APPENDIX 19: GRADE EVIDENCE PROFILES

1 CLINICAL EVIDENCE PROFILES

1.1 PSYCHOSOCIAL INTERVENTIONS AIMED AT CORE FEATURES OF AUTISM (OVERALL AUTISTIC BEHAVIOURS)

1.1.1 Behavioural interventions aimed at overall autistic behaviours as an indirect outcome

Early Start Denver Model versus treatment-as-usual for overall autistic behaviours as an indirect outcome

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 Treatment-as-usual	With Behavioural intervention		Risk with Treatment-as-usual	Risk difference with Behavioural intervention (95% CI)
Overall autistic behaviours (measured with: Autism Diagnostic Observation Schedule (ADOS/ADOS-G): Standardised severity score; Better indicated by lower values)											
45 (1 study) 104 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	21	24	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.16 standard deviations lower** (0.75 lower to 0.43 higher)
DSM-IV Clinical Diagnosis (assessed with: Number of participants who showed improvement in diagnosis from autistic disorder to PDD-NOS)											
45 (1 study) 104 weeks	serious[2]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	1/21 (4.8%)	7/24 (29.2%)	**RR 8.24** (0.92 to 73.79)	**Study population**	
										48 per 1000	**345 more per 1000** (from 4 fewer to 1000 more)

									Moderate	
									N/A	N/A

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear/unknown as blinding of outcome assessment is unclear
[3] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)

1.1.2 Educational interventions aimed at overall autistic behaviours as a direct outcome

COMPASS versus treatment-as-usual for overall autistic behaviours as a direct outcome

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 Treatment-as-usual	**With Teacher consultation and training according to the collaborative model for promoting competence and success (COMPASS)**		**Risk with Treatment-as-usual**	**Risk difference with Teacher consultation and training according to the collaborative model for promoting competence and success (COMPASS)** (95% CI)
IEP goal attainment for targeted objectives (measured with: Behavioural observation; Better indicated by lower values)											
32 (1 study) 39 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to risk of bias, imprecision	15	17	N/A	N/A	The mean IEP goal attainment for targeted objectives in the intervention groups was **1.42 standard deviations higher** (0.63 to 2.2 higher)

[1] High risk of performance bias as intervention administrators were non-blind. There was also a high risk of detection bias as the primary outcome assessor was the non-blind investigator with a blinded secondary outcome assessor only rating 20% of behavioural observations. In addition, because only 20% of observations were double-coded and a standardized observation measure was not used the relaibility and validity of this outcome measure is unclear
[2] N<400

LEAP training versus manual-only control for overall autistic behaviours as a direct outcome

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 Intervention-manual-only control	**With Inclusive educational intervention (LEAP) training**		**Risk with Intervention-manual-only control**	**Risk difference with Inclusive educational intervention (LEAP) training** (95% CI)
Overall autistic behaviours (measured with: Childhood Autism Rating Scale (CARS): Total; Better indicated by lower values)											
294 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to risk of bias, imprecision	117	177	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.42 standard deviations lower** (0.66 to 0.19 lower)

[1] High risk of performance and response bias as intervention administrators and participants non-blind. In addition, risk of detection bias is unclear/unknown as identity and blinding of outcome assessors not reported
[2] N<400

1.1.3 Parent training interventions aimed at overall autistic behaviours as a direct or indirect outcome

Parent training versus treatment-as-usual for overall autistic behaviours as an indirect outcome

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 Parent training versus treatment-as-usual for overall autistic behaviours as an indirect outcome		Risk with Control	Risk difference with Parent training versus treatment-as-usual for overall autistic behaviours as an indirect outcome (95% CI)
Overall autistic behaviours (PEC+PEBM combined) (measured with: Developmental Behaviour Checklist (DBC): Autism Screening Algorithm (ASA); Better indicated by lower values)											
103 (1 study) 46 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	35	68	N/A	N/A	The mean overall autistic behaviours (PEC+PEBM combined) in the intervention groups was **0.06 standard deviations lower** (0.47 lower to 0.34 higher)
Overall autistic behaviours (measured with: Childhood Autism Rating Scale (CARS): Total; Better indicated by lower values)											
102 (2 studies) 13-46 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	51	51	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.42 standard deviations lower** (0.81 to 0.03 lower)
Overall autistic behaviours (PEBM group) (measured with: Childhood Autism Rating Scale (CARS): Total; Better indicated by lower values)											
70 (1 study) 46 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[3,4] due to imprecision,	35	35	N/A	N/A	The mean overall autistic behaviours (PEBM group) in the intervention groups was

						publication bias					**0.44 standard deviations lower** (0.92 lower to 0.03 higher)
Overall autistic behaviours (measured with: Childhood Autism Rating Scale (CARS): Total; Better indicated by lower values)											
32 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3] due to imprecision	16	16	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.38 standard deviations lower** (1.08 lower to 0.32 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as parent-rated and parents were non-blind and involved in the intervention
[2] N<400
[3] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[4] Risk of selective reporting bias in TONGE2006/2012 as trial protocol is not registered on ClinicalTrials.gov or ISRCTN and there is a potential conflict of interest as the manuals used in this study have been published by Jessica Kingsley Publishers, and the authors receive royalties (5%) from sales

Parent and day-care staff training versus standard day-care for overall autistic behaviours as a direct outcome

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 Standard day-care	**With Parent and day-care staff training**		**Risk with Standard day-care**	**Risk difference with Parent and day-care staff training** (95% CI)
Overall autistic behaviour (measured with: Autism Behaviour Checklist (ABC): Total; Better indicated by lower values)											
35 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	19	16	N/A	N/A	The mean overall autistic behaviour in the intervention groups was **0.4 standard deviations lower** (1.08 lower to 0.27 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.1.4 Social-communication interventions aimed at overall autistic behaviours as an indirect outcome

Child's Talk versus treatment-as-usual for overall autistic behaviours as an indirect outcome

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 Treatment-as-usual	With Caregiver-mediated social-communication intervention (Child's Talk)		Risk with Treatment-as-usual	Risk difference with Caregiver-mediated social-communication intervention (Child's Talk) (95% CI)
Overall autistic behaviours (measured with: Autism Diagnostic Observation Schedule (ADOS/ADOS-G): Total score; Better indicated by lower values)											
28 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	14	14	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.76 standard deviations lower** (1.53 lower to 0.01 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.2 PSYCHOSOCIAL INTERVENTIONS AIMED AT THE CORE AUTISM FEATURE OF IMPAIRED RECIPROCAL SOCIAL COMMUNICATION AND INTERACTION

1.2.1 AAC intervention aimed at the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

PECS training for teachers versus treatment-as-usual for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

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 No treatment	**With Picture Exchange Communication System (PECS) training for teachers**		**Risk with No treatment**	**Risk difference with Picture Exchange Communication System (PECS) training for teachers** (95% CI)
Communication (assessed with: Autism Diagnostic Observation Schedule (ADOS/ADOS-G): Communication (odds of being in a higher severity category on ADOS-G))											
84 (1 study) 33 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	N/A	N/A	**OR 0.52** (0.24 to 1.12)	**Study population**	
										N/A	N/A
										Moderate	
										0 per 1000	N/A
Social interaction (assessed with: Autism Diagnostic Observation Schedule (ADOS/ADOS-G): Social Interaction (odds of being in a higher severity category on ADOS-G))											
84 (1 study) 33 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	N/A	N/A	**OR 0.55** (0.25 to 1.2)	**Study population**	
										N/A	N/A
										Moderate	

									0 per 1000	N/A
Social interaction (assessed with: Autism Diagnostic Observation Schedule (ADOS/ADOS-G): Social Interaction (odds of being in a higher severity category on ADOS-G))										
53 (1 study) 78 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	N/A	N/A	**OR 0.28** (0.09 to 0.88)	**Study population**
										N/A N/A
										Moderate
										0 per 1000 N/A

[1] High risk of performance, response and detection bias as intervention administrators, participants and outcome assessors were non-blind
[2] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm
[3] Events<300

1.2.2 Animal-based intervention aimed at the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

Horseback riding versus waitlist control for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

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 Waitlist control	**With Horseback riding**		**Risk with Waitlist control**	**Risk difference with Horseback riding** (95% CI)
Social impairment (measured with: Social Responsiveness Scale (SRS): Total; Better indicated by lower values)											
34 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision,	15	19	N/A	N/A	The mean social impairment in the intervention groups was **0.73 standard**

						publication bias				**deviations lower** (1.43 to 0.03 lower)
Social cognition (measured with: Social Responsiveness Scale (SRS): Social Cognition ; Better indicated by lower values)										
34 (1 study) 12 weeks	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	15	19	N/A	N/A	The mean social cognition in the intervention groups was **0.44 standard deviations lower** (1.13 lower to 0.24 higher)
Social awareness (measured with: Social Responsiveness Scale (SRS): Social Awareness ; Better indicated by lower values)										
34 (1 study) 12 weeks	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	15	19	N/A	N/A	The mean social awareness in the intervention groups was **0.4 standard deviations lower** (1.08 lower to 0.28 higher)
Social motivation (measured with: Social Responsiveness Scale (SRS): Social Motivation ; Better indicated by lower values)										
34 (1 study) 12 weeks	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	15	19	N/A	N/A	The mean social motivation in the intervention groups was **0.58 standard deviations lower** (1.27 lower to 0.12 higher)

[1] High risk of performance and response bias as intervention administrators and participants non-blind. There is also a high risk of detection bias as outcome measures are parent-rated and parents non-blind
[2] N<400
[3] High risk of selective reporting bias as data not reported for selected subscales: the social communication and autistic mannerisms subscales of the Social Responsiveness Scale (SRS)
[4] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.2.3 Arts-based intervention aimed at the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

RMT versus waitlist control for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

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 Waitlist control	With Music therapy		Risk with Waitlist control	Risk difference with Music therapy (95% CI)
Social communication (measured with: Childhood Autism Rating Scale (CARS): Social communication (composite score from imitation, verbal and non-verbal communication, consistency of intellectual responses and general impressions); Better indicated by lower values)											
24 (1 study) 30 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	12	12	N/A	N/A	The mean social communication in the intervention groups was **0.23 standard deviations higher** (0.58 lower to 1.03 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.2.4 Behavioural intervention aimed at the core autism feature of impaired reciprocal social communication and interaction as a direct or indirect outcome

RIT versus treatment-as-usual for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

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 Behaviour-focused intervention versus treatment-as-usual for the core autism		Risk with Control	Risk difference with Behaviour-focused intervention versus treatment-as-usual for the core

							feature of impaired reciprocal social communication and interaction as direct outcome				**autism feature of impaired reciprocal social communication and interaction as direct outcome** (95% CI)
Examiner-child joint/shared attention (measured with: EScs (Early Social Communication Scales): Initiating Joint Attention (IJA); Better indicated by lower values)											
27 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	13	14	N/A	N/A	The mean examiner-child joint/shared attention in the intervention groups was **0.89 standard deviations higher** (0.09 to 1.68 higher)
Examiner-child joint/shared attention (Copy) (measured with: EScs (Early Social Communication Scales): Initiating Joint Attention (IJA); Better indicated by lower values)											
27 (1 study) 23 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	13	14	N/A	N/A	The mean examiner-child joint/shared attention (copy) in the intervention groups was **0.86 standard deviations higher** (0.06 to 1.65 higher)
Social and emotional development (measured with: Bayley Scales of Infant Development: Social-Emotional ; Better indicated by lower values)											
27 (1 study) 23 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to risk of bias, imprecision	13	14	N/A	N/A	The mean social and emotional development in the intervention groups was **0.41 standard deviations higher** (0.36 lower to 1.17 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and the risk of detection bias is also high as outcome assessors were not blinded
[2] N<400
[3] High risk of performance and response bias as intervention administrators and participants were non-blind, and the risk of detection bias is also high as parent-report measue and parents non-blind
[4] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

P-ESDM versus treatment-as-usual for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

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 Treatment-as-usual	With Parent-mediated and brief Early Start Denver Model (P-ESDM)		Risk with Treatment-as-usual	Risk difference with Parent-mediated and brief Early Start Denver Model (P-ESDM) (95% CI)
Social affect (measured with: Autism Diagnostic Observation Schedule for Toddlers (ADOS-T): Social Affect; Better indicated by lower values)											
98 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	49	49	N/A	N/A	The mean social affect in the intervention groups was **0.07 standard deviations lower** (0.46 lower to 0.33 higher)
Imitation (measured with: Imitation tasks (Rogers et al., 2003): Imitative sequences; Better indicated by lower values)											
98 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to risk of bias, imprecision	49	49	N/A	N/A	The mean imitation in the intervention groups was **0.24 standard deviations higher** (0.16 lower to 0.63 higher)
Orienting to social stimuli (measured with: Social engagement task (Dawson et al., 2004): Mean Social Orient I; Better indicated by lower values)											
98 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to risk of bias, imprecision	49	49	N/A	N/A	The mean orienting to social stimuli in the intervention groups was **0.13 standard deviations higher** (0.27 lower to 0.52 higher)
Orienting to joint attention (measured with: Social engagement task (Dawson et al., 2004): Mean Orient to Joint Attention; Better indicated by lower values)											
98 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,3] due to risk of	49	49	N/A	N/A	The mean orienting to joint attention in the intervention groups was

						bias, imprecision				**0 standard deviations higher** (0.4 lower to 0.4 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear/unknown as outcome assessor reported only as 'laboratory personnel' with no information about blinding
[2] N<400
[3] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear/unknown as the identity and blinding of outcome assessors not reported and reliability and validity of outcome measure unclear
[4] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.2.5 Cognitive interventions aimed at the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

ERT versus treatment-as-usual for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

