.34)	**Study population**	
										689 per 1000	**48 more per 1000** (from 103 fewer to 234 more)

									Moderate	
									689 per 1000	**48 more per 1000** (from 103 fewer to 234 more)
Contact with developmental/rehabilitation specialist post-treatment (service utilisation at endpoint) – available case analysis (assessed with: infant service use: contact with developmental/rehabilitation specialist)										
95 (1 study) 25 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	30/49 (61.2%)	30/46 (65.2%)	**RR 1.07** (0.78 to 1.45)	**Study population**
									612 per 1000	**43 more per 1000** (from 135 fewer to 276 more)
									Moderate	
									612 per 1000	**43 more per 1000** (from 135 fewer to 275 more)
Any medication post-treatment (medication use [in past week] at endpoint) – ITT analysis (assessed with: infant service use: medication)										
121 (1 study) 25 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	45/61 (73.8%)	51/60 (85%)	**RR 1.15** (0.96 to 1.38)	**Study population**
									738 per 1000	**111 more per 1000** (from 30 fewer to 280 more)
									Moderate	
									738 per 1000	**111 more per 1000** (from 30 fewer to 280 more)
Any medication post-treatment (past medication use measured at endpoint) – available case analysis (assessed with: infant service use: medication)										
95 (1 study) 25 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	33/49 (67.3%)	37/46 (80.4%)	**RR 1.19** (0.94 to 1.52)	**Study population**
									673 per 1000	**128 more per 1000** (from 40 fewer to 350 more)
									Moderate	

									674 per 1000	**128 more per 1000** (from 40 fewer to 350 more)
Surgery post-treatment (service utilisation at endpoint) – ITT analysis (assessed with: infant service use: surgery after discharge from NICU)										
109 (1 study) 25 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	19/49 (38.8%)	20/60 (33.3%)	**RR 0.86** (0.52 to 1.42)	**Study population**
									388 per 1000	**54 fewer per 1000** (from 186 fewer to 163 more)
									Moderate	
									388 per 1000	**54 fewer per 1000** (from 186 fewer to 163 more)
Surgery post-treatment (service utilisation at endpoint) – available case analysis (assessed with: infant service use: surgery after discharge from NICU)										
95 (1 study) 25 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	7/49 (14.3%)	6/46 (13%)	**RR 0.91** (0.33 to 2.52)	**Study population**
									143 per 1000	**13 fewer per 1000** (from 96 fewer to 217 more)
									Moderate	
									143 per 1000	**13 fewer per 1000** (from 96 fewer to 217 more)
Oxygen therapy post-treatment (service utilisation at endpoint) – ITT analysis (assessed with: infant service use: oxygen therapy)										
121 (1 study) 25 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	14/61 (23%)	16/60 (26.7%)	**RR 1.16** (0.62 to 2.17)	**Study population**
									230 per 1000	**37 more per 1000** (from 87 fewer to 269 more)
									Moderate	
									230 per 1000	**37 more per 1000** (from 87 fewer to 269 more)
Oxygen therapy post-treatment (service utilisation at endpoint) – available case analysis (assessed with: infant service use: oxygen therapy)										

95 (1 study) 25 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	2/49 (4.1%)	2/46 (4.3%)	**RR 1.07** (0.16 to 7.25)	**Study population**	
										41 per 1000	**3 more per 1000** (from 34 fewer to 255 more)
										Moderate	
										41 per 1000	**3 more per 1000** (from 34 fewer to 256 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.107 Infant physical health: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant physical health: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU**		**Risk with control**	**Risk difference with infant physical health: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU** (95% CI)
Underweight post-treatment (underweight at endpoint or first measurement) – ITT analysis (assessed with: child is considered underweight if growth is less than the anthropometric cut-off of – 2 SD below the median WAZ and HAZ scores of the National Center for Health Statistics/WHO international references)											
903 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	318/440 (72.3%)	318/463 (68.7%)	**RR 0.95** (0.87 to 1.03)	**Study population**	
										723 per 1000	**36 fewer per 1000** (from 94 fewer to 22 more)
										Moderate	
										723 per 1000	**36 fewer per 1000** (from 94 fewer to 22 more)

Underweight post-treatment (underweight at endpoint or first measurement) – available case analysis (assessed with: child is considered underweight if growth is less than the anthropometric cut-off of – 2 SD below the median WAZ and HAZ scores of the National Center for Health Statistics/WHO international references)										
705 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	223/345 (64.6%)	215/360 (59.7%)	**RR 0.92** (0.82 to 1.04)	**Study population** **646 per 1000** **52 fewer per 1000** (from 116 fewer to 26 more) **Moderate** **646 per 1000** **52 fewer per 1000** (from 116 fewer to 26 more)
Weight-for-age post-treatment (mean *z* score at endpoint or first measurement) – available case analysis (measured with: weight-for-age Z score; better indicated by lower values)										
705 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	345	360	-	The mean weight-for-age post-treatment (mean *z* score at endpoint or first measurement) – available case analysis in the intervention groups was **0.13 standard deviations higher** (0.02 lower to 0.28 higher)
Stunted height post-treatment (short-for-age at endpoint or first measurement) – ITT analysis (assessed with: child is considered stunted if growth is less than the anthropometric cut-off of – 2 SD below the median WAZ and HAZ scores of the National Center for Health Statistics/WHO international references)										
903 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	176/440 (40%)	169/463 (36.5%)	**RR 0.91** (0.77 to 1.08)	**Study population** **400 per 1000** **36 fewer per 1000** (from 92 fewer to 32 more) **Moderate** **400 per 1000** **36 fewer per 1000** (from 92 fewer to 32 more)

Stunted height post-treatment (short-for-age at endpoint or first measurement) – available case analysis (assessed with: child is considered stunted if growth is less than the anthropometric cut-off of – 2 SD below the median WAZ and HAZ scores of the National Center for Health Statistics/WHO international references)										
705 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	81/345 (23.5%)	66/360 (18.3%)	**RR 0.78** (0.58 to 1.04)	**Study population** **235 per 1000** **52 fewer per 1000** (from 99 fewer to 9 more) **Moderate** **235 per 1000** **52 fewer per 1000** (from 99 fewer to 9 more)
Height-for-age post-treatment (mean *z* score at endpoint or first measurement) – available case analysis (measured with: height-for-age Z score; better indicated by lower values)										
705 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	345	360	-	The mean height-for-age post-treatment (mean *z* score at endpoint or first measurement) – available case analysis in the intervention groups was **0.24 standard deviations higher** (0.09 to 0.39 higher)
Diarrhoea post-treatment (≥1 diarrhoea episodes [in past 2 weeks] at endpoint or first measurement) – ITT analysis (assessed with: Diarrhoea was defined as ≥3 unformed stools passed in 24h, and a diarrhoeal episode was defined as being separated from another episode by at least 3 diarrhoea-free days)										
903 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	244/440 (55.5%)	219/463 (47.3%)	**RR 0.85** (0.75 to 0.97)	**Study population** **555 per 1000** **83 fewer per 1000** (from 17 fewer to 139 fewer) **Moderate** **555 per 1000** **83 fewer per 1000** (from 17 fewer to 139 fewer)

Diarrhoea post-treatment (≥1 diarrhoea episodes [in past 2 weeks] at endpoint or first measurement) – available case analysis (assessed with: Diarrhoea was defined as ≥3 unformed stools passed in 24h, and a diarrhoeal episode was defined as being separated from another episode by at least 3 diarrhoea-free days)										
705 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ **HIGH**	149/345 (43.2%)	116/360 (32.2%)	**RR 0.75** (0.62 to 0.9)	**Study population**
										432 per 1000 / **108 fewer per 1000** (from 43 fewer to 164 fewer)
										Moderate
										432 per 1000 / **108 fewer per 1000** (from 43 fewer to 164 fewer)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)

[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.108 Infant physical health: IPT versus support group

Quality assessment							**Summary of findings**				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant physical health: IPT versus support group**		**Risk with control**	**Risk difference with infant physical health: IPT versus support group** (95% CI)
Gestational age post-treatment (mean score at endpoint or first measurement) – available case analysis (Better indicated by lower values)											
44 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	22	22	-		The mean gestational age post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.3 standard deviations lower** (0.89 lower to 0.3 higher)
Birth weight post-treatment (mean score at endpoint or first measurement) – available case analysis (Better indicated by lower values)											

44 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to risk of bias, imprecision	22	22	-	The mean birth weight post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.08 standard deviations lower** (0.67 lower to 0.51 higher)

[1] High risk of selection bias due to unclear allocation concealment and statistically significant baseline differences with the control group showing a higher SES score/lower income and higher depression (CES-D) mean score
[2] Total population size is less than 400 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.109 Infant physical health: listening visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant physical health: listening visits versus TAU**		**Risk with control**	**Risk difference with infant physical health: listening visits versus TAU** (95% CI)
Ill health post-treatment (maternal concerns about child health at endpoint or first measurement) – ITT analysis (assessed with: child health and development concerns (maternal assessment): child's health)											
731 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1,2]	reporting bias strongly suspected[3]	⨁⨁⨀⨀ **LOW**[1,2,3] due to imprecision, publication bias	216/548 (39.4%)	60/183 (32.8%)	**RR 0.83** (0.66 to 1.05)	**Study population**	
										394 per 1000	**67 fewer per 1000** (from 134 fewer to 20 more)
										Moderate	
										394 per 1000	**67 fewer per 1000** (from 134 fewer to 20 more)
Ill health post-treatment (maternal concerns about child health at endpoint or first measurement) – available case analysis (assessed with: child health and development concerns (maternal assessment): child's health)											

650 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1,2]	reporting bias strongly suspected[3]	⨁⨁⨀⨀ **LOW**[1,2,3] due to imprecision, publication bias	156/488 (32%)	39/162 (24.1%)	**RR 0.75** (0.56 to 1.02)	**Study population**	
										320 per 1000	**80 fewer per 1000** (from 141 fewer to 6 more)
										Moderate	
										320 per 1000	**80 fewer per 1000** (from 141 fewer to 6 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data

1.3.110 Infant physical health: social support versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant physical health: social support versus TAU**		**Risk with control**	**Risk difference with infant physical health: social support versus TAU** (95% CI)
Infant cortisol levels post-treatment (mean score at endpoint or first measurement) – available case analysis (Better indicated by lower values)											
23 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	14	9	-		The mean infant cortisol levels post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.28 standard deviations higher** (0.56 lower to 1.12 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.111 Infant physical health: psychologically (CBT/IPT)-informed psychoeducation versus TAU/enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With infant physical health: psychologically (CBT/IPT)-informed psychoeducation versus TAU/enhanced TAU		Risk with control	Risk difference with infant physical health: psychologically (CBT/IPT)-informed psychoeducation versus TAU/enhanced TAU (95% CI)
Infant stress post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: Visual Analogue Scale (VAS): infant stress; better indicated by lower values)											
46 (1 study) 101 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	22	24	-		The mean infant stress post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.25 standard deviations higher** (0.33 lower to 0.83 higher)
Infant cortisol levels post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: Average (morning/evening) cortisol (log scores); better indicated by lower values)											
53 (1 study) 49 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	29	24	-		The mean infant cortisol levels post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.27 standard deviations lower** (0.82 lower to 0.27 higher)
Infant cortisol levels long follow-up (mean score at >24-week follow-up) – available case analysis (measured with: Average (morning/evening) cortisol (log scores); better indicated by lower values)											

46 (1 study) 101 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	reporting bias strongly suspected[4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, imprecision, publication bias	22	24	-		The mean infant cortisol levels long follow-up (mean score at >24-week follow-up) – available case analysis in the intervention groups was **0.11 standard deviations lower** (0.69 lower to 0.47 higher)

[1] High risk of selection bias due to statistically significant baseline/mid-treatment difference in average maternal salivary cortisol levels (0.62 in intervention group and 0.75 in control group)
[2] Total population size is less than 400 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] Paper omits data

1.3.112 Infant physical health: mother–infant relationship intervention (and guided self-help) versus listening visits (and guided self-help)

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant physical health: mother–infant relationship intervention (and guided self-help) versus listening visits (and guided self-help)**		**Risk with control**	**Risk difference with infant physical health: mother–infant relationship intervention (and guided self-help) versus listening visits (and guided self-help)** (95% CI)
Weight-for-age post-treatment (mean *z* score at endpoint or first measurement) – available case analysis (measured with: weight-for-age Z score; better indicated by lower values)											
77 (1 study) 35 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	39	38	-		The mean weight-for-age post-treatment (mean *z* score at endpoint or first measurement) – available case analysis in the intervention groups was **0.12 standard deviations lower** (0.56 lower to 0.33 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.113 Infant physical development: CBT versus listening visits

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With infant physical development: CBT versus listening visits		Risk with control	Risk difference with infant physical development: CBT versus listening visits (95% CI)
Infant motor development post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: Bayley Scales of Infant Development – Psychomotor development index; better indicated by lower values)											
34 (1 study)	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	14	20	-		The mean infant motor development post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.21 standard deviations higher** (0.47 lower to 0.9 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.114 Infant physical development: listening visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With infant physical development: listening visits versus TAU		Risk with control	Risk difference with infant physical development: listening visits versus TAU (95% CI)
Infant eating habits post-treatment (maternal concerns at endpoint or first measurement) – ITT analysis (assessed with: child health and development concerns (maternal assessment): child's eating habits)											

731 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1,2]	reporting bias strongly suspected[3]	⊕⊕⊖⊖ **LOW**[1,2,3] due to imprecision, publication bias	202/548 (36.9%)	61/183 (33.3%)	**RR 0.9** (0.72 to 1.14)	**Study population**	
										369 per 1000	**37 fewer per 1000** (from 103 fewer to 52 more)
										Moderate	
										369 per 1000	**37 fewer per 1000** (from 103 fewer to 52 more)
Infant eating habits post-treatment (maternal concerns at endpoint or first measurement) – available case analysis (assessed with: child health and development concerns (maternal assessment): child's eating habits)											
591 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,3] due to imprecision, publication bias	102/448 (22.8%)	21/143 (14.7%)	**RR 0.65** (0.42 to 0.99)	**Study population**	
										228 per 1000	**80 fewer per 1000** (from 2 fewer to 132 fewer)
										Moderate	
										228 per 1000	**80 fewer per 1000** (from 2 fewer to 132 fewer)
Infant sleeping habits post-treatment (maternal concerns at endpoint or first measurement) – ITT analysis (assessed with: child health and development concerns (maternal assessment): child's sleeping habits)											
731 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1,2]	reporting bias strongly suspected[3]	⊕⊕⊖⊖ **LOW**[1,2,3] due to imprecision, publication bias	159/548 (29%)	56/183 (30.6%)	**RR 1.05** (0.82 to 1.36)	**Study population**	
										290 per 1000	**15 more per 1000** (from 52 fewer to 104 more)
										Moderate	
										290 per 1000	**14 more per 1000** (from 52 fewer to 104 more)

Infant sleep problems post-treatment (maternal report at endpoint or first measurement) – available case analysis (assessed with: child health and development concerns (maternal assessment): child's sleeping habits)											
591 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	59/448 (13.2%)	16/143 (11.2%)	**RR 0.85** (0.51 to 1.43)	**Study population**	
										132 per 1000	**20 fewer per 1000** (from 65 fewer to 57 more)
										Moderate	
										132 per 1000	**20 fewer per 1000** (from 65 fewer to 57 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data

1.3.115 Infant physical development: home visits versus TAU

Quality assessment							**Summary of findings**				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant physical development: home visits versus TAU**		**Risk with control**	**Risk difference with infant physical development: home visits versus TAU** (95% CI)
Infant motor development post-treatment (below threshold at endpoint or first measurement) – ITT analysis (assessed with: Bayley Scales of Infant Development – Psychomotor development index<85)											
364 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	87/185 (47%)	72/179 (40.2%)	**RR 0.86** (0.68 to 1.08)	**Study population**	
										470 per 1000	**66 fewer per 1000** (from 150 fewer to 38 more)
										Moderate	
										470 per 1000	**66 fewer per 1000** (from 150 fewer to 38 more)