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 Emotion recognition training versus treatment-as-usual for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome**		**Risk with Control**	**Risk difference with Emotion recognition training versus treatment-as-usual for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome** (95% CI)
Emotion recognition (measured with: Assessment of Perception of Emotion from Facial Expression (Spence, 1995) or Situation-Facial Expression Matching (SEM): Distant generalization (study-specific) or Ekman emotion recognition photographs (Ekman & Friesen, 1976); Better indicated by lower values)											
119 (3 studies) 4-8 weeks	serious[1]	very serious[2]	no serious indirectness	serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, inconsistency, imprecision	53	66	N/A	N/A	The mean emotion recognition in the intervention groups was **0.65 standard deviations higher** (0.27 to 1.03 higher)

Recognising emotion from posture (measured with: Assessment of Perception of Emotion from Posture Cues (Spence, 1995); Better indicated by lower values)										
49 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,4] due to risk of bias, imprecision	23 26	N/A	N/A	The mean recognising emotion from posture in the intervention groups was **0.17 standard deviations higher** (0.4 lower to 0.73 higher)
Emotion understanding (measured with: Emotional vocabulary (study-specific); Better indicated by lower values)										
38 (1 study) 4 weeks	serious[5]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3,5] due to risk of bias, imprecision	18 20	N/A	N/A	The mean emotion understanding in the intervention groups was **1.02 standard deviations higher** (0.34 to 1.7 higher)
Emotion regulation and social skills (measured with: Emotion Regulation and Social Skills Questionnaire (ERSSQ; study-specific): Total; Better indicated by lower values)										
49 (1 study) 8 weeks	serious[6]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3,6] due to risk of bias, imprecision	23 26	N/A	N/A	The mean emotion regulation and social skills in the intervention groups was **1.39 standard deviations higher** (0.76 to 2.02 higher)
Anxiety coping skills (measured with: James and the Maths Test (Attwood, 2004); Better indicated by lower values)										
49 (1 study) 8 weeks	serious[7]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3,7] due to risk of bias, imprecision	23 26	N/A	N/A	The mean anxiety coping skills in the intervention groups was **1.23 standard deviations higher** (0.62 to 1.85 higher)
Bullying coping skills (measured with: Dylan is Being Teased (Attwood, 2004); Better indicated by lower values)										
49 (1 study) 8 weeks	serious[7]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3,7] due to risk of bias, imprecision	23 26	N/A	N/A	The mean bullying coping skills in the intervention groups was **1.29 standard deviations**

										higher (0.67 to 1.91 higher)
Social skills (measured with: Social Skills Questionnaire (Spence, 1995): Total ; Better indicated by lower values)										
49 (1 study) 8 weeks	serious[8]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3,8] due to risk of bias, imprecision	23	26	N/A	N/A	The mean social skills in the intervention groups was **1.42 standard deviations higher** (0.79 to 2.05 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind and risk of detection bias is unclear/unknown as blinding of outcome assessors is unclear
[2] Substantial to considerable heterogeneity (I-squared value of 77%, p = 0.01)
[3] N<400
[4] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[5] High risk of performance and response bias as intervention adminsitrators and participants were non-blind and high risk of detection bias as outcome assessor was non-blind investigator and study-specific outcome measure with no independent measures of reliability or validity data
[6] High risk of performance and response bias as intervention adminsitrators and participants were non-blind and high risk of detection bias as outcome assessor was non-blind parent and study-specific outcome measure with no independent measures of reliability or validity data
[7] High risk of performance and response bias as intervention adminsitrators and participants were non-blind and high risk of detection bias as only 33% of responses were independently and blindly coded
[8] High risk of performance, response and detection bias. The questionnaire was parent-rated and parents were not blind and participated in the intervention

FRT versus waitlist control for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

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 Face recognition training versus treatment-as-usual for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome**		**Risk with Control**	**Risk difference with Face recognition training versus treatment-as-usual for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome** (95% CI)
Face recognition (measured with: The Let's Face It! Skills Battery: Matching identity across masked features (percent correct); Better indicated by lower values)											

78 (1 study) 19 weeks	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	41	N/A	N/A	The mean face recognition in the intervention groups was **0.07 standard deviations lower** (0.52 lower to 0.37 higher)
Face recognition (measured with: The Let's Face It! Skills Battery: Featural and configural face dimensions (percent correct); Better indicated by lower values)											
78 (1 study) 19 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	37	41	N/A	N/A	The mean face recognition in the intervention groups was **0.02 standard deviations lower** (0.47 lower to 0.42 higher)
Face recognition (measured with: The Let's Face It! Skills Battery: Matching identity across expression (percent correct); Better indicated by lower values)											
79 (1 study) 19 weeks	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	42	N/A	N/A	The mean face recognition in the intervention groups was **0.43 standard deviations lower** (0.88 lower to 0.02 higher)
Face recognition (measured with: The Let's Face It! Skills Battery: Parts/whole identity (percent correct); Better indicated by lower values)											
77 (1 study) 19 weeks	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	36	41	N/A	N/A	The mean face recognition in the intervention groups was **0.06 standard deviations higher** (0.39 lower to 0.51 higher)
Face recognition (measured with: The Let's Face It! Skills Battery: Immediate memory for faces (percent correct); Better indicated by lower values)											
77 (1 study) 19 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias,	36	41	N/A	N/A	The mean face recognition in the intervention groups was **0.26 standard deviations**

						imprecision				**lower** (0.71 lower to 0.19 higher)

[1] High risk of performance and response bias as intervention administrator and participants non-blind, and risk of detection bias unclear/unknown as identity and blinding of outcome assessors not reported and no independent reliability or validity data for outcome measure
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] The paper states that other experimental measures were taken that are not reported
[4] N<400

ToM versus waitlist control for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

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 **With Theory of Mind training versus waitlist control for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome**		**Risk with**	**Risk difference with Theory of Mind training versus waitlist control for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome** (95% CI)
Theory of mind (measured with: Theory of Mind (ToM) Test: Total; Better indicated by lower values)										
36 (1 study) 16 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	17	19	N/A	N/A	The mean theory of mind in the intervention groups was **0.04 standard deviations higher** (0.61 lower to 0.7 higher)
Empathy (measured with: Index of Empathy for Children and Adolescents: Total; Better indicated by lower values)										
36 (1 study) 16 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	17	19	N/A	N/A	The mean empathy in the intervention groups was **0.17 standard deviations lower** (0.82 lower to 0.49 higher)
Emotional awareness (measured with: Levels of Emotional Awareness Scale for Children (LEAS-C): Total; Better indicated by lower values)										
36 (1 study) 16 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	17	19	N/A	N/A	The mean emotional awareness in the intervention groups was **0.46 standard deviations higher** (0.2 lower to 1.13 higher)
Maladaptive social behaviour (measured with: Children's Social Behavior Questionnaire (CSBQ): Total; Better indicated by lower values)										
36	serious[4]	no serious	no serious	very	undetected	⊕⊖⊖⊖	17	19	N/A	N/A	The mean maladaptive social

(1 study) 16 weeks		inconsistency	indirectness	serious[2]		**VERY LOW**[2,4] due to risk of bias, imprecision				behaviour in the intervention groups was **0.31 standard deviations lower** (0.97 lower to 0.35 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear/unknown as identity and blinding of outcome assessor not reported
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] High risk of performance and response bias as intervention administrators and participants non-blind, and high risk of detection bias as self-completed
[4] High risk of performance and response bias as intervention administrators and participants non-blind, and high risk of detection bias as parent-completed and parents non-blind

Computer-based ERT versus software training (attention-placebo) for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

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 Emotion recognition training (computer-based) versus attention-placebo (computer software training) for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome**		**Risk with Control**	**Risk difference with Emotion recognition training (computer-based) versus attention-placebo (computer software training) for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome** (95% CI)
Emotion recognition (IQ <70 and >70 combined) (measured with: Ekman emotion recognition photographs; Better indicated by lower values)											
49 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	25	24	N/A	N/A	The mean emotion recognition (iq <70 and >70 combined) in the intervention groups was **0.96 standard deviations higher** (0.37 to 1.56 higher)
Emotion recognition (IQ <70 and >70 combined) (measured with: Study-specific emotion recognition in drawings test; Better indicated by lower values)											
49	serious[3]	no serious	no serious	serious[2]	undetected	⊕⊕⊖⊖	25	24	N/A	N/A	The mean emotion

(1 study) 8 weeks		inconsistency	indirectness			**LOW**[2,3] due to risk of bias, imprecision				recognition (iq <70 and >70 combined) in the intervention groups was **1.1 standard deviations higher** (0.5 to 1.7 higher)
Emotion recognition (IQ <70 and >70 combined) (measured with: Composite score from Ekman emotion recognition photographs and study-specific emotion recognition from drawings test; Better indicated by lower values)										
49 (1 study) 8 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,3] due to risk of bias, imprecision	25	24	N/A	N/A	The mean emotion recognition (iq <70 and >70 combined) in the intervention groups was **1.09 standard deviations higher** (0.48 to 1.69 higher)
Face recognition (IQ <70 and >70 combined) (measured with: Benton Facial Recognition Test: Short Form; Better indicated by lower values)										
49 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	25	24	N/A	N/A	The mean face recognition (iq <70 and >70 combined) in the intervention groups was **0.88 standard deviations higher** (0.29 to 1.47 higher)
Face recognition (IQ <70 and >70 combined) (measured with: Benton Facial Recognition Test: Long Form; Better indicated by lower values)										
49 (1 study) 8 weeks	serious[4]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,4] due to risk of bias, imprecision	25	24	N/A	N/A	The mean face recognition (iq <70 and >70 combined) in the intervention groups was **1.13 standard deviations higher** (0.53 to 1.74 higher)
Social skills (IQ <70 and >70 combined) (measured with: Social Skills Rating System (SSRS): Social skills (standardized score); Better indicated by lower values)										
49 (1 study) 8 weeks	no serious risk of bias	very serious[5]	no serious indirectness	very serious[6]	undetected	⊕⊖⊖⊖ **VERY LOW**[5,6] due to inconsistency,	25	24	N/A	N/A	The mean social skills (iq <70 and >70 combined) in the intervention groups was **0.29 standard deviations**

						imprecision					**higher** (0.29 lower to 0.88 higher)
Social skills (IQ <70) (measured with: Social Skills Rating System (SSRS): Social skills (standardized score); Better indicated by lower values)											
25 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨁⨀ **MODERATE**[2] due to imprecision	14	11	N/A	N/A	The mean social skills (iq <70) in the intervention groups was **0.92 standard deviations higher** (0.08 to 1.75 higher)
Social skills (IQ >70) (measured with: Social Skills Rating System (SSRS): Social skills (standardized score); Better indicated by lower values)											
24 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[6]	undetected	⨁⨁⨀⨀ **LOW**[6] due to imprecision	11	13	N/A	N/A	The mean social skills (iq >70) in the intervention groups was **0.29 standard deviations lower** (1.09 lower to 0.52 higher)
Positive social interactions (IQ <70 and >70 combined) (measured with: Behavioural observation: Initiating or maintaining social interactions; Better indicated by lower values)											
49 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨁⨀ **MODERATE**[2] due to imprecision	24	25	N/A	N/A	The mean positive social interactions (iq <70 and >70 combined) in the intervention groups was **0.6 standard deviations higher** (0.02 to 1.17 higher)
Positive social Interactions (IQ <70 and >70 combined) (measured with: Behavioural observation: Social intention without inititating interaction (e.g. proximity); Better indicated by lower values)											
49 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[6]	undetected	⨁⨁⨀⨀ **LOW**[6] due to imprecision	25	24	N/A	N/A	The mean positive social interactions (iq <70 and >70 combined) in the intervention groups was **0.12 standard deviations lower**

										(0.68 lower to 0.45 higher)
Negative social interactions (IQ <70 and >70 combined) (measured with: Behavioural observation: Negative social interaction behaviours; Better indicated by lower values)										
49 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨁⨀ **MODERATE**[2] due to imprecision	25	24	N/A	N/A	The mean negative social interactions (iq <70 and >70 combined) in the intervention groups was **0.88 standard deviations lower** (1.47 to 0.29 lower)

[1] High risk of performance bias as intervention administrator non-blind and risk of detection bias is unclear/unknown as identity of outcome assessor is not reported
[2] N<400
[3] High risk of performance bias as intervention administrator non-blind and risk of detection bias is unclear/unknown as identity of outcome assessor is not reported and no independent reliability or validity data for this outcome measure
[4] High risk of performance bias as intervention administrator non-blind and risk of detection bias is unclear/unknown as identity of outcome assessor is not reported and no reliability or validity data for the long form
[5] Substantial to considerable heterogeneity with an I-squared value of 76% (p = 0.04)
[6] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Enhanced ERT versus standard ERT for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

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 Enhanced emotion recognition training (DVD-based) versus standard emotion recognition training (DVD-based) for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome**		**Risk with Control**	**Risk difference with Enhanced emotion recognition training (DVD-based) versus standard emotion recognition training (DVD-based) for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome** (95% CI)
Emotion recognition (measured with: Faces task (Baron-Cohen et al., 1997): Emotion recognition photographs; Better indicated by lower values)										

25 (1 study) 3 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	12	13	N/A	N/A	The mean emotion recognition in the intervention groups was **1.2 standard deviations higher** (0.34 to 2.07 higher)
Emotion recognition (measured with: Developmental Neuropsychological Assessment (NEPSY-II): Affect Recognition; Better indicated by lower values)											
25 (1 study) 3 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	12	13	N/A	N/A	The mean emotion recognition in the intervention groups was **1.55 standard deviations higher** (0.63 to 2.46 higher)
Positive social behaviours (measured with: Social Communication Questionnaire (SCQ): Social peer interest; Better indicated by lower values)											
25 (1 study) 3 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to risk of bias, imprecision	12	13	N/A	N/A	The mean positive social behaviours in the intervention groups was **0.33 standard deviations higher** (0.46 lower to 1.12 higher)
Positive social behaviours (measured with: Social Communication Questionnaire (SCQ): Eye contact; Better indicated by lower values)											
25 (1 study) 3 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to risk of bias, imprecision	12	13	N/A	N/A	The mean positive social behaviours in the intervention groups was **0.04 standard deviations higher** (0.74 lower to 0.83 higher)
Gaze aversion (measured with: Social Communication Questionnaire (SCQ): Gaze aversion; Better indicated by lower values)											
25 (1 study) 3 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to risk of bias, imprecision	12	13	N/A	N/A	The mean gaze aversion in the intervention groups was **0.14 standard deviations lower** (0.93 lower to 0.64 higher)