Infant motor development post-treatment (below threshold at endpoint or first measurement) – available case analysis (assessed with: Bayley Scales of Infant Development – Psychomotor development index<85)											
249 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	25/123 (20.3%)	19/126 (15.1%)	**RR 0.74** (0.43 to 1.28)	**Study population**	
										203 per 1000	**53 fewer per 1000** (from 116 fewer to 57 more)
										Moderate	
										203 per 1000	**53 fewer per 1000** (from 116 fewer to 57 more)
Infant feeding problems post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: study-specific child health questionnaire: Feeding problems; better indicated by lower values)											
138 (1 study) 52 weeks	serious[4]	no serious inconsistency	no serious indirectness	very serious[3,5]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4,5] due to risk of bias, imprecision	70	68	-		The mean infant feeding problems post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.2 standard deviations higher** (0.14 lower to 0.53 higher)
Infant sleep problems post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: study-specific child health questionnaire: sleeping problems; better indicated by lower values)											
138 (1 study) 52 weeks	serious[4]	no serious inconsistency	no serious indirectness	very serious[3,5]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4,5] due to risk of bias, imprecision	70	68	-		The mean infant sleep problems post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.18 standard deviations higher** (0.15 lower to 0.52 higher)

[1] High risk of selection bias due to unclear allocation concealment and statistically significant baseline differences in poor psychological resources (37% intervention group versus 50% control) and in prenatal enrolment (41% intervention group and 53% control)

[2] Total number of events is less than 300 (a threshold rule-of-thumb)

[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

[4] High risk of selection bias due to statistically significant baseline group differences in: parity (54% of intervention group primiparous versus 33% of control); identification as indigenous Australian (9% of intervention versus 2% of control); mental illness of partner (3% of intervention versus 14% of control); history of postnatal depression (11% of intervention versus 28% of control); physical domestic abuse (2% of intervention versus 10% of control); potential for child abuse (mean CAPI score in intervention was 123 versus 159 in control, and elevated CAPI

score for 12% of intervention group versus 30% of control group)
[5] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.116 Infant physical development: mother–infant relationship interventions versus TAU/enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With infant physical development: mother–infant relationship interventions versus TAU/enhanced TAU		Risk with control	Risk difference with infant physical development: mother–infant relationship interventions versus TAU/enhanced TAU (95% CI)
Infant motor development post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: Bayley Scales of Infant Development-Motor; better indicated by lower values)											
96 (1 study) 25 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	50	46	-		The mean infant motor development post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.12 standard deviations lower** (0.52 lower to 0.28 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.117 Infant physical development: infant sleep training (controlled crying) versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With infant physical development: infant sleep training (controlled crying) versus TAU		Risk with control	Risk difference with infant physical development: infant sleep training (controlled crying) versus TAU (95% CI)

Infant sleep problems post-treatment (maternal report at endpoint or first measurement) – available case analysis (assessed with: maternal report: infant sleep problem – Treatment non-response (no further detail reported))										
189 (2 studies) 9-13 weeks	no serious risk of bias	very serious[1]	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to inconsistency, imprecision	63/93 (67.7%) 41/96 (42.7%)	**RR 0.55** (0.25 to 1.19)	**Study population**	
									677 per 1000	**305 fewer per 1000** (from 508 fewer to 129 more)
									Moderate	
									661 per 1000	**297 fewer per 1000** (from 496 fewer to 126 more)
Infant sleep problems short follow-up (maternal report at 9-16-week follow-up) – available case analysis (assessed with: maternal report: infant sleep problem – Treatment non-response (no further detail reported))										
184 (2 studies) 17-22 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	52/88 (59.1%) 41/96 (42.7%)	**RR 0.73** (0.55 to 0.97)	**Study population**	
									591 per 1000	**160 fewer per 1000** (from 18 fewer to 266 fewer)
									Moderate	
									577 per 1000	**156 fewer per 1000** (from 17 fewer to 260 fewer)
Infant sleep problems long follow-up (maternal report at >24-week follow-up) – available case analysis (assessed with: maternal report: infant sleep problem – Treatment non-response (no further detail re)										
272 (1 study) 74 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊕⊖⊖ **LOW**[2,3] due to imprecision	42/129 (32.6%) 39/143 (27.3%)	**RR 0.84** (0.58 to 1.21)	**Study population**	
									326 per 1000	**52 fewer per 1000** (from 137 fewer to 68 more)
									Moderate	
									326 per 1000	**52 fewer per 1000** (from 137 fewer to 68 more)

[1] There was evidence of substantial to considerable heterogeneity between effect sizes
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.118 Infant cognitive development: CBT versus listening visits

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With infant cognitive development: CBT versus listening visits		Risk with control	Risk difference with infant cognitive development: CBT versus listening visits (95% CI)
Infant cognitive development post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: Bayley Scales of Infant Development – Mental development index; better indicated by lower values)											
34 (1 study)	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	14	20	-		The mean infant cognitive development post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.59 standard deviations higher** (0.11 lower to 1.29 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.119 Infant cognitive development: listening visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							With control	With infant cognitive development: listening visits versus TAU		Risk with control	Risk difference with infant cognitive development: listening visits versus TAU (95% CI)

Infant cognitive development post-treatment (maternal concerns/below threshold at endpoint or first measurement) – ITT analysis (assessed with: child health and development concerns (maternal assessment): child's development)										
731 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	93/548 (17%)	29/183 (15.8%)	**RR 0.93** (0.64 to 1.37)	**Study population** **170 per 1000**: **12 fewer per 1000** (from 61 fewer to 63 more) **Moderate** **170 per 1000**: **12 fewer per 1000** (from 61 fewer to 63 more)
Infant cognitive development post-treatment (maternal concerns/below threshold at endpoint or first measurement) – available case analysis (assessed with: child health and development concerns (maternal assessment): child's development)										
640 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	23/478 (4.8%)	8/162 (4.9%)	**RR 1.03** (0.47 to 2.25)	**Study population** **48 per 1000**: **1 more per 1000** (from 26 fewer to 60 more) **Moderate** **48 per 1000**: **1 more per 1000** (from 25 fewer to 60 more)
Infant verbal development post-treatment (maternal concerns at endpoint or first measurement) – ITT analysis (assessed with: child health and development concerns (maternal assessment): child's speech)										
731 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	166/548 (30.3%)	49/183 (26.8%)	**RR 0.88** (0.67 to 1.16)	**Study population** **303 per 1000**: **36 fewer per 1000** (from 100 fewer to 48 more) **Moderate**

										303 per 1000	**36 fewer per 1000** (from 100 fewer to 48 more)
Infant verbal development post-treatment (maternal concerns at endpoint or first measurement) – available case analysis (assessed with: child health and development concerns (maternal assessment): child's speech)											
591 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected[3]	⨁⊖⊖⊖ **VERY LOW**[1,3] due to imprecision, publication bias	66/448 (14.7%)	9/143 (6.3%)	**RR 0.43** (0.22 to 0.84)	**Study population**	
										147 per 1000	**84 fewer per 1000** (from 24 fewer to 115 fewer)
										Moderate	
										147 per 1000	**84 fewer per 1000** (from 24 fewer to 115 fewer)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data

1.3.120 Infant cognitive development: social support versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant cognitive development: social support versus TAU**		**Risk with control**	**Risk difference with infant cognitive development: social support versus TAU** (95% CI)
Infant cognitive development post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: Bayley Scales of Infant Development – Mental development index; better indicated by lower values)											
48 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⊖⊖ **LOW**[1,2] due to imprecision	27	21	-		The mean infant cognitive development post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention

									groups was **0.21 standard deviations lower** (0.78 lower to 0.36 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.121 Infant cognitive development: home visits versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant cognitive development: home visits versus TAU**		**Risk with control**	**Risk difference with infant cognitive development: home visits versus TAU** (95% CI)
Infant cognitive development post-treatment (maternal concerns/below threshold at endpoint or first measurement) – ITT analysis (assessed with: Bayley Scales of Infant Development – Mental development index<85)											
364 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2,3]	undetected	⨁⨁⨀⨀ **LOW**[1,2,3] due to risk of bias, imprecision	126/185 (68.1%)	106/179 (59.2%)	**RR 0.87** (0.74 to 1.02)	**Study population**	
										681 per 1000	**89 fewer per 1000** (from 177 fewer to 14 more)
										Moderate	
										681 per 1000	**89 fewer per 1000** (from 177 fewer to 14 more)
Infant cognitive development post-treatment (maternal concerns/below threshold at endpoint or first measurement) – available case analysis (assessed with: Bayley Scales of Infant Development – Mental development index<85)											
249 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⨀⨀⨀ **VERY LOW**[1,2,3] due to risk of bias, imprecision	64/123 (52%)	53/126 (42.1%)	**RR 0.81** (0.62 to 1.05)	**Study population**	
										520 per 1000	**99 fewer per 1000** (from 198 fewer to 26 more)
										Moderate	

									520 per 1000	**99 fewer per 1000** (from 198 fewer to 26 more)

[1] High risk of selection bias due to unclear allocation concealment and statistically significant baseline differences in poor psychological resources (37% intervention group versus 50% control) and in prenatal enrolment (41% intervention group and 53% control)
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.122 Infant cognitive development: mother–infant relationship interventions versus TAU/enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant cognitive development: mother–infant relationship interventions versus TAU/enhanced TAU**		**Risk with control**	**Risk difference with infant cognitive development: mother–infant relationship interventions versus TAU/enhanced TAU** (95% CI)
Infant cognitive development post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: Bayley Scales of Infant Development – Cognitive; better indicated by lower values)											
96 (1 study) 25 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	50	46	-		The mean infant cognitive development post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.07 standard deviations higher** (0.33 lower to 0.47 higher)
Infant verbal development post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: peabody Picture Vocabulary Test – Revised (PPVT-R): VIQ or Bayley Scales of Infant Development – Language; better indicated by lower values)											
154 (2 studies) 25-271 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	79	75	-		The mean infant verbal development post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was

										0.1 standard deviations higher (0.25 lower to 0.45 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.123 Infant emotional development: social support versus TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant emotional development: social support versus TAU**		**Risk with control**	**Risk difference with infant emotional development: social support versus TAU** (95% CI)
Infant 'difficult' temperament post-treatment (maternal-rated mean score at endpoint or first measurement) – available case analysis (measured with: infant Characteristics Questionnaire; better indicated by lower values)											
51 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to imprecision	28	23	-		The mean infant 'difficult' temperament post-treatment (maternal-rated mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.33 standard deviations higher** (0.23 lower to 0.88 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

1.3.124 Infant emotional development: home visits versus TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant emotional development: home visits versus TAU**		**Risk with control**	**Risk difference with infant emotional development: home visits versus TAU** (95% CI)

Infant externalizing post-treatment (symptomatology – above threshold at endpoint or first measurement) – ITT analysis (assessed with: child Behaviour Checklist (CBCL/1.5-5): Externalising)										
364 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	90/185 (48.6%)	76/179 (42.5%)	**RR 0.87** (0.7 to 1.09)	**Study population**
										486 per 1000 — **63 fewer per 1000** (from 146 fewer to 44 more)
										Moderate
										487 per 1000 — **63 fewer per 1000** (from 146 fewer to 44 more)
Infant externalizing post-treatment (symptomatology – above threshold at endpoint or first measurement) – available case analysis (assessed with: child Behaviour Checklist (CBCL/1.5-5): Externalising)										
249 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	28/123 (22.8%)	23/126 (18.3%)	**RR 0.8** (0.49 to 1.31)	**Study population**
										228 per 1000 — **46 fewer per 1000** (from 116 fewer to 71 more)
										Moderate
										228 per 1000 — **46 fewer per 1000** (from 116 fewer to 71 more)
Infant internalizing post-treatment (symptomatology – above threshold at endpoint or first measurement) – ITT analysis (assessed with: child Behaviour Checklist (CBCL/1.5-5): Internalising)										
364 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	88/185 (47.6%)	69/179 (38.5%)	**RR 0.81** (0.64 to 1.03)	**Study population**
										476 per 1000 — **90 fewer per 1000** (from 171 fewer to 14 more)
										Moderate
										476 per 1000 — **90 fewer per 1000** (from 171 fewer to 14 more)
Infant internalizing post-treatment (symptomatology – above threshold at endpoint or first measurement) – available case analysis (assessed with: child Behaviour Checklist (CBCL/1.5-5): Internalising)										

249 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	26/123 (21.1%)	16/126 (12.7%)	**RR 0.6** (0.34 to 1.06)	**Study population**
										211 per 1000 — **85 fewer per 1000** (from 140 fewer to 13 more)
										Moderate
										211 per 1000 — **84 fewer per 1000** (from 139 fewer to 13 more)
Infant social withdrawal post-treatment (symptomatology – above threshold at endpoint or first measurement) – ITT analysis (assessed with: Alarm Distress Baby Scale (ADBB) ≥5)										
440 (1 study) 87 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊕⊖⊖ **LOW**[2,3] due to imprecision	79/218 (36.2%)	69/222 (31.1%)	**RR 0.86** (0.66 to 1.12)	**Study population**
										362 per 1000 — **51 fewer per 1000** (from 123 fewer to 43 more)
										Moderate
										362 per 1000 — **51 fewer per 1000** (from 123 fewer to 43 more)
Infant social withdrawal post-treatment (symptomatology – above threshold at endpoint or first measurement) – available case analysis (assessed with: Alarm Distress Baby Scale (ADBB) ≥5)										
367 (1 study) 87 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊕⊖⊖ **LOW**[2,3] due to imprecision	44/183 (24%)	31/184 (16.8%)	**RR 0.7** (0.46 to 1.06)	**Study population**
										240 per 1000 — **72 fewer per 1000** (from 130 fewer to 14 more)
										Moderate
										240 per 1000 — **72 fewer per 1000** (from 130 fewer to 14 more)
Infant social withdrawal post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: Alarm Distress Baby Scale (ADBB); better indicated by lower values)										

160 (1 study) 87 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4] due to imprecision	84	76	-		The mean infant social withdrawal post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0 standard deviations higher** (0.31 lower to 0.31 higher)

[1] High risk of selection bias due to unclear allocation concealment and statistically significant baseline differences in poor psychological resources (37% intervention group versus 50% control) and in prenatal enrolment (41% intervention group and 53% control)
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.125 Infant emotional development: mother–infant relationship interventions versus TAU/enhanced TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant emotional development: mother–infant relationship interventions versus TAU/enhanced TAU**		**Risk with control**	**Risk difference with infant emotional development: mother–infant relationship interventions versus TAU/enhanced TAU** (95% CI)
Infant adaptive behaviour post-treatment (treatment response at endpoint or first measurement) – ITT analysis (assessed with: Ages and Stages Questionnaire: social-Emotional (ASQ:SE): Treatment response (improvement-reliable change index))											
80 (1 study) 26 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	7/40 (17.5%)	9/40 (22.5%)	**RR 1.29** (0.53 to 3.12)	**Study population**	
										175 per 1000	**51 more per 1000** (from 82 fewer to 371 more)
										Moderate	
										175 per 1000	**51 more per 1000** (from 82 fewer to 371 more)