[1] High risk of performance bias as intervention adminstered by non-blind parents and risk of detection bias is unclear/unknown as identity (beyond stating 'researcher') and blinding of

outcome assessor unclear and the reliability and validity of this outcome measure is unclear
[2] N<400
[3] High risk of performance and detection bias as parents were non-blind and were intervention administrators and outcome assessors
[4] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.2.6 Educational interventions aimed at the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

LEAP training versus manual-only control for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

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 Intervention-manual-only control	**With Inclusive educational intervention (LEAP) training**		**Risk with Intervention-manual-only control**	**Risk difference with Inclusive educational intervention (LEAP) training** (95% CI)
Social skills (measured with: Social Skills Rating System (SSRS): Positive social skills (percentile rank score); Better indicated by lower values)											
294 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to risk of bias, imprecision	117	177	N/A	N/A	The mean social skills in the intervention groups was **0.76 standard deviations higher** (0.52 to 1 higher)

[1] High risk of performance and response bias as intervention administrators and participants non-blind. In addition, risk of detection bias is unclear/unknown as identity and blinding of outcome assessors not reported
[2] N<400

Combined TeachTown and IBI versus IBI-only for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

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 IBI-only	**With Combined computer-assisted educational intervention and intensive behavioural intervention (IBI) day class program**		**Risk with IBI-only**	**Risk difference with Combined computer-assisted educational intervention and intensive behavioural intervention (IBI) day class program** (95% CI)
Social skills (measured with: Brigance Inventory of Child Development: Social skills; Better indicated by lower values)											
46 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	24	22	N/A	N/A	The mean social skills in the intervention groups was **0.1 standard deviations lower** (0.68 lower to 0.48 higher)
Social skills (preschool subgroup analysis) (measured with: Brigance Inventory of Child Development: Social skills; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	12	11	N/A	N/A	The mean social skills (preschool subgroup analysis) in the intervention groups was **0.18 standard deviations lower** (1 lower to 0.64 higher)
Social skills (K-1 subgroup analysis) (measured with: Brigance Inventory of Child Development: Social skills; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	12	11	N/A	N/A	The mean social skills (k-1 subgroup analysis) in the intervention groups was **0.03 standard deviations lower** (0.85 lower to 0.79 higher)

[1] High risk of performance and response bias as intervention administrators and participants non-blind. Risk of detection bias is unclear/unknown as the identity and blinding of outcome assessors not reported. In addition, for the Brigance Inventory of Child Development scale there are no independent reliability and/or validity data reported
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.2.7 Parent training interventions aimed at the core autism feature of impaired reciprocal social communication and interaction as a direct or indirect outcome

Parent training versus treatment-as-usual for the core autism feature of impaired reciprocal social communication and interaction as a direct or indirect outcome

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 Parent training versus treatment-as-usual for the core autism feature of impaired reciprocal social communication and interaction		Risk with Control	Risk difference with Parent training versus treatment-as-usual for the core autism feature of impaired reciprocal social communication and interaction (95% CI)
Reciprocal social interaction (direct outcome) (measured with: Autism Diagnostic Interview-Revised (ADI-R): Reciprocal Social Interaction; Better indicated by lower values)											
24 (1 study) 52 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	12	12	N/A	N/A	The mean reciprocal social interaction (direct outcome) in the intervention groups was **0.38 standard deviations lower** (1.19 lower to 0.43 higher)
Nonverbal communication (direct outcome) (measured with: Autism Diagnostic Interview-Revised (ADI-R): Nonverbal Communication; Better indicated by lower values)											
24 (1 study) 52 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	12	12	N/A	N/A	The mean nonverbal communication (direct outcome) in the intervention groups was **0.37 standard deviations lower** (1.18 lower to 0.44 higher)

Social skills (indirect outcome) (measured with: Social Skills Questionnaire (Spence, 1995): Total or Scales of Independent Behavior-Revised (SIB): Social interaction; Better indicated by lower values)										
71 (2 studies) 10-12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[3,4] due to risk of bias, imprecision	25	46	N/A	N/A	The mean social skills (indirect outcome) in the intervention groups was **0.77 standard deviations higher** (0.25 to 1.28 higher)
Social skills (indirect outcome; combined workshop + individual sessions) (measured with: Social Skills Questionnaire (Spence, 1995): Total ; Better indicated by lower values)										
51 (1 study) 10 weeks	serious[5]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4,5] due to risk of bias, imprecision	15	36	N/A	N/A	The mean social skills (indirect outcome; combined workshop + individual sessions) in the intervention groups was **0.98 standard deviations higher** (0.34 to 1.61 higher)
Social skills (indirect outcome) (measured with: Scales of Independent Behavior-Revised (SIB-R): Social interaction; Better indicated by lower values)										
20 (1 study) 12 weeks	serious[6]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,6] due to risk of bias, imprecision	10	10	N/A	N/A	The mean social skills (indirect outcome) in the intervention groups was **0.37 standard deviations higher** (0.52 lower to 1.25 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome assessors were non-blind
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias high or unclear as either parent-rated and parents were non-blind and involved in the intervention or the identity and blinding of the outcome assessor was not reported
[4] N<400
[5] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as parent-rated and parents were non-blind and involved in the intervention
[6] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias unclear as the identity and blinding of the outcome assessor was not reported

1.2.8 Social-communication interventions aimed at the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

Caregiver- or preschool-teacher- mediated social-communication interventions versus treatment-as-usual for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

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 Treatment-as-usual	**With Caregiver- or preschool-teacher-mediated social-communication interventions**		**Risk with Treatment-as-usual**	**Risk difference with Caregiver- or preschool-teacher-mediated social-communication interventions** (95% CI)
Social interaction (Caregiver-mediated social communication intervention) (measured with: Autism Diagnostic Observation Schedule (ADOS/ADOS-G): Social Interaction; Better indicated by lower values)											
180 (2 studies) 52-56 weeks	no serious risk of bias	serious[1]	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to inconsistency, imprecision	89	91	N/A	N/A	The mean social interaction (caregiver-mediated social communication intervention) in the intervention groups was **0.29 standard deviations lower** (0.59 lower to 0 higher)
Communication (Caregiver-mediated social communication intervention) (measured with: Autism Diagnostic Observation Schedule (ADOS/ADOS-G): Communication ; Better indicated by lower values)											
152 (1 study) 56 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊕⊖⊖ **LOW**[2,3] due to imprecision, publication bias	75	77	N/A	N/A	The mean communication (caregiver-mediated social communication intervention) in the intervention groups was **0.03 standard deviations lower** (0.35 lower to 0.29 higher)

Social interaction and communication (Caregiver-mediated social communication intervention) (measured with: Autism Diagnostic Observation Schedule (ADOS/ADOS-G): Communication & Social Interaction; Better indicated by lower values)										
202 (2 studies) 39-56 weeks	no serious risk of bias	serious[1]	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to inconsistency, imprecision, publication bias	100	102	N/A	N/A	The mean social interaction and communication (caregiver-mediated social communication intervention) in the intervention groups was **0 standard deviations higher** (0.28 lower to 0.27 higher)
Parent-rated social-communication (Caregiver-mediated social communication intervention) (measured with: Communication and Symbolic Behavior Scales Developmental Profile (CSBS DP): Social composite; Better indicated by lower values)										
152 (1 study) 56 weeks	serious[4]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,4] due to risk of bias, imprecision	75	77	N/A	N/A	The mean parent-rated social-communication (caregiver-mediated social communication intervention) in the intervention groups was **0.39 standard deviations higher** (0.06 to 0.71 higher)
Communication acts (Caregiver-mediated social communication intervention) (measured with: Behavioural observation: Child communication acts or Parent-Child Free Play Procedure (PCFP): Frequency of intentional communication (weighted); Better indicated by lower values)										
223 (3 studies) 22-56 weeks	no serious risk of bias	serious[1]	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to inconsistency, imprecision	108	115	N/A	N/A	The mean communication acts (caregiver-mediated social communication intervention) in the intervention groups was **0.37 standard deviations higher** (0.1 to 0.64 higher)
Examiner-child joint/shared attention (Caregiver- or preschool-teacher- mediated social-communication intervention) (measured with: ESCs (Early Social Communication Scales): Initiating Joint Attention (IJA); Better indicated by lower values)										

111 (2 studies) 8-22 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	50	61	N/A	N/A	The mean examiner-child joint/shared attention (caregiver- or preschool-teacher- mediated social-communication intervention) in the intervention groups was **0.06 standard deviations lower** (0.43 lower to 0.32 higher)
Examiner-child joint/shared attention (Caregiver-mediated social communication intervention) (measured with: EScs (Early Social Communication Scales): Initiating Joint Attention (IJA); Better indicated by lower values)											
51 (1 study) 22 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	23	28	N/A	N/A	The mean examiner-child joint/shared attention (caregiver-mediated social communication intervention) in the intervention groups was **0.12 standard deviations lower** (0.68 lower to 0.43 higher)
Examiner-child joint/shared attention (Preschool-teacher-mediated social communication intervention) (measured with: EScs (Early Social Communication Scales): Initiating Joint Attention (IJA); Better indicated by lower values)											
60 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	27	33	N/A	N/A	The mean examiner-child joint/shared attention (preschool-teacher-mediated social communication intervention) in the intervention groups was **0 standard deviations higher** (0.51 lower to 0.51 higher)
Parent-child joint/shared attention (Caregiver- or preschool-teacher- mediated social-communication intervention) (measured with: Behavioural observation: Parent-child joint/shared attention; Better indicated by lower values)											
302	no	no serious	no serious	serious[2]	undetected	⊕⊕⊕⊖	147	155	N/A	N/A	The mean parent-child

(5 studies) 8-56 weeks	serious risk of bias	inconsistency	indirectness			**MODERATE**[2] due to imprecision					joint/shared attention (caregiver- or preschool-teacher- mediated social-communication intervention) in the intervention groups was **0.30 standard deviations higher** (0.07 to 0.53 higher)
Parent-child joint/shared attention (Caregiver-mediated social communication intervention) (measured with: Behavioural observation: Parent-child joint/shared attention; Better indicated by lower values)											
241 (4 studies) 8-56 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	120	121	N/A	N/A	The mean parent-child joint/shared attention (caregiver-mediated social communication intervention) in the intervention groups was **0.33 standard deviations higher** (0.07 to 0.59 higher)
Parent-child joint/shared attention -(Preschool-teacher-mediated social communication intervention) (measured with: Behavioural observation: Parent-child joint/shared attention; Better indicated by lower values)											
61 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	27	34	N/A	N/A	The mean parent-child joint/shared attention - (preschool-teacher-mediated social communication intervention) in the intervention groups was **0.17 standard deviations higher** (0.33 lower to 0.68 higher)
Parent-child joint attention responses (Caregiver-mediated social communication intervention) (measured with: Behavioural observation: Joint attention responses; Better indicated by lower values)											
61 (2 studies)	no serious	very serious[6]	no serious indirectness	serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,6]	31	30	N/A	N/A	The mean parent-child joint attention responses

8-52 weeks	risk of bias					due to inconsistency, imprecision				(caregiver-mediated social communication intervention) in the intervention groups was **2.25 standard deviations higher** (1.57 to 2.93 higher)
Parent-child joint engagement (Caregiver- or preschool-teacher- mediated social-communication intervention) (measured with: Behavioural observation: Joint engagement; Better indicated by lower values)										
99 (2 studies) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	46	53	N/A	N/A	The mean parent-child joint engagement (caregiver- or preschool-teacher-mediated social-communication intervention) in the intervention groups was **0.55 standard deviations higher** (0.14 to 0.95 higher)
Parent-child joint engagement (Caregiver-mediated social communication intervention) (measured with: Behavioural observation: Joint engagement; Better indicated by lower values)										
38 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	19	19	N/A	N/A	The mean parent-child joint engagement (caregiver-mediated social communication intervention) in the intervention groups was **0.85 standard deviations higher** (0.18 to 1.52 higher)
Parent-child joint engagement (Preschool-teacher-mediated social communication intervention) (measured with: Behavioural observation: Joint engagement; Better indicated by lower values)										
61 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	27	34	N/A	N/A	The mean parent-child joint engagement (preschool-teacher-mediated social communication

										intervention) in the intervention groups was **0.37 standard deviations higher** (0.14 lower to 0.88 higher)
Teacher-child joint/shared attention (Preschool-teacher-mediated social communication intervention) (measured with: Behavioural observation (Preschool teacher-child play): Joint attention; Better indicated by lower values)										
61 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	27	34	N/A	N/A	The mean teacher-child joint/shared attention (preschool-teacher-mediated social communication intervention) in the intervention groups was **0.57 standard deviations higher** (0.05 to 1.08 higher)
Teacher-child joint engagement (Preschool-teacher-mediated social communication intervention) (measured with: Behavioural observation (Preschool teacher-child play): Joint engagement; Better indicated by lower values)										
61 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	27	34	N/A	N/A	The mean teacher-child joint engagement (preschool-teacher-mediated social communication intervention) in the intervention groups was **0.31 standard deviations lower** (0.81 lower to 0.2 higher)
Behaviour requests (Caregiver-mediated social communication intervention) (measured with: EScs (Early Social Communication Scales): Initiating Behavioural Requests (IBR); Better indicated by lower values)										
51 (1 study) 22 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	23	28	N/A	N/A	The mean behaviour requests (caregiver-mediated social communication intervention) in the

											intervention groups was **0.18 standard deviations higher** (0.37 lower to 0.73 higher)
Behaviour requests (Caregiver-mediated social communication intervention) (Copy) (measured with: EScs (Early Social Communication Scales): Initiating Behavioural Requests (IBR); Better indicated by lower values)											
49 (1 study) 39 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	25	24	N/A	N/A	The mean behaviour requests (caregiver-mediated social communication intervention) (copy) in the intervention groups was **0.07 standard deviations higher** (0.49 lower to 0.63 higher)
Non-verbal communication (Caregiver-mediated social communication intervention) (measured with: Parent Interview for Autism-Clinical Version (PIA-CV): Nonverbal communication; Better indicated by lower values)											
47 (1 study) 22 weeks	serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	20	27	N/A	N/A	The mean non-verbal communication (caregiver-mediated social communication intervention) in the intervention groups was **0.09 standard deviations lower** (0.67 lower to 0.49 higher)
Non-verbal communication (Caregiver-mediated social communication intervention) (measured with: Parent Interview for Autism-Clinical Version (PIA-CV): Nonverbal communication; Better indicated by lower values)											
47 (1 study) 39 weeks	serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	24	23	N/A	N/A	The mean non-verbal communication (caregiver-mediated social communication intervention) in the intervention groups was **0.04 standard deviations lower**