Infant adaptive behaviour post-treatment (treatment response at endpoint or first measurement) – available case analysis (assessed with: Ages and Stages Questionnaire: social-Emotional (ASQ:SE): Treatment response (improvement-reliable change index))										
75 (1 study) 26 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	7/37 (18.9%)	9/38 (23.7%)	**RR 1.25** (0.52 to 3.01)	**Study population**
										189 per 1000 / **47 more per 1000** (from 91 fewer to 380 more)
										Moderate
										189 per 1000 / **47 more per 1000** (from 91 fewer to 380 more)
Infant adaptive behaviour post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: Ages and Stages Questionnaire: social-Emotional (ASQ:SE) or Infant Toddler Social and Emotional Assessment: Competence; better indicated by lower values)										
146 (2 studies) 26-57 weeks	serious[1]	very serious[4]	no serious indirectness	very serious[3,5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3,4,5] due to risk of bias, inconsistency, imprecision	73	73	-	The mean infant adaptive behaviour post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.21 standard deviations higher** (0.59 lower to 1 higher)
Infant externalizing post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: infant Toddler Social and Emotional Assessment: Externalizing; better indicated by lower values)										
71 (1 study) 57 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3,5]	undetected	⊕⊕⊖⊖ **LOW**[3,5] due to imprecision	36	35	-	The mean infant externalizing post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.09 standard deviations higher** (0.38 lower to 0.55 higher)
Infant internalizing post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: infant Toddler Social and Emotional Assessment: Internalizing; better indicated by lower values)										

71 (1 study) 57 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3,5]	undetected	⊕⊕⊖⊖ **LOW**[3,5] due to imprecision	36	35	-		The mean infant internalizing post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.3 standard deviations higher** (0.17 lower to 0.77 higher)
Infant dysregulation post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: infant Toddler Social and Emotional Assessment: Dysregulation; better indicated by lower values)											
71 (1 study) 57 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3,5]	undetected	⊕⊕⊖⊖ **LOW**[3,5] due to imprecision	36	35	-		The mean infant dysregulation post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.08 standard deviations lower** (0.54 lower to 0.39 higher)
Infant self-esteem post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: puppet Interview: child self-esteem; better indicated by lower values)											
58 (1 study) 271 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	29	29	-		The mean infant self-esteem post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **1.46 standard deviations higher** (0.88 to 2.05 higher)
Infant externalizing Very long follow-up (mean score at >104-week follow-up) – available case analysis (measured with: child Behaviour Checklist (CBCL/1.5-5): Externalising; better indicated by lower values)											
58 (1 study) 271 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3,5]	undetected	⊕⊕⊖⊖ **LOW**[3,5] due to imprecision	29	29	-		The mean infant externalizing very long follow-up (mean score at

										>104-week follow-up) – available case analysis in the intervention groups was **0.14 standard deviations lower** (0.65 lower to 0.38 higher)
Infant internalizing Very long follow-up (mean score at >104-week follow-up) – available case analysis (measured with: child Behaviour Checklist (CBCL/1.5-5): Internalising; better indicated by lower values)										
58 (1 study) 271 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⨁⨁⨀⨀ **LOW**[5] due to imprecision	29	29	-	The mean infant internalizing very long follow-up (mean score at >104-week follow-up) – available case analysis in the intervention groups was **1.79 standard deviations higher** (1.17 to 2.4 higher)

[1] High risk of selection bias due to statistically significant baseline difference in the age of infants (4.4 months old in intervention group versus 5.9 months old in TAU group)
[2] Total number of events is less than 300 (a threshold rule-of-thumb)
[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[4] There was evidence of substantial to considerable heterogeneity between effect sizes
[5] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.126 Infant emotional development: infant sleep training (controlled crying) versus TAU

Quality assessment							Summary of findings				
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With infant emotional development: infant sleep training (controlled crying) versus TAU**		**Risk with control**	**Risk difference with infant emotional development: infant sleep training (controlled crying) versus TAU** (95% CI)
Infant externalizing post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: child Behaviour Check List (CBCL) – Externalising; better indicated by lower values)											

268 (1 study) 74 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⨁⨁⨁⨀ **MODERATE**[1] due to imprecision	126	142	-		The mean infant externalizing post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.07 standard deviations higher** (0.17 lower to 0.31 higher)
Infant internalizing post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: child Behaviour Check List (CBCL) – Internalising; better indicated by lower values)											
268 (1 study) 74 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⨁⨁⨁⨀ **MODERATE**[1] due to imprecision	126	142	-		The mean infant internalizing post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.02 standard deviations higher** (0.22 lower to 0.26 higher)

[1] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.127 Prevention of neglect or abuse of the infant: listening visits versus TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**	
							With control	**With Prevention of neglect or abuse of the infant: listening visits versus TAU**		**Risk with control**	**Risk difference with Prevention of neglect or abuse of the infant: listening visits versus TAU** (95% CI)
Child injury post-treatment (injury requiring medical attention at endpoint or first measurement) – ITT analysis (assessed with: child health service use – Injury requiring medical attention)											
				serious[1,2]						**Study population**	

731 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness		reporting bias strongly suspected[3]	⊕⊕⊖⊖ **LOW**[1,2,3] due to imprecision, publication bias	128/548 (23.4%)	43/183 (23.5%)	**RR 1.01** (0.74 to 1.36)	**234 per 1000**	**2 more per 1000** (from 61 fewer to 84 more)
										Moderate	
										234 per 1000	**2 more per 1000** (from 61 fewer to 84 more)
Child injury post-treatment (injury requiring medical attention at endpoint or first measurement) – available case analysis (assessed with: child health service use – Injury requiring medical attention)											
651 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	67/487 (13.8%)	24/164 (14.6%)	**RR 1.06** (0.69 to 1.64)	**Study population**	
										138 per 1000	**8 more per 1000** (from 43 fewer to 88 more)
										Moderate	
										138 per 1000	**8 more per 1000** (from 43 fewer to 88 more)
Child injury long follow-up (injury requiring medical attention at >24-week follow-up) – ITT analysis (assessed with: child health service use – Injury requiring medical attention)											
731 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1,2]	reporting bias strongly suspected[3]	⊕⊕⊖⊖ **LOW**[1,2,3] due to imprecision, publication bias	138/548 (25.2%)	55/183 (30.1%)	**RR 1.19** (0.92 to 1.55)	**Study population**	
										252 per 1000	**48 more per 1000** (from 20 fewer to 139 more)
										Moderate	
										252 per 1000	**48 more per 1000** (from 20 fewer to 139 more)

Child injury long follow-up (injury requiring medical attention at >24-week follow-up) – by intervention (assessed with: child health service use – Injury requiring medical attention)										
596 (1 study) 78 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	reporting bias strongly suspected[3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to imprecision, publication bias	41/451 (9.1%)	12/145 (8.3%)	**RR 0.91** (0.49 to 1.68)	**Study population**
										91 per 1000 / **8 fewer per 1000** (from 46 fewer to 62 more)
										Moderate
										91 per 1000 / **8 fewer per 1000** (from 46 fewer to 62 more)

[1] Total number of events is less than 300 (a threshold rule-of-thumb)
[2] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)
[3] Paper omits data

1.3.128 Prevention of neglect or abuse of the infant: home visits versus TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**
							With control	**With Prevention of neglect or abuse of the infant: home visits versus TAU**		**Risk with control** / **Risk difference with Prevention of neglect or abuse of the infant: home visits versus TAU** (95% CI)
Child injury post-treatment (injury requiring medical attention at endpoint or first measurement) – ITT analysis (assessed with: medical record: child injuries requiring medical care)										
364 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	92/185 (49.7%)	86/179 (48%)	**RR 0.97** (0.78 to 1.19)	**Study population**
										497 per 1000 / **15 fewer per 1000** (from 109 fewer to 94 more)
										Moderate
										497 per 1000 / **15 fewer per 1000** (from 109 fewer to 94 more)

Child injury post-treatment (injury requiring medical attention at endpoint or first measurement) – available case analysis (assessed with: medical record: child injuries requiring medical care)											
268 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	44/137 (32.1%)	38/131 (29%)	**RR 0.9** (0.63 to 1.3)	**Study population**	
										321 per 1000	**32 fewer per 1000** (from 119 fewer to 96 more)
										Moderate	
										321 per 1000	**32 fewer per 1000** (from 119 fewer to 96 more)
Ingestion of poison post-treatment (incidence during trial measured at endpoint or first measurement) – available case analysis (assessed with: study-specific child health questionnaire: Ingestion of poison)											
138 (1 study) 52 weeks	serious[4]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3,4] due to risk of bias, imprecision	4/70 (5.7%)	0/68 (0%)	**RR 0.11** (0.01 to 2.08)	**Study population**	
										57 per 1000	**51 fewer per 1000** (from 57 fewer to 62 more)
										Moderate	
										57 per 1000	**51 fewer per 1000** (from 56 fewer to 62 more)
Child protective service reports (all types) post-treatment (substantiated reports during trial measured at endpoint or first measurement) – ITT analysis (assessed with: Child Protective Services' reports: substantiated reports of all types)											
364 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	61/185 (33%)	56/179 (31.3%)	**RR 0.95** (0.7 to 1.28)	**Study population**	
										330 per 1000	**16 fewer per 1000** (from 99 fewer to 92 more)
										Moderate	
										330 per 1000	**17 fewer per 1000** (from 99 fewer to 92 more)

Child protective service reports (all types) post-treatment (substantiated reports during trial measured at endpoint or first measurement) – available case analysis (assessed with: Child Protective Services' reports: substantiated reports of all types)											
297 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	26/150 (17.3%)	24/147 (16.3%)	**RR 0.94** (0.57 to 1.56)	**Study population**	
										173 per 1000	**10 fewer per 1000** (from 75 fewer to 97 more)
										Moderate	
										173 per 1000	**10 fewer per 1000** (from 74 fewer to 97 more)
Child protective service reports (neglect) post-treatment (substantiated reports during trial measured at endpoint or first measurement) – ITT analysis (assessed with: Child Protective Services' reports: substantiated reports of neglect)											
364 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	55/185 (29.7%)	50/179 (27.9%)	**RR 0.94** (0.68 to 1.3)	**Study population**	
										297 per 1000	**18 fewer per 1000** (from 95 fewer to 89 more)
										Moderate	
										297 per 1000	**18 fewer per 1000** (from 95 fewer to 89 more)
Child protective service reports (neglect) post-treatment (substantiated reports during trial measured at endpoint or first measurement) – available case analysis (assessed with: Child Protective Services' reports: substantiated reports of neglect)											
297 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision	20/150 (13.3%)	18/147 (12.2%)	**RR 0.92** (0.51 to 1.66)	**Study population**	
										133 per 1000	**11 fewer per 1000** (from 65 fewer to 88 more)
										Moderate	
										133 per 1000	**11 fewer per 1000** (from 65 fewer to 88 more)

Maternal use of punishment post-treatment (corporal/verbal punishment used anytime in past week measured at endpoint or first measurement) – ITT analysis (assessed with: Straus's parent–child Conflict Tactics Scale (CTS-PC): corporal/verbal punishment)											
364 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	146/185 (78.9%)	136/179 (76%)	**RR 0.96** (0.86 to 1.08)	**Study population**	
										789 per 1000	**32 fewer per 1000** (from 110 fewer to 63 more)
										Moderate	
										789 per 1000	**32 fewer per 1000** (from 110 fewer to 63 more)
Maternal use of punishment post-treatment (corporal/verbal punishment used anytime in past week measured at endpoint or first measurement) – available case analysis (assessed with: Straus's parent–child Conflict Tactics Scale (CTS-PC): corporal/verbal punishment)											
249 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	84/123 (68.3%)	83/126 (65.9%)	**RR 0.96** (0.81 to 1.15)	**Study population**	
										683 per 1000	**27 fewer per 1000** (from 130 fewer to 102 more)
										Moderate	
										683 per 1000	**27 fewer per 1000** (from 130 fewer to 102 more)
Potential for child abuse post-treatment (mean score at endpoint or first measurement) – available case analysis (measured with: Child Abuse Potential Inventory (CAPI); better indicated by lower values)											
124 (1 study) 78 weeks	serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	63	61	-		The mean potential for child abuse post-treatment (mean score at endpoint or first measurement) – available case analysis in the intervention groups was **0.36 standard deviations lower** (0.71 lower to 0 higher)

[1] High risk of selection bias due to unclear allocation concealment and statistically significant baseline differences in poor psychological resources (37% intervention group versus 50% control) and in prenatal enrolment (41% intervention group and 53% control)

[2] Total number of events is less than 300 (a threshold rule-of-thumb)

[3] 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD – 0.5/0.5 or RR 0.75/1.25)

[4] High risk of selection bias due to statistically significant baseline group differences in: parity (54% of intervention group primiparous versus 33% of control); identification as indigenous

Australian (9% of intervention versus 2% of control); mental illness of partner (3% of intervention versus 14% of control); history of postnatal depression (11% of intervention versus 28% of control); physical domestic abuse (2% of intervention versus 10% of control); potential for child abuse (mean CAPI score in intervention was 123 versus 159 in control, and elevated CAPI score for 12% of intervention group versus 30% of control group)

[5] Total population size is less than 400 (a threshold rule-of-thumb)

1.3.129 Optimal infant care: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU

Quality assessment							Summary of findings			
Participants (studies) Follow-up	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect** (95% CI)	**Anticipated absolute effects**
							With control	**With optimal infant care: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU**		**Risk with control**	**Risk difference with optimal infant care: structured psychological interventions (CBT or IPT) versus TAU/enhanced TAU** (95% CI)
Immunisation post-treatment (complete immunisation at endpoint or first measurement) – ITT analysis (assessed with: optimal infant care: complete immunisation)											
903 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⨁⨁⨁⨁ **HIGH**	294/440 (66.8%)	339/463 (73.2%)	**RR 1.1** (1.01 to 1.19)	**Study population**	
										668 per 1000	**67 more per 1000** (from 7 more to 127 more)
										Moderate	
										668 per 1000	**67 more per 1000** (from 7 more to 127 more)
Immunisation post-treatment (complete immunisation at endpoint or first measurement) – available case analysis (assessed with: optimal infant care: complete immunisation)											
705 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⨁⨁⨁⨁ **HIGH**	294/345 (85.2%)	339/360 (94.2%)	**RR 1.11** (1.05 to 1.16)	**Study population**	
										852 per 1000	**94 more per 1000** (from 43 more to 136 more)
										Moderate	

									852 per 1000	**94 more per 1000** (from 43 more to 136 more)

2 ECONOMIC EVIDENCE PROFILES

2.1 CASE IDENTIFICATION AND ASSESSMENT OF MENTAL HEALTH PROBLEMS IN PREGNANCY OR THE POSTNATAL PERIOD

2.1.1 PHQ-3 versus standard care case identification

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1]	Incremental effect	ICER (£/effect)	Uncertainty
Campbell et al, 2008 New Zealand	Potentially serious limitations[2]	Partially applicable[3]	Cost-effectiveness and cost-utility Measure of outcome: cases of depression detected; cases of depression resolved; QALYs Time horizon: 12 months	£1,083,600	7,420 cases of depression detected 5,330 cases of depression resolved 616 QALYs	£146 per case of depression detected £203 per case of depression resolved £1,759/QALY	Results sensitive to proportion of women that were identified with depression and that accessed and initiated treatment

1. Costs converted to UK pounds using purchasing power parities (PPP) exchange rates (http://www.oecd.org/std/ppp); all costs uplifted to 2013/2014 UK pounds using the UK HCHS inflation index
2. Effectiveness based on published sources and authors' assumptions; resource use based on national recommendations, international guidance, other published sources, expert opinion and authors' assumptions; utility values for general depression population treated with antidepressant medication; assumes that GPs will identify all cases correctly (that is, false positive rate associated with GP assessment is assumed to be zero)
3. Study conducted in New Zealand with healthcare system sufficiently similar to UK NHS; model heavily relies on the previous *Antenatal and postnatal Mental Health* guideline (NICE, 2007; NCCMH, 2007); QALYs as one of the outcomes; however standard was not very well defined