											(0.62 lower to 0.53 higher)
Focusing on faces (Caregiver-mediated social communication intervention) (measured with: Behavioural observation (PJAM): Focusing on faces; Better indicated by lower values)											
23 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	12	11	N/A	N/A	The mean focusing on faces (caregiver-mediated social communication intervention) in the intervention groups was **1.87 standard deviations higher** (0.86 to 2.88 higher)
Focusing on faces (Caregiver-mediated social communication intervention) (measured with: Behavioural observation (PJAM): Focusing on faces; Better indicated by lower values)											
23 (1 study) 60 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	12	11	N/A	N/A	The mean focusing on faces (caregiver-mediated social communication intervention) in the intervention groups was **0.91 standard deviations higher** (0.05 to 1.78 higher)
Turn-taking (Caregiver-mediated social communication intervention) (measured with: Behavioural observation (PJAM): Turn-Taking; Better indicated by lower values)											
23 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	12	11	N/A	N/A	The mean turn-taking (caregiver-mediated social communication intervention) in the intervention groups was **0.73 standard deviations higher** (0.12 lower to 1.58 higher)
Turn-taking Caregiver-mediated social communication intervention) (measured with: Behavioural observation (PJAM): Turn-Taking; Better indicated by lower values)											
23	no	no serious	no serious	very	undetected	⊕⊕⊖⊖	12	11	N/A	N/A	The mean turn-taking

(1 study) 60 weeks	serious risk of bias	inconsistency	indirectness	serious[5]		**LOW**[5] due to imprecision				caregiver-mediated social communication intervention) in the intervention groups was **0.14 standard deviations lower** (0.96 lower to 0.68 higher)

[1] Moderate to substantial heterogeneity
[2] N<400
[3] High risk of selective reporting bias as data could not be extracted from ALDRED2001/2004 for the ADOS communication subdomain
[4] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome measure was parent-reported and parents were non-blind and involved in the delivery of the intervention
[5] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[6] Substantial to considerable heterogeneity

Peer-mediated (and/or therapist-mediated) social-communication interventions versus treatment-as-usual for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

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 Treatment-as-usual	**With Peer-mediated (and/or therapist-mediated) social-communication interventions**		**Risk with Treatment-as-usual**	**Risk difference with Peer-mediated (and/or therapist-mediated) social-communication interventions** (95% CI)
Peer-child joint engagement (Peer-mediated social-communication intervention) (measured with: Behavioural observation: Number of intervals of child-initiated social interaction with unfamiliar TD peer or Behavioural observation: % time in joint engagement in playground; Better indicated by lower values)											
114 (2 studies) 6-15 weeks	no serious risk of bias	very serious[1]	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to inconsistency, imprecision, publication bias	51	63	N/A	N/A	The mean peer-child joint engagement (peer-mediated social-communication intervention) in the intervention groups was **0.7 standard deviations higher** (0.31 to 1.08 higher)
Peer-child joint engagement (Therapist-mediated social-communication intervention) (measured with: Behavioural observations of % time in joint engagement in playground ; Better indicated by lower values)											
29 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4] due to imprecision	14	15	N/A	N/A	The mean peer-child joint engagement (therapist-mediated social-communication intervention) in the intervention groups was **0.03 standard deviations higher** (0.7 lower to 0.76 higher)

Peer-child joint engagement (Peer-mediated social-communication intervention) (measured with: Behavioural observations of % time in joint engagement in playground; Better indicated by lower values)											
29 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4] due to imprecision	14	15	N/A	N/A	The mean peer-child joint engagement (peer-mediated social-communication intervention) in the intervention groups was **0.12 standard deviations higher** (0.61 lower to 0.84 higher)
Peer-child joint engagement (Both therapist- and peer- mediated social-communication intervention) (measured with: Behavioural observations of % time in joint engagement in playground; Better indicated by lower values)											
29 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4] due to imprecision	14	15	N/A	N/A	The mean peer-child joint engagement (both therapist- and peer- mediated social-communication intervention) in the intervention groups was **0 standard deviations higher** (0.73 lower to 0.73 higher)
Peer-child joint engagement (Therapist-mediated social-communication intervention) (measured with: Behavioural observations of % time in joint engagement in playground; Better indicated by lower values)											
30 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4] due to imprecision	15	15	N/A	N/A	The mean peer-child joint engagement (therapist-mediated social-communication intervention) in the intervention groups was **0.13 standard deviations higher** (0.59 lower to 0.85 higher)

Peer-child joint engagement (Peer-mediated social-communication intervention) (measured with: Behavioural observations of % time in joint engagement in playground; Better indicated by lower values)											
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4] due to imprecision	15	14	N/A	N/A	The mean peer-child joint engagement (peer-mediated social-communication intervention) in the intervention groups was **0.75 standard deviations higher** (0 to 1.51 higher)
Peer-child joint engagement (Both therapist- and peer- mediated social-communication intervention) (measured with: Behavioural observations of % time in joint engagement in playground; Better indicated by lower values)											
30 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	15	15	N/A	N/A	The mean peer-child joint engagement (both therapist- and peer- mediated social-communication intervention) in the intervention groups was **0.86 standard deviations higher** (0.11 to 1.62 higher)
Child-initiated social interactions (Peer-mediated social-communication intervention) (measured with: Behavioural observations of number of child-initiated social interactions with familiar TD peer; Better indicated by lower values)											
85 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊕⊖⊖ **LOW**[2,3] due to imprecision, publication bias	37	48	N/A	N/A	The mean child-initiated social interactions (peer-mediated social-communication intervention) in the intervention groups was **0.65 standard deviations higher** (0.21 to 1.09 higher)

Child-initiated social interactions (Peer-mediated social-communication intervention) (measured with: Behavioural observations of number of child-initiated social interactions with unfamiliar TD peer; Better indicated by lower values)										
85 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊕⊖⊖ **LOW**[2,3] due to imprecision, publication bias	37	48	N/A	N/A	The mean child-initiated social interactions (peer-mediated social-communication intervention) in the intervention groups was **0.68 standard deviations higher** (0.24 to 1.12 higher)
Social network salience (Therapist-mediated social-communication intervention) (measured with: Social Network Survey (SNS): Social Network Salience Ratio; Better indicated by lower values)										
30 (1 study) 6 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	15	15	N/A	N/A	The mean social network salience (therapist-mediated social-communication intervention) in the intervention groups was **0.05 standard deviations lower** (0.77 lower to 0.66 higher)
Social network salience (Peer-mediated social-communication intervention) (measured with: Social Network Survey (SNS): Social Network Salience Ratio; Better indicated by lower values)										
30 (1 study) 6 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	15	15	N/A	N/A	The mean social network salience (peer-mediated social-communication intervention) in the intervention groups was **0.42 standard deviations higher** (0.3 lower to 1.15 higher)
Social network salience (Both therapist-mediated and peer-mediated social-communication intervention) (measured with:										

Social Network Survey (SNS): Social Network Salience Ratio; Better indicated by lower values)											
30 (1 study) 6 weeks	serious[5]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,5] due to risk of bias, imprecision	15	15	N/A	N/A	The mean social network salience (both therapist-mediated and peer-mediated social-communication intervention) in the intervention groups was **1.15 standard deviations higher** (0.37 to 1.93 higher)
Social network salience (Therapist-mediated social-communication intervention) (measured with: Social Network Survey (SNS): Social Network Salience Ratio; Better indicated by lower values)											
29 (1 study) 12 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	15	14	N/A	N/A	The mean social network salience (therapist-mediated social-communication intervention) in the intervention groups was **0.51 standard deviations lower** (1.25 lower to 0.23 higher)
Social network salience (Peer-mediated social-communication intervention) (measured with: Social Network Survey (SNS): Social Network Salience Ratio; Better indicated by lower values)											
30 (1 study) 12 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	15	15	N/A	N/A	The mean social network salience (peer-mediated social-communication intervention) in the intervention groups was **0.03 standard deviations higher** (0.68 lower to 0.75 higher)
Social network salience (Both therapist-mediated and peer-mediated social-communication intervention) (measured with:											

Social Network Survey (SNS): Social Network Salience Ratio; Better indicated by lower values)											
30 (1 study) 12 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	15	15	N/A	N/A	The mean social network salience (both therapist-mediated and peer-mediated social-communication intervention) in the intervention groups was **0.32 standard deviations higher** (0.4 lower to 1.04 higher)
Number of received friendship nominations (Therapist-mediated social-communication intervention) (measured with: Social Network Survey (SNS): Number of received friendship nominations (Indegrees); Better indicated by lower values)											
30 (1 study) 6 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	15	15	N/A	N/A	The mean number of received friendship nominations (therapist-mediated social-communication intervention) in the intervention groups was **0.18 standard deviations lower** (0.9 lower to 0.54 higher)
Number of received friendship nominations (Peer-mediated social-communication intervention) (measured with: Social Network Survey (SNS): Number of received friendship nominations (Indegrees); Better indicated by lower values)											
30 (1 study) 6 weeks	serious[5]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,5] due to risk of bias, imprecision	15	15	N/A	N/A	The mean number of received friendship nominations (peer-mediated social-communication intervention) in the intervention groups was **0.96 standard deviations higher** (0.19 to 1.72 higher)

Number of received friendship nominations (Both therapist-mediated and peer-mediated social-communication intervention) (measured with: Social Network Survey (SNS): Number of received friendship nominations (Indegrees); Better indicated by lower values)											
30 (1 study) 6 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	15	15	N/A	N/A	The mean number of received friendship nominations (both therapist-mediated and peer-mediated social-communication intervention) in the intervention groups was **0.51 standard deviations higher** (0.22 lower to 1.24 higher)
Number of received friendship nominations (Therapist-mediated social-communication intervention) (measured with: Social Network Survey (SNS): Number of received friendship nominations (Indegrees); Better indicated by lower values)											
29 (1 study) 12 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	15	14	N/A	N/A	The mean number of received friendship nominations (therapist-mediated social-communication intervention) in the intervention groups was **0.1 standard deviations lower** (0.83 lower to 0.63 higher)
Number of received friendship nominations (Peer-mediated social-communication intervention) (measured with: Social Network Survey (SNS): Number of received friendship nominations (Indegrees); Better indicated by lower values)											
30 (1 study) 12 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	15	15	N/A	N/A	The mean number of received friendship nominations (peer-mediated social-communication intervention) in the intervention groups was **0.33 standard deviations**

										higher (0.39 lower to 1.05 higher)
Number of received friendship nominations (Both therapist-mediated and peer-mediated social-communication intervention) (measured with: Social Network Survey (SNS): Number of received friendship nominations (Indegrees); Better indicated by lower values)										
30 (1 study) 12 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	15	15	N/A	N/A	The mean number of received friendship nominations (both therapist-mediated and peer-mediated social-communication intervention) in the intervention groups was **0.25 standard deviations higher** (0.47 lower to 0.97 higher)
Number of times child identified as someone other children don't like to 'hang out with' (Therapist-mediated social-communication intervention) (measured with: Social Network Survey (SNS): Rejections; Better indicated by lower values)										
27 (1 study) 6 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	14	13	N/A	N/A	The mean number of times child identified as someone other children don't like to 'hang out with' (therapist-mediated social-communication intervention) in the intervention groups was **0.44 standard deviations higher** (0.32 lower to 1.21 higher)
Number of times child identified as someone other children don't like to 'hang out with' (Peer-mediated social-communication intervention) (measured with: Social Network Survey (SNS): Rejections; Better indicated by lower values)										
29 (1 study)	serious[5]	no serious	no serious	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,5]	14	15	N/A	N/A	The mean number of times child identified as someone