2.1.2 Formal case identification (BDI or EPDS) versus standard care case identification

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1]	Incremental effect	ICER (£/effect)	Uncertainty
Hewitt et al, 2009; Paulden et al, 2009 UK	Potentially serious limitations[2]	Directly applicable[3]	• Cost-utility • Time horizon: 12 months • Incremental costs and outcomes are relative to the next more expensive strategy (after excluding dominated or extendedly dominated strategies)	BDI (cut-off 10) £86 EPDS (cut-off 16-8) £29-197	BDI (cut-off 10) 0.0013 EPDS (cut-off 16-8) 0.0006-0.0017	>£40,000 for all formal identification strategies	False positives correctly diagnosed with single GP consultation as opposed to receiving 'additional care', EPDS (cut-off 10) ICER of £34,616/QALY when compared with SC; using EPDS (cut-off 13) with confirmatory structured clinical interview, ICER of £40,060/QALY when compared with SC; Whooley questions as identification method ICER of £55,197/QALY when compared with EPDS (cut-off 16); women with major depression only, ICER EPDS (cut-off 16) £27,511/QALY

1. All costs uplifted to 2013/2014 UK pounds using the UK HCHS inflation index
2. Effectiveness data based on meta-analysis of diagnostic studies and other published sources; resource use based on assumptions and other published sources; some unit costs derived from published studies; decision model doesn't adequately reflect the management of depression in the postnatal period in the UK (that is, no further assessment of positive cases considered, treatment of positives cases limited to intensive psychological therapy; assumption that no false positives were found following standard care case identification)
3. UK study; NHS and PSS perspective; QALYs as an outcome measure, however utility values are for general depression population treated with antidepressant medication

2.1.3 Formal case identification (Whooley, EPDS or PHQ-9) versus standard care case identification

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1] versus standard care	Incremental QALY versus standard care	ICER (£/QALY)	Uncertainty
Guideline economic analysis UK	Potentially serious limitations[2]	Directly applicable[3]	• Cost-utility • Time horizon: 12 months	*Per 1000 women* Whooley and PHQ-9: – £35,915 Whooley and EPDS: – £30,752 EPDS only: – £4,918	*Per 1000 women* Whooley and PHQ-9: 2.82 Whooley and EPDS: 2.93 EPDS only: 1.60	Whooley and EPDS versus Whooley and PHQ-9 ICER £45,593/QALY; EPDS only and standard care dominated	EPDS only or standard care case identification were never the preferred options. *ICER of Whooley and EPDS versus Whooley and PHQ-9* ICER was sensitive to diagnostic characteristics associated with EPDS and PHQ-9; the model was robust to other inputs including prevalence of depression, proportion of moderate to severe depression, treatment relative risks, costs associated with false positives, treatment costs; whether assessment was performed by GP or HV; whether standard care case identification was performed by GP or HV.

1. All costs uplifted to 2013/2014 UK pounds using the UK HCHS inflation index
2. Effectiveness data for EDPS taken from guideline meta-analysis of diagnostic studies, however for PHQ-9 and Whooley questions only single studies were available; PHQ-9 study reporting diagnostic characteristics was for antenatal population; sensitivity and specificity of first and second (that is, subsequent tool) was assumed to be independent of each other; resource use based on published data and GDG expert opinion; national unit costs used; deterministic sensitivity analysis, PSA not possible
3. NHS and PSS perspective, QALYs based on EQ-5D UK tariff; utility data taken from general population with depression and not from women with depression in postnatal period

2.2 PSYCHOLOGICAL AND PSYCHOSOCIAL INTERVENTIONS FOR THE PREVENTION OF DEVELOPING MENTAL HEALTH PROBLEMS IN PREGNANCY OR THE POSTNATAL PERIOD

2.2.1 Home visiting versus standard care

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1]	Incremental effect	ICER (£/effect)	Uncertainty
Aracena et al, 2009 Chile	Potentially serious limitations[2]	Partially applicable[3]	• Cost-effectiveness • Measure of outcome: Goldberg's depression scale score • Time horizon: 15 months	£30.9	-2.91	£10.4	None reported, but benefit significantly higher for intervention
Barlow et al, 2007; McIntosh et al, 2009 UK	Minor limitations[4]	Partially applicable[5]	• Cost-effectiveness • Measure of outcome: proportion of infants identified as being ill-treated; improvement in maternal sensitivity and infant cooperativeness CARE index scores; time exposed to abuse and neglect • Time horizon: 18 months; 5 years when time exposed to abuse and neglect used as an outcome	£3,110 from healthcare payer perspective	0.059 proportion of infants being ill treated 1.07 maternal sensitivity index 1.43 infant co-operativeness index -1.92 months	From healthcare payer perspective: £52,718 per infant identifed as being ill treated £2,871 per extra unit of improvement on maternal sensitivity index £2,136 per extra unit of improvement in infant co-operativeness index £1,229 for a reduction in infant exposure to abuse and neglect by one month	From healthcare payer perspective: at WTP of £18,320 per unit improvement on maternal sensitivity index probability of intervention being cost effective was 0.95; at WTP of £3,558 per unit improvement on infant cooperativeness index probability that intervention was cost effective was 0.95

1. In non-UK studies costs converted to UK pounds using purchasing power parities (PPP) exchange rates (http://www.oecd.org/std/ppp); all costs uplifted to 2013/2014 UK pounds using the UK HCHS inflation index
2. Effectiveness based on one RCT (n=90); not clear what type of healthcare costs were included; resource use estimates from registries of health centres; source of unit costs unclear; the use of Goldberg's depression scale as a primary outcome may mean that other important aspects of HRQoL may not be captured
3. Study conducted in Chile; healthcare payer perspective; non-QALY outcome; standard care may not be representative of routine and best practice in the NHS

4. Effectiveness based on one RCT (n=131); some of the resource use from published sources; a mixture of national and local unit costs
5. UK study; non-QALY outcome; base-case analysis from societal perspective but also reports costs from healthcare perspective; unclear if analysis from healthcare perspective includes all relevant costs to NHS and PSS

2.2.2 Infant sleep training intervention versus standard care

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1]	Incremental effect	ICER (£/effect)	Uncertainty
Hiscock et al, 2007 Australia	Minor limitations[2]	Partially applicable[3]	Cost-effectiveness Measure of outcome: percent of mothers reporting infant sleep problem; depression symptoms (mesured using EPDS); SF-12 mental health domain scores Time horizon: 12 months	-£10.45	-16% per cent of mothers reporting infant sleep problem -1.7 reduction in EPDS score 3.9 point improvement on SF-12 mental health domain	Intervention dominant	Difference of – 16% of mothers reporting infant sleep problem (p = 0.004); difference of – 1.7 points in EPDS scores (p = 0.001); 3.9 point improvement on SF-12 mental health domain scores (p < 0.001); reduction in costs of £10.45 (p = 0.55)

1. Costs converted to UK pounds using purchasing power parities (PPP) exchange rates (http://www.oecd.org/std/ppp); all costs uplifted to 2013/2014 UK pounds using the UK HCHS inflation index
2. Source of unit costs unclear
3. Australian study with healthcare system sufficiently similar to UK NHS; non-QALY outcome however intervention was dominant; healthcare perspective plus informal care
1. All costs uplifted to 2013/2014 UK pounds using the UK HCHS inflation index
2. A mix of local and national unit costs
3. UK study; non-QALY outcome; includes cost categories not relevant to NHS and PSS perspective (that is, informal care); discount rate of 6% for costs and 1.5% for health effects

2.2.3 Listening visits versus standard care

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1]	Incremental effect	ICER (£/effect)	Uncertainty

Petrou et al, 2006 UK	Minor limitations[2]	Partially applicable[3]	• Cost-effectiveness • Measure of outcome: number of months in depression avoided • Time horizon: 18 months	£179	0.49	£365	Community service utilisation increased by 10-30%, ICER ranged from £632-1,170; per diem cost for inpatient care ±20%, ICER ranged from £62-669; discount rate for cost and health effects ranged from 0-10%, ICER ranged from £526-296; discount rate for costs and health effects 3%, ICER £453; at WTP of £1,000 and £2,000 per additional month of depression avoided, probability intervention being cost effective was 0.71 and 0.77, respectively.

1. All costs uplifted to 2013/2014 UK pounds using the UK HCHS inflation index
2. A mix of local and national unit costs
3. UK study; non-QALY outcome; includes cost categories not relevant to NHS and PSS perspective (that is, informal care); discount rate of 6% for costs and 1.5% for health effects

2.3 PSYCHOLOGICAL AND PSYCHOSOCIAL INTERVENTIONS FOR THE TREATMENT OF MENTAL HEALTH PROBLEMS IN PREGNANCY OR THE POSTNATAL PERIOD

2.3.1 Social support versus standard care

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1]	Incremental effect	ICER (£/effect)	Uncertainty
Dukhovny et al, 2013 Canada	Potentially serious limitations[2]	Partially applicable[3]	• Cost-effectiveness • Measure of outcome: cases with EPDS score ≤12 • Time horizon: 12 weeks	£361 from healthcare payer perspective £658 from societal perspective	0.1116	£3,286 from healthcare payer perspective £5,892 from societal perspective	From societal perspective: as healthcare visits are varied between 50-400%, ICER ranges from £5,693 to £5,363; ICER sensitive to cost of running programme, volunteer time, family/friend and partner work absence; at WTP per case with EPDS score ≤12 of £11,889, probability intervention CE is 0.95

1. Costs converted to UK pounds using purchasing power parities (PPP) exchange rates (http://www.oecd.org/std/ppp); all costs uplifted to 2013/2014 UK pounds using the UK HCHS inflation index
2. Time horizon only 12 weeks which may not be sufficiently long to reflect all important differences in costs and outcomes; a mixture of national and local unit costs; sensitivity analysis only reported from societal perspective
3. Canadian study (healthcare system sufficiently similar to UK NHS); non-QALY outcome; main analysis conducted from societal perspective, but also analysis considering costs from a healthcare perspective included

2.3.2 Structured psychological therapy, listening visits and standard care

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1]	Incremental QALY	ICER (£/QALY)	Uncertainty
Hewitt et al, 2009; Paulden et al, 2009 UK	Minor limitations[2]	Directly applicable[3]	• Cost-utility • Time horizon: 12 months	*Versus standard care:* Structured psychological therapy: £939 Listening visits: £1,123 Listening visits versus structured psychological therapy: £154	*Versus standard care:* Structured psychological therapy: 0.05 Listening visits: 0.0477 Listening visits versus structured psychological therapy: 0.0024	*Versus standard care:* Structured psychological therapy: £20,732 Listening visits: £23,534 Listening visits versus structured psychological therapy £78,606	*Structured psychological therapy versus standard care* At cost per QALY of £20,000-£30,000 probability structured psychological therapy is CE is 0.504-0.549; however this probability comes from the comparison of 3 options. The probability would be higher if only two of the options were compared. *Listening visits versus standard care* The ICER was estimated based on the data reported in the publication. Sensitivity analysis was not relevant in this comparison because the intervention was not cost-effective. *Listening visits versus structured psychological therapy* At the cost per QALY of £20,000-£30,000 probability listening home visits CE is 0.276-0.414; however this probability comes from the comparison of 3 options. The probability would be expected to be higher if only two of the options were compared.

1. All costs uplifted to 2013/2014 UK pounds using the UK HCHS inflation index
2. Some of resource use informed by expert opinion; costs associated with infant care excluded; the relative effect between listening visits and structured psychological therapy was based on indirect comparisons between treatments, using standard care as the baseline common comparator, due to lack of head-to-head comparisons between the two interventions
3. UK study; NHS and PSS perspective; QALYs used as an outcome; however utility scores are relevant to the general depression population treated with antidepressant medication

2.3.3 Listening visits versus structured psychological therapy; structured psychological therapy based on person-centred approach (PCA) versus cognitive behavioural approach (CBA)

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1]	Incremental QALY	ICER (£/QALY)	Uncertainty
Morrell et al, 2009 UK	Minor limitations[2]	Directly applicable[3]	• Cost-utility • Time horizon: 6 and 12 months	*Structured psychological therapy versus standard care:* -£59 at 6 months -£127 at 12 months *PCA versus CBA* £32 at 6 months No difference at 12 months	*Structured psychological therapy versus standard care:* 0.004 at 6 months 0.025 at 12 months *PCA versus CBA* -0.002 at 6 months No difference at 12 months	*Structured psychological therapy versus standard care:* Intervention dominant at 6 and 12 months *PCA versus CBA* CBA dominant at 6 months No difference between CBA and PCA at 12 months	*Structured psychological therapy versus standard care:* At WTP of £20,000-£30,000/QALY probability of intervention being cost effective is >0.70 and >0.80 at 6 and 12 months, respectively. *PCA versus CBA* At WTP of £20,000-£30,000/QALY probability of CBA being cost effective was >0.70 at 6 months. However, PSA included SC (that is, three comparators); if only two comparators were included this probability would be expected to be higher.

1. All costs uplifted to 2013/2014 UK pounds using the UK HCHS inflation index
2. Some of resource use estimates informed by expert opinion and authors' assumptions; high attrition rate in RCT may have resulted in analysis being underpowered to detect differences between treatments
3. UK study; NHS and PSS perspective; QALYs used as an outcome (utility values derived using mapping technique)

2.3.4 CBT-informed psychoeducation versus standard care

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1]	Incremental QALY	ICER (£/QALY)	Uncertainty
Stevenson et al, 2010 (A); Stevenson et al, 2010 (B) UK	Potentially serious limitations[2]	Directly applicable[3]	• Cost-utility • Time horizon: 12 months	£1,729	0.032	£53,563	Cost of intervention per woman decreased to £865 ICER of £26,781/QALY; increased to £2,306 ICER of £71,416/QALY; lower estimate of efficacy ICER of £65,280/QALY; upper estimate of efficacy ICER £45,515/QALY; linear decline in advantage of intervention extended to 18 months ICER of £39,637/QALY; additional QALY gain of 0.02 assumed, ICER of £33,255/QALY; when cost of intervention per woman decreased to £1,112, EPDS decrease of 4.3 assumed, and linear decline in advantage extended to 18 months ICER of £22,169/QALY.

1. All costs uplifted to 2013/2014 UK pounds using the UK HCHS inflation index
2. Effectiveness derived from a small RCT (n=45) and extrapolated to 12 months using conceptual model based on authors' assumptions; some of resource use estimates informed by expert opinion and authors' assumptions; hasn't included additional running costs associated with intervention (that is, room hire and crèche facilities); some unit costs derived from RCT
3. UK study; NHS and PSS perspective; QALYs used as an outcome (utility values derived using mapping technique)

2.3.5 Facilitated guided self-help, listening visits, and standard care

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1] versus standard care	Incremental QALY versus standard care	ICER (£/QALY)	Uncertainty
Guideline economic analysis UK	Minor limitations[2]	Directly applicable[3]	• Cost-utility • Time horizon: 12 months	*Per woman versus standard care* Facilitated guided self-help: £179 Listening visits: £490 Facilitated guided self-help versus listening visits:-£311	*Per woman versus standard care* Facilitated guided self-help: 0.014 Listening visits: 0.0021 Facilitated guided self-help versus listening visits: 0.012	*versus standard care* ICER of facilitated guided self-help £12,675/QALY. ICER of listening visits £233,912/QALY. Facilitated guided self-help versus listening visits: facilitated guided self-help dominant	*Listening visits versus standard care* were never the preferred treatment option *ICER of facilitated guided self-help versus standard care* Utility score associated with subthreshold/minor to moderate depression varied from 0.5 to 0.7 ICER £4,225-£9,506/QALY; cost of providing facilitated guided self-help varied from £100-£300 ICER £3,845-£17,982/QALY; absolute risk of no improvement varied from 0.5-0.8 ICER £16,158-£8,890/QALY. At WTP of £20,000-£30,000/QALY probability of facilitated guided self-help being cost effective was 0.59-0.72. For facilitated guided self-help versus listening visits: sensitivity analysis not undertaken for this specific comparison, as listening visits was not cost effective among the options assessed, and thus this comparison was not relevant.

1. All costs uplifted to 2013/2014 UK pounds using the UK HCHS inflation index
2. Effectiveness data taken from guideline meta-analysis; resource use based on published data and GDG expert opinion; national unit costs used; probabilistic and deterministic sensitivity analysis
3. NHS and PSS perspective, QALYs based on EQ-5D UK tariff