6 weeks		inconsistency	indirectness			due to risk of bias, imprecision					other children don't like to 'hang out with' (peer-mediated social-communication intervention) in the intervention groups was **0.94 standard deviations higher** (0.17 to 1.72 higher)
Number of times child identified as someone other children don't like to 'hang out with' (Both therapist-mediated and peer-mediated social-communication intervention) (measured with: Social Network Survey (SNS): Rejections; Better indicated by lower values)											
29 (1 study) 6 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	14	15	N/A	N/A	The mean number of times child identified as someone other children don't like to 'hang out with' (both therapist-mediated and peer-mediated social-communication intervention) in the intervention groups was **0.35 standard deviations higher** (0.38 lower to 1.09 higher)
Number of times child identified as someone other children don't like to 'hang out with' (Therapist-mediated social-communication intervention) (measured with: Social Network Survey (SNS): Rejections; Better indicated by lower values)											
26 (1 study) 12 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	14	12	N/A	N/A	The mean number of times child identified as someone other children don't like to 'hang out with' (therapist-mediated social-communication intervention) in the intervention groups was **0.17 standard deviations**

										lower (0.94 lower to 0.61 higher)
Number of times child identified as someone other children don't like to 'hang out with' (Peer-mediated social-communication intervention) (measured with: Social Network Survey (SNS): Rejections; Better indicated by lower values)										
29 (1 study) 12 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	14	15	N/A	N/A	The mean number of times child identified as someone other children don't like to 'hang out with' (peer-mediated social-communication intervention) in the intervention groups was **0.14 standard deviations higher** (0.59 lower to 0.87 higher)
Number of times child identified as someone other children don't like to 'hang out with' (Both therapist-mediated and peer-mediated social-communication intervention) (measured with: Social Network Survey (SNS): Rejections; Better indicated by lower values)										
29 (1 study) 12 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	14	15	N/A	N/A	The mean number of times child identified as someone other children don't like to 'hang out with' (both therapist-mediated and peer-mediated social-communication intervention) in the intervention groups was **0.42 standard deviations higher** (0.32 lower to 1.15 higher)
Teacher-rated social skills (Therapist-mediated social-communication intervention) (measured with: Teacher Perception of Social Skills (TPSS): Total; Better indicated by lower values)										

26 (1 study) 6 weeks	serious[6]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,6] due to risk of bias, imprecision	13	13	N/A	N/A	The mean teacher-rated social skills (therapist-mediated social-communication intervention) in the intervention groups was **0.11 standard deviations lower** (0.88 lower to 0.66 higher)
Teacher-rated social skills (Peer-mediated social-communication intervention) (measured with: Teacher Perception of Social Skills (TPSS): Total; Better indicated by lower values)											
28 (1 study) 6 weeks	serious[6]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,6] due to risk of bias, imprecision	13	15	N/A	N/A	The mean teacher-rated social skills (peer-mediated social-communication intervention) in the intervention groups was **0.36 standard deviations higher** (0.39 lower to 1.11 higher)
Teacher-rated social skills (Both therapist-mediated and peer-mediated social-communication intervention) (measured with: Teacher Perception of Social Skills (TPSS): Total; Better indicated by lower values)											
28 (1 study) 6 weeks	serious[6]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,6] due to risk of bias, imprecision	13	15	N/A	N/A	The mean teacher-rated social skills (both therapist-mediated and peer-mediated social-communication intervention) in the intervention groups was **0.32 standard deviations higher** (0.43 lower to 1.06 higher)
Teacher-rated social skills (Therapist-mediated social-communication intervention) (measured with: Teacher Perception of Social Skills (TPSS): Total; Better indicated by lower values)											

25 (1 study) 12 weeks	serious[6]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,6] due to risk of bias, imprecision	14	11	N/A	N/A	The mean teacher-rated social skills (therapist-mediated social-communication intervention) in the intervention groups was **0.02 standard deviations lower** (0.81 lower to 0.77 higher)
Teacher-rated social skills (Peer-mediated social-communication intervention) (measured with: Teacher Perception of Social Skills (TPSS): Total; Better indicated by lower values)											
29 (1 study) 12 weeks	serious[6]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,6] due to risk of bias, imprecision	14	15	N/A	N/A	The mean teacher-rated social skills (peer-mediated social-communication intervention) in the intervention groups was **0.14 standard deviations higher** (0.59 lower to 0.87 higher)
Teacher-rated social skills (Both therapist-mediated and peer-mediated social-communication intervention) (measured with: Teacher Perception of Social Skills (TPSS): Total; Better indicated by lower values)											
29 (1 study) 12 weeks	serious[6]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,6] due to risk of bias, imprecision	14	15	N/A	N/A	The mean teacher-rated social skills (both therapist-mediated and peer-mediated social-communication intervention) in the intervention groups was **0.48 standard deviations higher** (0.26 lower to 1.22 higher)

[1] Substantial heterogeneity
[2] N<400
[3] High risk of selective reporting bias for ROEYERS1996 as data cannot be extracted for the Social Behavior Rating Scale which was designed to measure generalization of gains in social behaviour to larger school setting
[4] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

[5] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear as blinding of the typically-developing peer completers was not reported
[6] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear as teacher-rated and blinding of teachers was not reported

Joint attention training and EBI/EIBI versus EBI/EIBI only for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

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 EBI/EIBI only	**With Combined joint attention training and EBI/EIBI**		**Risk with EBI/EIBI only**	**Risk difference with Combined joint attention training and EBI/EIBI** (95% CI)
Examiner-child joint attention (Child-initiated JA) (measured with: EScs (Early Social Communication Scales): Coordinated JA looks; Better indicated by lower values)											
37 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	20	N/A	N/A	The mean examiner-child joint attention (child-initiated ja) in the intervention groups was **0.09 standard deviations lower** (0.74 lower to 0.56 higher)
Examiner-child joint attention (Child-initiated JA) (measured with: EScs (Early Social Communication Scales): Showing; Better indicated by lower values)											
37 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	20	N/A	N/A	The mean examiner-child joint attention (child-initiated ja) in the intervention groups was **0.55 standard deviations higher** (0.11 lower to 1.21 higher)
Examiner-child joint attention (Child-initiated JA) (measured with: EScs (Early Social Communication Scales): Pointing; Better indicated by lower values)											
37 (1 study)	no serious risk of	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to	17	20	N/A	N/A	The mean examiner-child joint attention (child-initiated ja) in the intervention groups was

6 weeks	bias					imprecision				**0.69 standard deviations higher** (0.02 to 1.36 higher)
Examiner-child joint attention (Child-initiated JA) (measured with: EScs (Early Social Communication Scales): Giving; Better indicated by lower values)										
37 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	20	N/A	N/A	The mean examiner-child joint attention (child-initiated ja) in the intervention groups was **0.48 standard deviations higher** (0.18 lower to 1.14 higher)
Examiner-child joint attention (Child-initiated JA) (measured with: Communication and Symbolic Behavior Scales Developmental Profile (CSBS DP): Initiating joint attention (IJA); Better indicated by lower values)										
48 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	24	24	N/A	N/A	The mean examiner-child joint attention (child-initiated ja) in the intervention groups was **0.31 standard deviations higher** (0.26 lower to 0.88 higher)
Examiner-child joint attention (Child-initiated JA) (measured with: Communication and Symbolic Behavior Scales Developmental Profile (CSBS DP): Initiating joint attention (IJA); Better indicated by lower values)										
48 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	24	24	N/A	N/A	The mean examiner-child joint attention (child-initiated ja) in the intervention groups was **0.44 standard deviations higher** (0.14 lower to 1.01 higher)
Examiner-child joint attention (Child responding to JA) (measured with: EScs (Early Social Communication Scales): Responding to Joint Attention (RJA); Better indicated by lower values)										

37 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	17	20	N/A	N/A	The mean examiner-child joint attention (child responding to ja) in the intervention groups was **1.11 standard deviations higher** (0.41 to 1.81 higher)
Examiner-child shared positive affect (measured with: EScs (Early Social Communication Scales): JA & shared positive affect or Communication and Symbolic Behavior Scales Developmental Profile (CSBS DP): Shared positive affect (SPA); Better indicated by lower values)											
84 (2 studies) 6-26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	40	44	N/A	N/A	The mean examiner-child shared positive affect in the intervention groups was **0.04 standard deviations higher** (0.39 lower to 0.47 higher)
Examiner-child shared positive affect (measured with: EScs (Early Social Communication Scales): JA & shared positive affect or Communication and Symbolic Behavior Scales Developmental Profile (CSBS DP): Shared positive affect (SPA); Better indicated by lower values)											
84 (2 studies) 26-52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	44	N/A	N/A	The mean examiner-child shared positive affect in the intervention groups was **0.43 standard deviations higher** (0 to 0.87 higher)
Examiner-child shared positive affect (measured with: EScs (Early Social Communication Scales): JA & shared positive affect; Better indicated by lower values)											
36 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	16	20	N/A	N/A	The mean examiner-child shared positive affect in the intervention groups was **0.6 standard deviations higher** (0.08 lower to 1.27 higher)
Examiner-child joint attention, shared positive affect & utterance (measured with: EScs (Early Social Communication Scales): JA & shared positive											

affect & utterance; Better indicated by lower values)										
36 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	16	20	N/A	N/A	The mean examiner-child joint attention, shared positive affect & utterance in the intervention groups was **0.04 standard deviations higher** (0.62 lower to 0.7 higher)
Examiner-child joint attention, shared positive affect & utterance (measured with: EScs (Early Social Communication Scales): JA & shared positive affect & utterance; Better indicated by lower values)										
36 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	16	20	N/A	N/A	The mean examiner-child joint attention, shared positive affect & utterance in the intervention groups was **0.56 standard deviations higher** (0.12 lower to 1.23 higher)
Examiner-child joint attention, shared positive affect & utterance (measured with: EScs (Early Social Communication Scales): JA & shared positive affect & utterance; Better indicated by lower values)										
36 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	16	20	N/A	N/A	The mean examiner-child joint attention, shared positive affect & utterance in the intervention groups was **0.77 standard deviations higher** (0.09 to 1.46 higher)
Examiner-child socially engaged imitation (measured with: Behavioural observation: Socially engaged imitation (SEI); Better indicated by lower values)										
48 (1 study) 26 weeks	no serious risk of	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to	24	24	N/A	N/A	The mean examiner-child socially engaged imitation in the intervention groups was **0.29 standard deviations**

	bias					imprecision					**higher** (0.28 lower to 0.86 higher)
Examiner-child socially engaged imitation (measured with: Behavioural observation: Socially engaged imitation (SEI); Better indicated by lower values)											
48 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	24	24	N/A	N/A	The mean examiner-child socially engaged imitation in the intervention groups was **0.73 standard deviations higher** (0.15 to 1.32 higher)
Mother-child joint attention (Child-initiated JA) (measured with: Behavioural observation: Mother-child interaction (Coordinated JA looks); Better indicated by lower values)											
37 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	20	N/A	N/A	The mean mother-child joint attention (child-initiated ja) in the intervention groups was **0.48 standard deviations higher** (0.18 lower to 1.13 higher)
Mother-child joint attention (Child-initiated JA) (measured with: Behavioural observation: Mother-child interaction (Showing); Better indicated by lower values)											
37 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	20	N/A	N/A	The mean mother-child joint attention (child-initiated ja) in the intervention groups was **0.51 standard deviations higher** (0.15 lower to 1.16 higher)
Mother-child joint attention (Child-initiated JA) (measured with: Behavioural observation: Mother-child interaction (Pointing); Better indicated by lower values)											
37 (1 study) 6 weeks	no serious risk of	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to	17	20	N/A	N/A	The mean mother-child joint attention (child-initiated ja) in the intervention groups was **0.39 standard deviations**

	bias					imprecision					**lower** (1.04 lower to 0.27 higher)
Mother-child joint attention (Child-initiated JA) (measured with: Behavioural observation: Mother-child interaction (Giving); Better indicated by lower values)											
37 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	20	N/A	N/A	The mean mother-child joint attention (child-initiated ja) in the intervention groups was **0.36 standard deviations higher** (0.3 lower to 1.01 higher)
Mother-child joint attention (Child-initiated JA) (measured with: Behavioural observation: Mother-child interaction – Duration of JA (seconds); Better indicated by lower values)											
37 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	17	20	N/A	N/A	The mean mother-child joint attention (child-initiated ja) in the intervention groups was **0.77 standard deviations higher** (0.1 to 1.45 higher)
Mother-child joint attention (Child-initiated JA) (measured with: Behavioural observation: Mother-child interaction – Duration of JA (seconds); Better indicated by lower values)											
37 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	20	N/A	N/A	The mean mother-child joint attention (child-initiated ja) in the intervention groups was **0.19 standard deviations higher** (0.46 lower to 0.83 higher)
Mother-child joint attention (Child-initiated JA) (measured with: Behavioural observation: Mother-child interaction – Duration of JA (seconds); Better indicated by lower values)											
36 (1 study)	no serious	no serious	no serious	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2]	16	20	N/A	N/A	The mean mother-child joint attention (child-initiated ja) in

52 weeks	risk of bias	inconsistency	indirectness			due to imprecision					the intervention groups was **0.81 standard deviations higher** (0.13 to 1.5 higher)
Examiner-child and mother-child joint attention (JA initiation composite) (measured with: EScs and mother-child interaction observations: JA initiation composite; Better indicated by lower values)											
37 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	20	N/A	N/A	The mean examiner-child and mother-child joint attention (ja initiation composite) in the intervention groups was **0.51 standard deviations higher** (0.15 lower to 1.17 higher)
Examiner-child and mother-child joint attention (JA initiation composite) (measured with: EScs and mother-child interaction observations: JA initiation composite; Better indicated by lower values)											
37 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	20	N/A	N/A	The mean examiner-child and mother-child joint attention (ja initiation composite) in the intervention groups was **0.53 standard deviations higher** (0.13 lower to 1.18 higher)
Examiner-child and mother-child joint attention (JA initiation composite) (measured with: EScs and mother-child interaction observations: JA initiation composite; Better indicated by lower values)											
36 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	16	20	N/A	N/A	The mean examiner-child and mother-child joint attention (ja initiation composite) in the intervention groups was **0.99 standard deviations higher** (0.29 to 1.69 higher)

Examiner-child and mother-child joint attention (JA responses composite) (measured with: EScs and mother-child interaction observations: JA responses composite; Better indicated by lower values)										
37 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	17	20	N/A	N/A	The mean examiner-child and mother-child joint attention (ja responses composite) in the intervention groups was **1.11 standard deviations higher** (0.41 to 1.81 higher)
Examiner-child and mother-child joint attention (JA responses composite) (measured with: EScs and mother-child interaction observations: JA responses composite; Better indicated by lower values)										
37 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	17	20	N/A	N/A	The mean examiner-child and mother-child joint attention (ja responses composite) in the intervention groups was **0.8 standard deviations higher** (0.12 to 1.47 higher)
Examiner-child and mother-child joint attention (JA responses composite) (measured with: EScs and mother-child interaction observations: JA responses composite; Better indicated by lower values)										
36 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	16	20	N/A	N/A	The mean examiner-child and mother-child joint attention (ja responses composite) in the intervention groups was **0.17 standard deviations higher** (0.49 lower to 0.83 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] N<400

LEGO® therapy versus SULP for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

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 Social Use of Language Programme (SULP)	**With LEGO therapy**		**Risk with Social Use of Language Programme (SULP)**	**Risk difference with LEGO therapy** (95% CI)
Social interaction (measured with: Gilliam Autism Rating Scale (GARS): Social interaction; Better indicated by lower values)											
31 (1 study) 18 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	15	16	N/A	N/A	The mean social interaction in the intervention groups was **0.73 standard deviations lower** (1.46 lower to 0 higher)
Frequency of child-initiated social interactions with TD peers (measured with: Behavioural observation; Better indicated by lower values)											
21 (1 study) 18 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to risk of bias, imprecision	10	11	N/A	N/A	The mean frequency of child-initiated social interactions with td peers in the intervention groups was **0.23 standard deviations higher** (0.63 lower to 1.09 higher)
Duration of all social interactions with TD peers (measured with: Behavioural observation; Better indicated by lower values)											
21 (1 study) 18 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to risk of bias,	10	11	N/A	N/A	The mean duration of all social interactions with td peers in the intervention groups was

						imprecision				**0.27 standard deviations higher** (0.59 lower to 1.13 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear as parent-rated and blinding of parents was not reported
[2] N<400
[3] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias due to non-blinded behavioural observations which were carried out by the investigator and there was no reliability or validity data reported for observation measures
[4] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Social skills group versus treatment-as-usual for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

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 Treatment-as-usual	**With Social skills group**		**Risk with Treatment-as-usual**	**Risk difference with Social skills group** (95% CI)
Social skills (measured with: Social Skills Rating System (SSRS): Assertion or Social Skills Rating System (SSRS): Social skills (standardized score) or Behavior Assessment System for Children, 2nd ed., parent rated (BASC-2-PRS): Social skills; Better indicated by lower values)											
137 (3 studies) 6-12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to risk of bias, imprecision	67	70	N/A	N/A	The mean social skills in the intervention groups was **0.6 standard deviations higher** (0.26 to 0.95 higher)

Social impairment (measured with: Social Responsiveness Scale (SRS): Total; Better indicated by lower values)										
35 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	17	18	N/A	N/A	The mean social impairment in the intervention groups was **0.69 standard deviations lower** (1.37 lower to 0 higher)
Adaptive social behaviour (measured with: Social Competence Inventory (SCI): Pro-social index; Better indicated by lower values)										
41 (1 study) 16 weeks	serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	18	23	N/A	N/A	The mean adaptive social behaviour in the intervention groups was **0.11 standard deviations higher** (0.51 lower to 0.73 higher)
Capacity for social interactions (measured with: Social Competence Inventory (SCI): Social initiation index; Better indicated by lower values)										
41 (1 study) 16 weeks	serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	18	23	N/A	N/A	The mean capacity for social interactions in the intervention groups was **0.03 standard deviations lower** (0.65 lower to 0.58 higher)
Study-specific targeted social skills (measured with: Adapted Skillstreaming Checklist (ASC): Total; Better indicated by lower values)										
36 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	18	18	N/A	N/A	The mean study-specific targeted social skills in the intervention groups was **0.9 standard deviations higher** (0.21 to 1.59 higher)
Social skills knowledge (self-rated or researcher-rated) (measured with: Test of Adolescent Social Skills Knowledge (TASSK): Total or Skillstreaming										

Knowledge Assessment: Total; Better indicated by lower values)										
69 (2 studies) 6-12 weeks	serious[6]	very serious[7]	no serious indirectness	serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,6,7] due to risk of bias, inconsistency, imprecision	34	35	N/A	N/A	The mean social skills knowledge (self-rated or researcher-rated) in the intervention groups was **1.58 standard deviations higher** (1.03 to 2.14 higher)
Social skills knowledge (self-rated) (measured with: Test of Adolescent Social Skills Knowledge (TASSK): Total; Better indicated by lower values)										
33 (1 study) 12 weeks	serious[6]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,6] due to risk of bias, imprecision	16	17	N/A	N/A	The mean social skills knowledge (self-rated) in the intervention groups was **2.17 standard deviations higher** (1.29 to 3.06 higher)
Social skills knowledge (researcher-rated) (measured with: Skillstreaming Knowledge Assessment: Total; Better indicated by lower values)										
36 (1 study) 6 weeks	serious[6]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,6] due to risk of bias, imprecision	18	18	N/A	N/A	The mean social skills knowledge (researcher-rated) in the intervention groups was **1.19 standard deviations higher** (0.48 to 1.91 higher)
Feelings of loneliness (measured with: Loneliness Scale: Total; Better indicated by lower values)										
67 (1 study) 12 weeks	serious[8]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,8] due to risk of bias, imprecision	32	35	N/A	N/A	The mean feelings of loneliness in the intervention groups was **0.67 standard deviations lower**

										(1.16 to 0.18 lower)
Popularity (self-rated) (measured with: Piers-Harris Self-Concept Scale (PHS): Popularity; Better indicated by lower values)										
68 (1 study) 12 weeks	serious[8]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,8] due to risk of bias, imprecision	33	35	N/A	N/A	The mean popularity (self-rated) in the intervention groups was **0.56 standard deviations higher** (0.07 to 1.04 higher)
Number of times child invited to a play date (parent-rated) (measured with: Quality of Play Questionnaire (QPQ): Host; Better indicated by lower values)										
97 (2 studies) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	45	52	N/A	N/A	The mean number of times child invited to a play date (parent-rated) in the intervention groups was **0.36 standard deviations higher** (0.04 lower to 0.77 higher)
Number of times child invited to a play date (Self-rated) (measured with: Quality of Play Questionnaire (QPQ): Host; Better indicated by lower values)										
33 (1 study) 12 weeks	serious[8]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[5,8] due to risk of bias, imprecision	16	17	N/A	N/A	The mean number of times child invited to a play date (self-rated) in the intervention groups was **0.26 standard deviations lower** (0.95 lower to 0.42 higher)
Time spent in interactive activities (measured with: Quality of Play Questionnaire (QPQ): Engage; Better indicated by lower values)										
62 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias,	27	35	N/A	N/A	The mean time spent in interactive activities in the intervention groups was **0.2 standard deviations**

						imprecision					**higher** (0.31 lower to 0.7 higher)
Time spent in minimally interactive activities (measured with: Quality of Play Questionnaire (QPQ): Disengage; Better indicated by lower values)											
62 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to risk of bias, imprecision	27	35	N/A	N/A	The mean time spent in minimally interactive activities in the intervention groups was **1.31 standard deviations lower** (1.87 to 0.75 lower)
Quality of friendships (self-rated) (measured with: Friendship Qualities Scale (FQS): Total; Better indicated by lower values)											
33 (1 study) 12 weeks	serious[8]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⨁⨀⨀⨀ **VERY LOW**[5,8] due to risk of bias, imprecision	16	17	N/A	N/A	The mean quality of friendships (self-rated) in the intervention groups was **0.14 standard deviations higher** (0.55 lower to 0.82 higher)
Positive treatment response (assessed with: Dichotomous measure of number of participants 'much improved/very improved' on Clinical Global Impression-Improvement (CGI-I))											
41 (1 study) 16 weeks	serious[9]	no serious inconsistency	no serious indirectness	serious[10]	undetected	⨁⨁⨀⨀ **LOW**[9,10] due to risk of bias, imprecision	0/18 (0%)	16/23 (69.6%)	**RR 26.12** (1.67 to 407.99)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Emotion recognition (measured with: Diagnostic Analysis of Nonverbal Accuracy 2 (DANVA2): Child faces; Better indicated by lower values)											

36 (1 study) 6 weeks	serious[11]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[5,11] due to risk of bias, imprecision	18	18	N/A	N/A	The mean emotion recognition in the intervention groups was **0.44 standard deviations higher** (0.22 lower to 1.1 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome measures were parent-rated and parents were non-blind and involved in the intervention
[2] N<400
[3] High risk of selective reporting bias as LOPATA2010 did not report data for the waitlist control group for the staff-rated version of this outcome measure
[4] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome measures were parent-rated and parents were non-blind
[5] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[6] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome assessors (self-completed or researcher) were non-blind
[7] Moderate to substantial heterogeneity
[8] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as self-rated
[9] High risk of performance and response bias as intervention administrator and participants were non-blind, and high risk of detection bias as although the rater of the CGI was blind this measure was based on interview with parents who were non-blind
[10] Events<300
[11] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome assessors (researchers) were non-blind and high levels of variability for this outcome measure were dealt with by administering the test twice at each time point and taking the average score

Social skills group modified for autism versus standard social skills group for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

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 Standard social skills group	**With Social skills group specifically modified for individuals with high-functioning autism**		**Risk with Standard social skills group**	**Risk difference with Social skills group specifically modified for individuals with high-functioning autism** (95% CI)
Social skills (measured with: Social Responsiveness Scale (SRS): Social Awareness (standardized change score); Better indicated by lower values)											
50 (1 study) 19 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	26	24	N/A	N/A	The mean social skills in the intervention groups was **0.68 standard deviations lower** (1.26 to 0.11 lower)
Social skills (measured with: Social Responsiveness Scale (SRS): Social Cognition (standardized change score); Better indicated by lower values)											
50 (1 study) 19 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	26	24	N/A	N/A	The mean social skills in the intervention groups was **0.33 standard deviations lower** (0.89 lower to 0.23 higher)
Social skills (measured with: Social Responsiveness Scale (SRS): Social Communication (standardized change score); Better indicated by lower values)											
50 (1 study)	serious[1]	no serious	no serious	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2]	26	24	N/A	N/A	The mean social skills in the intervention groups

19 weeks		inconsistency	indirectness			due to risk of bias, imprecision					was **0.93 standard deviations lower** (1.52 to 0.34 lower)
Social skills (measured with: Social Responsiveness Scale (SRS): Social Motivation (standardized change score); Better indicated by lower values)											
50 (1 study) 19 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	26	24	N/A	N/A	The mean social skills in the intervention groups was **0.66 standard deviations lower** (1.23 to 0.08 lower)
Social skills (measured with: Social Responsiveness Scale (SRS): Autistic Mannerisms (standardized change score); Better indicated by lower values)											
50 (1 study) 19 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	26	24	N/A	N/A	The mean social skills in the intervention groups was **0.67 standard deviations lower** (1.24 to 0.1 lower)
Social self-efficacy (self-rated) (measured with: Social Self-efficacy Scale (standardized change score); Better indicated by lower values)											
52 (1 study) 19 weeks	serious[4]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to risk of bias, imprecision	28	24	N/A	N/A	The mean social self-efficacy (self-rated) in the intervention groups was **0.12 standard deviations lower** (0.67 lower to 0.42 higher)
Feelings of loneliness (measured with: Social Dissatisfaction Questionnaire (standardized change score); Better indicated by lower values)											
52 (1 study)	serious[4]	no serious	no serious	very	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4]	28	24	N/A	N/A	The mean feelings of loneliness in the

19 weeks		inconsistency	indirectness	serious[3]		due to risk of bias, imprecision				intervention groups was **0.15 standard deviations higher** (0.4 lower to 0.69 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as parent-completed and parents were non-blind and involved in the intervention
[2] N<400
[3] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[4] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome measure self-rated

1.3 PSYCHOSOCIAL INTERVENTIONS AIMED AT THE CORE AUTISM FEATURE OF RESTRICTED INTERESTS AND RIGID AND REPETITIVE BEHAVIOURS

1.3.1 Behavioural interventions aimed at the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

ESDM or P-ESDM versus treatment-as-usual for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

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 Treatment-as-usual	**With ESDM or P-ESDM**		**Risk with Treatment-as-usual**	**Risk difference with ESDM or P-ESDM** (95% CI)
Repetitive behaviour (ESDM or P-ESDM) (measured with: Repetitive Behavior Scale (RBS): Total or Autism Diagnostic Observation Schedule for Toddlers (ADOS-T): Restricted, Repetitive Behaviours; Better indicated by lower values)											
143 (2 studies) 12-104 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	70	73	N/A	N/A	The mean repetitive behaviour (esdm or p-esdm) in the intervention groups was **0.06 standard deviations lower**

										(0.39 lower to 0.27 higher)
Repetitive behaviour (ESDM) (measured with: Repetitive Behavior Scale (RBS): Total; Better indicated by lower values)										
45 (1 study) 104 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to risk of bias, imprecision	21	24	N/A	N/A	The mean repetitive behaviour (esdm) in the intervention groups was **0.35 standard deviations lower** (0.95 lower to 0.24 higher)
Repetitive behaviour (P-ESDM) (measured with: Autism Diagnostic Observation Schedule for Toddlers (ADOS-T): Restricted, Repetitive Behaviours; Better indicated by lower values)										
98 (1 study) 12 weeks	serious[5]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,5] due to risk of bias, imprecision	49	49	N/A	N/A	The mean repetitive behaviour (p-esdm) in the intervention groups was **0.07 standard deviations higher** (0.32 lower to 0.47 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and unclear/unknown risk of detection bias as blinding ofoutcome assessors was either not reported or the outcome measure was parent-completed and parents were non-blind and involved in the intervention
[2] N<400
[3] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as this outcome measure was parent-completed and parents were non-blind and involved in the intervention
[4] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[5] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear/unknown as the outcome assessor reported as 'laboratory personnel' with no detail regarding blinding of outcome assessors reported

1.3.2 Cognitive intervention aimed at the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

Enhanced ERT versus standard ERT for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

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 Enhanced emotion recognition training (DVD-based) versus standard emotion recognition training (DVD-based) for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome		Risk with Control	Risk difference with Enhanced emotion recognition training (DVD-based) versus standard emotion recognition training (DVD-based) for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome (95% CI)
Stereotyped behaviour (measured with: Social Communication Questionnaire (SCQ): Stereotyped behaviour; Better indicated by lower values)											
25 (1 study) 3 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	12	13	N/A	N/A	The mean stereotyped behaviour in the intervention groups was **0.31 standard deviations lower** (1.1 lower to 0.48 higher)

[1] High risk of performance and detection bias as parents were non-blind and were intervention administrators and outcome assessors
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.3.3 Parent training intervention aimed at the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

Combined parent training and antipsychotic versus antipsychotic-only for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

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 Combined antipsychotic and parent training versus antipsychotic only for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome		Risk with Control	Risk difference with Combined antipsychotic and parent training versus antipsychotic only for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome (95% CI)
Compulsions (measured with: Children's Yale-Brown Obsessive Compulsive Scales-PDD (CYBOCS-PDD): Compulsions; Better indicated by lower values)											
95 (1 study) 24 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to risk of bias, imprecision	40	55	N/A	N/A	The mean compulsions in the intervention groups was **0.42 standard deviations lower** (0.83 to 0.01 lower)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear/unknown as outcome measure based on interview, but unclear who the interviewee is but if parental interview then non-blind. There was also a high risk of attrition bias due to higher dropout rates in the experimental (combined risperidone and parent training) group (N=20; 27% attrition) than the control (risperidone only) group (N=9; 18% attrition)
[2] N<400

1.3.4 Social-communication intervention aimed at the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

Caregiver-mediated social-communication intervention (PACT) versus treatment-as-usual for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

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 Treatment-as-usual	**With Caregiver-mediated social-communication intervention (PACT)**		**Risk with Treatment-as-usual**	**Risk difference with Caregiver-mediated social-communication intervention (PACT)** (95% CI)
Repetitive behaviours (measured with: Autism Diagnostic Observation Schedule-Generic (ADOS-G): Repetitive Behaviours; Better indicated by lower values)											
152 (1 study) 56 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⨁⨁⨀⨀ **LOW**[1] due to imprecision	75	77	N/A	N/A	The mean repetitive behaviours in the intervention groups was **0.3 standard deviations lower** (0.62 lower to 0.02 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.4 PHARMACOLOGICAL INTERVENTIONS AIMED AT CORE FEATURES OF AUTISM (OVERALL AUTISTIC BEHAVIOURS)

1.4.1 Anticonvulsants for overall autistic behaviours as an indirect outcome

Divalproex sodium versus placebo for overall autistic behaviours as an indirect outcome

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 Placebo	**With Anticonvulsants**		**Risk with Placebo**	**Risk difference with Anticonvulsants** (95% CI)
Overall autistic behaviours (global improvement) (assessed with: Positive treatment response (number of participants 'much improved/very improved' on Clinical Global Impression-Improvement [CGI-I]: Autism))											
27 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	0/11 (0%)	2/16 (12.5%)	**RR 3.53** (0.19 to 67.1)	**Study population**	
										0 per 1000	NA
										Moderate	
										0 per 1000	NA

[1] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)

1.4.2 Antidepressants for overall autistic behaviours as an indirect outcome

Fluoxetine versus placebo for overall autistic behaviours as an indirect outcome

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 Placebo	With Antidepressant		Risk with Placebo	Risk difference with Antidepressant (95% CI)
Overall autistic behaviours (global improvement) (measured with: Global Autism Composite Improvement (Clinical Global Improvement Scale Adapted to Global Autism [CGI-AD] and Children's Yale-Brown Obsessive-Compulsion Scale [CYBOCS] compulsions subscale change score); Better indicated by lower values)											
39 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	20	19	N/A	N/A	The mean overall autistic behaviours (global improvement) in the intervention groups was **0.35 standard deviations lower** (0.98 lower to 0.28 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.4.3 Antihistamines for overall autistic behaviours as an indirect outcome

Cyproheptadine and haloperidol versus placebo and haloperidol for overall autistic behaviours as an indirect outcome

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 Combined antipsychotic and placebo	**With Combined antihistamine and antipsychotic**		**Risk with Combined antipsychotic and placebo**	**Risk difference with Combined antihistamine and antipsychotic** (95% CI)
Overall autistic behaviours (measured with: Childhood Autism Rating Scale (CARS): Total [change score]; Better indicated by lower values)											
40 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	20	20	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.96 standard deviations lower** (1.62 to 0.3 lower)

[1] N<400

1.4.4 Antipsychotics for overall autistic behaviours as a direct or indirect outcome

Risperidone versus placebo for overall autistic behaviours as a direct or indirect outcome

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 Antipsychotics versus placebo for overall autistic behaviours**		**Risk with Control**	**Risk difference with Antipsychotics versus placebo for overall autistic behaviours** (95% CI)
Overall autistic behaviours (assessed with: Dichotomous: Positive treatment response (>20% improvement on CARS))											
39 (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	0/20 (0%)	12/19 (63.2%)	**RR 26.25** (1.66 to 414.57)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Overall autistic behaviours (assessed with: Dichotomous: Positive treatment response (>20% improvement on Children's Global Assessment Scale))											
39 (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	2/20 (10%)	17/19 (89.5%)	**RR 8.95** (2.38 to 33.62)	**Study population**	
										100 per 1000	**795 more per 1000** (from 138 more to 1000 more)
										Moderate	
										100 per 1000	**795 more per 1000** (from 138 more to 1000 more)
Overall autistic behaviours (measured with: Childhood Autism Rating Scale (CARS): Total or Ritvo-Freeman Real-life Rating Scale (RLRS): Total; Better indicated by lower											

values)										
124 (2 studies) 8-24 weeks	no serious risk of bias	very serious[3]	no serious indirectness	serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to inconsistency, imprecision	64	60	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.87 standard deviations lower** (1.25 to 0.5 lower)
Overall autistic behaviours (direct outcome) (measured with: Childhood Autism Rating Scale (CARS): Total; Better indicated by lower values)										
23 (1 study) 24 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[6]	undetected	⊕⊖⊖⊖ **VERY LOW**[5,6] due to risk of bias, imprecision	12	11	N/A	N/A	The mean overall autistic behaviours (direct outcome) in the intervention groups was **0.31 standard deviations higher** (0.51 lower to 1.14 higher)
Overall autistic behaviours (indirect outcome) (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Total; Better indicated by lower values)										
101 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	52	49	N/A	N/A	The mean overall autistic behaviours (indirect outcome) in the intervention groups was **1.19 standard deviations lower** (1.61 to 0.76 lower)

[1] Events<300
[2] High risk of selective reporting bias as mean and standard deviation data were not reported for continuous scale outcome measures
[3] Substantial to considerable heterogeneity with an I-squared value of 90%
[4] N<400
[5] High risk of selection bias as the allocation was unconcealed and the groups were not comparable at baseline for this outcome measure (the risperidone group showed significantly greater severity of autism symptoms as measured by the CARS)
[6] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Risperidone versus haloperidol for overall autistic behaviours as a direct outcome

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 Risperidone versus haloperidol for overall autistic behaviours as a direct outcome		Risk with Control	Risk difference with Risperidone versus haloperidol for overall autistic behaviours as a direct outcome (95% CI)
Overall autistic behaviours (measured with: Turgay DSM-IV PDD Rating Scale; Better indicated by lower values)											
28 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	15	13	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.35 standard deviations lower** (1.1 lower to 0.4 higher)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Social; Better indicated by lower values)											
28 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	15	13	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.26 standard deviations lower** (1 lower to 0.49 higher)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Motor; Better indicated by lower values)											
28 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	15	13	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.34 standard deviations lower** (1.09 lower to 0.41 higher)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Affective; Better indicated by lower values)											
28 (1 study)	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2]	15	13	N/A	N/A	The mean overall autistic behaviours in the

12 weeks						due to risk of bias, imprecision					intervention groups was **0.23 standard deviations lower** (0.98 lower to 0.52 higher)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Sensory; Better indicated by lower values)											
28 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	15	13	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.17 standard deviations lower** (0.92 lower to 0.57 higher)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Language; Better indicated by lower values)											
28 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	15	13	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.22 standard deviations higher** (0.53 lower to 0.96 higher)

[1] Paper states 'double-blind' but gives no further detail with regards to who is blinded, i.e. participant, parent, investigator, intervention administrator, outcome assessor
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.4.5 SNRIs for overall autistic behaviours as an indirect outcome

Atomoxetine versus placebo for overall autistic behaviours as an indirect outcome

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 Placebo	**With Selective noradrenaline reuptake inhibitors**		**Risk with Placebo**	**Risk difference with Selective noradrenaline reuptake inhibitors** (95% CI)
Overall autistic behaviours (measured with: Children's Social Behavior Questionnaire (CSBQ): Total; Better indicated by lower values)											

89 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	46	43	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.27 standard deviations lower** (0.68 lower to 0.15 higher)
[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)											

1.5 PHARMACOLOGICAL INTERVENTIONS AIMED AT THE CORE AUTISM FEATURE OF IMPAIRED RECIPROCAL SOCIAL COMMUNICATION AND INTERACTION

1.5.1 Antioxidants for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

N-acetylcysteine versus placebo for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

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 Placebo	**With Antioxidants**		**Risk with Placebo**	**Risk difference with Antioxidants** (95% CI)
Social impairment (measured with: Social Responsiveness Scale (SRS): Total; Better indicated by lower values)											
29 (1 study)	no serious	no serious	no serious	very	undetected	⊕⊕⊖⊖ **LOW**[1]	15	14	N/A	N/A	The mean social impairment in the intervention groups

12 weeks	risk of bias	inconsistency	indirectness	serious[1]		due to imprecision				was **0.14 standard deviations lower** (0.87 lower to 0.59 higher)
Social Awareness (measured with: Social Responsiveness Scale (SRS): Social Awareness ; Better indicated by lower values)										
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	14	N/A	N/A	The mean social awareness in the intervention groups was **0.45 standard deviations lower** (1.19 lower to 0.29 higher)
Social Cognition (measured with: Social Responsiveness Scale (SRS): Social Cognition ; Better indicated by lower values)										
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	14	N/A	N/A	The mean social cognition in the intervention groups was **0.02 standard deviations lower** (0.74 lower to 0.71 higher)
Social Communication (measured with: Social Responsiveness Scale (SRS): Social Communication ; Better indicated by lower values)										
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	14	N/A	N/A	The mean social communication in the intervention groups was **0.09 standard deviations lower** (0.82 lower to 0.64 higher)
Social Motivation (measured with: Social Responsiveness Scale (SRS): Social Motivation ; Better indicated by lower values)										
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to	15	14	N/A	N/A	The mean social motivation in the intervention groups was **0.24 standard deviations lower**

						imprecision				(0.97 lower to 0.49 higher)
Autistic Mannerisms (measured with: Social Responsiveness Scale (SRS): Autistic Mannerisms ; Better indicated by lower values)										
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15 14	N/A	N/A	The mean autistic mannerisms in the intervention groups was **0.64 standard deviations lower** (1.39 lower to 0.11 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.6 PHARMACOLOGICAL INTERVENTIONS AIMED AT THE CORE AUTISM FEATURE OF RESTRICTED INTERESTS AND RIGID AND REPETITIVE BEHAVIOURS

1.6.1 Antidepressants for the core autism feature of restricted interests and rigid and repetitive behaviours as a direct outcome

SSRIs versus placebo for the core autism feature of restricted interests and rigid and repetitive behaviours as a direct outcome

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 Placebo	**With Antidepressants**		**Risk with Placebo**	**Risk difference with Antidepressants** (95% CI)

Global positive treatment response (assessed with: Dichotomous: Positive treatment response ('much improved/very improved' on CGI-improvement))											
149 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	26/76 (34.2%)	24/73 (32.9%)	**RR 0.96** (0.61 to 1.51)	**Study population**	
										342 per 1000	**14 fewer per 1000** (from 133 fewer to 174 more)
										Moderate	
										342 per 1000	**14 fewer per 1000** (from 133 fewer to 174 more)
Global positive treatment response (assessed with: Dichotomous: Positive treatment response (>25% improvement on CYBOCS-PDD & 'much improved/very improved' on CGI-improvement))											
149 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	10/76 (13.2%)	15/73 (20.5%)	**RR 1.56** (0.75 to 3.25)	**Study population**	
										132 per 1000	**74 more per 1000** (from 33 fewer to 296 more)
										Moderate	
										132 per 1000	**74 more per 1000** (from 33 fewer to 297 more)
Compulsions (measured with: Children's Yale-Brown Obsessive Compulsive Scales-PDD (CYBOCS-PDD): Compulsions or Children's Yale-Brown Obsessive Compulsive Scale (CYBOCS): Compulsions; Better indicated by lower values)											
188 (2 studies) 8-12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	96	92	N/A	N/A	The mean compulsions in the intervention groups was **0.08 standard deviations lower** (0.36 lower to 0.21

											higher)
Compulsive (measured with: Repetitive Behavior Scale (RBS): Compulsive; Better indicated by lower values)											
149 (1 study) 12 days	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	76	73	N/A	N/A	The mean compulsive in the intervention groups was **0.09 standard deviations higher** (0.23 lower to 0.42 higher)
Restrictive (measured with: Repetitive Behavior Scale (RBS): Restrictive; Better indicated by lower values)											
149 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	76	73	N/A	N/A	The mean restrictive in the intervention groups was **0.34 standard deviations higher** (0.01 to 0.66 higher)
Ritualistic (measured with: Repetitive Behavior Scale (RBS): Ritualistic; Better indicated by lower values)											
149 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	76	73	N/A	N/A	The mean ritualistic in the intervention groups was **0 standard deviations higher** (0.32 lower to 0.32 higher)
Sameness (measured with: Repetitive Behavior Scale (RBS): Sameness; Better indicated by lower values)											
149 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	76	73	N/A	N/A	The mean sameness in the intervention groups was **0.05 standard deviations higher**

										(0.27 lower to 0.37 higher)
Self-injurious (measured with: Repetitive Behavior Scale (RBS): Self-injurious; Better indicated by lower values)										
149 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	76	73	N/A	N/A	The mean self-injurious in the intervention groups was **0.15 standard deviations higher** (0.17 lower to 0.47 higher)
Stereotyped (measured with: Repetitive Behavior Scale (RBS): Stereotyped; Better indicated by lower values)										
149 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	76	73	N/A	N/A	The mean stereotyped in the intervention groups was **0.13 standard deviations higher** (0.2 lower to 0.45 higher)

[1] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)
[2] N<400

1.6.2 Antioxidants for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

N-acetylcysteine versus placebo for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

Quality assessment							Summary of Findings		
Participants	**Risk of**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication**	**Overall quality**	**Study event rates**	**Relative**	**Anticipated absolute effects**

(studies) Follow up	bias				bias	of evidence	(%)		effect (95% CI)		
							With Placebo	With Antioxidants		Risk with Placebo	Risk difference with Antioxidants (95% CI)
Compulsions (measured with: Repetitive Behavior Scale (RBS): Compulsions; Better indicated by lower values)											
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	14	N/A	N/A	The mean compulsions in the intervention groups was **0.68 standard deviations lower** (1.43 lower to 0.08 higher)
Restricted (measured with: Repetitive Behavior Scale (RBS): Restricted; Better indicated by lower values)											
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	14	N/A	N/A	The mean restricted in the intervention groups was **0.42 standard deviations lower** (1.15 lower to 0.32 higher)
Rituals (measured with: Repetitive Behavior Scale (RBS): Rituals; Better indicated by lower values)											
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	14	N/A	N/A	The mean rituals in the intervention groups was **0.3 standard deviations lower** (1.03 lower to 0.44 higher)
Sameness (measured with: Repetitive Behavior Scale (RBS): Sameness; Better indicated by lower values)											
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	14	N/A	N/A	The mean sameness in the intervention groups was **0.46 standard deviations lower** (1.2 lower to 0.28 higher)

Self-injurious behaviour (measured with: Repetitive Behavior Scale (RBS): Self-injurious behaviour; Better indicated by lower values)											
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	14	N/A	N/A	The mean self-injurious behaviour in the intervention groups was **0.26 standard deviations lower** (0.99 lower to 0.48 higher)
Stereotypic behaviour (measured with: Repetitive Behavior Scale (RBS): Stereotypies; Better indicated by lower values)											
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	14	N/A	N/A	The mean stereotypic behaviour in the intervention groups was **0.51 standard deviations lower** (1.25 lower to 0.24 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.6.3 Antipsychotics for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

Antipsychotics versus placebo for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

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 Antipsychotics versus placebo for the core autism feature of restricted interests and rigid and repetitive**		**Risk with Control**	**Risk difference with Antipsychotics versus placebo for the core autism feature of restricted interests and rigid and**

							behaviours				**repetitive behaviours** (95% CI)
Compulsions (risperidone or aripiprazole) (measured with: Children's Yale-Brown Obsessive Compulsive Scale (CYBOCS): Compulsions (Endpoint or Change Score); Better indicated by lower values)											
385 (3 studies) 6-8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	130	255	N/A	N/A	The mean compulsions (risperidone or aripiprazole) in the intervention groups was **0.42 standard deviations lower** (0.64 to 0.2 lower)
Compulsions (risperidone) (measured with: Children's Yale-Brown Obsessive Compulsive Scale (CYBOCS): Compulsions; Better indicated by lower values)											
193 (2 studies) 6-8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	86	107	N/A	N/A	The mean compulsions (risperidone) in the intervention groups was **0.49 standard deviations lower** (0.79 to 0.20 lower)
Compulsions (aripiprazole) (measured with: Children's Yale-Brown Obsessive Compulsive Scale (CYBOCS): Compulsions (Change Score); Better indicated by lower values)											
192 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	44	148	N/A	N/A	The mean compulsions (aripiprazole) in the intervention groups was **0.31 standard deviations lower** (0.65 lower to 0.03 higher)

[1] N<400
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Low-dose antipsychotics versus placebo for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

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 Low dose antipsychotics versus placebo for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome**		**Risk with Control**	**Risk difference with Low dose antipsychotics versus placebo for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome** (95% CI)
Compulsions (risperidone or aripiprazole) (measured with: Children's Yale-Brown Obsessive Compulsive Scale (CYBOCS): Compulsions (Endpoint or Change Score); Better indicated by lower values)											
153 (2 studies) 6-8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	78	75	N/A	N/A	The mean compulsions (risperidone or aripiprazole) in the intervention groups was **0.27 standard deviations lower** (0.59 lower to 0.04 higher)
Compulsions (risperidone) (measured with: Children's Yale-Brown Obsessive Compulsive Scale (CYBOCS): Compulsions; Better indicated by lower values)											
63 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	34	29	N/A	N/A	The mean compulsions (risperidone) in the intervention groups was **0.29 standard deviations lower** (0.79 lower to 0.21 higher)
Compulsions (aripiprazole) (measured with: Children's Yale-Brown Obsessive Compulsive Scale (CYBOCS): Compulsions (Change Score); Better indicated by lower values)											

90 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	44	46		N/A	N/A	The mean compulsions (aripiprazole) in the intervention groups was **0.27 standard deviations lower** (0.68 lower to 0.15 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.7 BIOMEDICAL INTERVENTIONS AIMED AT CORE FEATURES OF AUTISM (OVERALL AUTISTIC BEHAVIOURS)

1.7.1 Complementary therapies for overall autistic behaviours as a direct or indirect outcome

Acupressure versus waitlist for overall autistic behaviours as a direct outcome

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 Waitlist	**With Acupressure**		**Risk with Waitlist**	**Risk difference with Acupressure** (95% CI)
Overall autistic behaviours (measured with: Study-specific parent-rated questionnaire: Total score; Better indicated by lower values)											
32 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	16	16	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.92 standard deviations higher** (0.19 to 1.66 higher)
Overall autistic behaviours (measured with: Study-specific parent-rated questionnaire: Language; Better indicated by lower values)											
32 (1 study)	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖	16	16	N/A	N/A	The mean overall autistic behaviours in the intervention

6 weeks						**LOW**[1,2] due to risk of bias, imprecision					groups was **1.33 standard deviations higher** (0.55 to 2.1 higher)
Overall autistic behaviours (measured with: Study-specific parent-rated questionnaire: Social interaction; Better indicated by lower values)											
32 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	16	16	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.98 standard deviations higher** (0.24 to 1.72 higher)
Overall autistic behaviours (measured with: Study-specific parent-rated questionnaire: Social interaction; Better indicated by lower values)											
32 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	16	16	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.23 standard deviations higher** (0.47 lower to 0.92 higher)
Overall autistic behaviours (measured with: tudy-specific parent-rated questionnaire: Motor functioning; Better indicated by lower values)											
32 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	16	16	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.45 standard deviations higher** (0.25 lower to 1.15 higher)

[1] High risk of performance and response bias as participants and intervention administrators were non-blind, and high risk of detection bias as outcome measure was parent-rated and parents were non-blind
[2] N<400
[3] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Acupuncture/electro-acupuncture and conventional educational programme versus conventional educational programme only for overall autistic behaviours as a direct outcome

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 Conventional educational programme only	With Acupuncture/electro-acupuncture and conventional educational programme		Risk with Conventional educational programme only	Risk difference with Acupuncture/electro-acupuncture and conventional educational programme (95% CI)
Overall autistic behaviours (measured with: Autism Evaluation Treatment Checklist (ATEC): Total; Better indicated by lower values)											
36 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	18	18	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.25 standard deviations higher** (0.41 lower to 0.9 higher)
Overall autistic behaviours (measured with: Autism Evaluation Treatment Checklist (ATEC): Speech/Language/Communication; Better indicated by lower values)											
36 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	18	18	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.06 standard deviations lower** (0.71 lower to 0.59 higher)
Overall autistic behaviours (measured with: Autism Evaluation Treatment Checklist (ATEC): Sociability; Better indicated by lower values)											
36 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	18	18	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.14 standard deviations higher** (0.51 lower to 0.8 higher)

Overall autistic behaviours (measured with: Autism Evaluation Treatment Checklist (ATEC): Sensory/Cognitive Awareness; Better indicated by lower values)											
36 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	18	18	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.42 standard deviations higher** (0.24 lower to 1.08 higher)
Overall autistic behaviours (measured with: Autism Evaluation Treatment Checklist (ATEC): Physical health & behaviour; Better indicated by lower values)											
36 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	18	18	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.18 standard deviations higher** (0.47 lower to 0.84 higher)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Total; Better indicated by lower values)											
65 (2 studies) 8 weeks	serious[1]	serious[3]	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, inconsistency, imprecision	32	33	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.28 standard deviations higher** (0.21 lower to 0.77 higher)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Motor; Better indicated by lower values)											
66 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	33	33	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.16 standard deviations higher** (0.33 lower to 0.64

										higher)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Social; Better indicated by lower values)										
66 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	33	33	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.2 standard deviations lower** (0.69 lower to 0.28 higher)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Affective; Better indicated by lower values)										
66 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	33	33	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.17 standard deviations higher** (0.32 lower to 0.66 higher)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Sensory; Better indicated by lower values)										
66 (2 studies) 8 weeks	serious[1]	serious[3]	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, inconsistency, imprecision	33	33	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.12 standard deviations higher** (0.36 lower to 0.61 higher)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Language; Better indicated by lower values)										
66 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	33	33	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.35 standard**

											deviations higher (0.13 lower to 0.84 higher)
Overall autistic behaviours (measured with: Clinical Global Impression Scale (CGI): Total; Better indicated by lower values)											
36 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⨁⨁⨀⨀ **LOW**[1,4] due to risk of bias, imprecision	18	18	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.9 standard deviations lower** (1.58 to 0.21 lower)
Overall autistic behaviours (measured with: Clinical Global Impression Scale (CGI): Response to social interaction; Better indicated by lower values)											
30 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⨁⨀⨀⨀ **VERY LOW**[1,2] due to risk of bias, imprecision	15	15	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.2 standard deviations lower** (0.91 lower to 0.52 higher)
Overall autistic behaviours (measured with: Clinical Global Impression Scale (CGI): Social initiation; Better indicated by lower values)											
30 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⨁⨀⨀⨀ **VERY LOW**[1,2] due to risk of bias, imprecision	15	15	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.1 standard deviations lower** (0.81 lower to 0.62 higher)
Overall autistic behaviours (measured with: Clinical Global Impression Scale (CGI): Use of speech; Better indicated by lower values)											
30 (1 study) 8 weeks	N/A	N/A	N/A	N/A	N/A	N/A	15	15	N/A	N/A	Not estimable

Overall autistic behaviours (measured with: Clinical Global Impression Scale (CGI): Repetitive behaviour; Better indicated by lower values)											
30 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[1,4] due to risk of bias, imprecision	15	15	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **1.11 standard deviations lower** (1.88 to 0.33 lower)
Overall autistic behaviours (measured with: Clinical Global Impression Scale (CGI): Behaviour problem; Better indicated by lower values)											
30 (1 study) 8 weeks	N/A	N/A	N/A	N/A	N/A	N/A	15	15	N/A	N/A	Not estimable
Overall autistic behaviours (measured with: Clinical Global Impression Scale (CGI): Activity level; Better indicated by lower values)											
30 (1 study) 8 weeks	N/A	N/A	N/A	N/A	N/A	N/A	15	15	N/A	N/A	Not estimable
Overall autistic behaviours (measured with: Clinical Global Impression Scale (CGI): Sleep problem; Better indicated by lower values)											
30 (1 study) 8 weeks	N/A	N/A	N/A	N/A	N/A	N/A	15	15	N/A	N/A	Not estimable
Overall autistic behaviours (measured with: Clinical Global Impression Scale (CGI): Digestive problem; Better indicated by lower values)											
30 (1 study) 8 weeks	N/A	N/A	N/A	N/A	N/A	N/A	15	15	N/A	N/A	Not estimable

[1] High risk of performance and response bias as intervention administrators and participants were non-blind and potential for care confounds as the conventional education programme differed for each participant which may introduce bias. There was also an unclear risk of detection bias as although all outcomes were measured by blinded assessors, some outcomes involved input from parents who were not blind to treatment allocation or confounding variables and systematic review from which data was extracted does not report which outcome measures relied on non-blind parental report
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] Moderate to substantial heterogeneity
[4] N<400

Acupuncture/electro-acupuncture versus sham acupuncture/electro-acupuncture for overall autistic behaviours as an indirect outcome

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 Sham acupuncture/electro-acupuncture	With Acupuncture/electro-acupuncture		Risk with Sham acupuncture/electro-acupuncture	Risk difference with Acupuncture/electro-acupuncture (95% CI)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Total (change scores); Better indicated by lower values)											
105 (2 studies) 4-9 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	50	55	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.3 standard deviations lower** (0.69 lower to 0.09 higher)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Motor (change scores); Better indicated by lower values)											
105 (2 studies) 4-9 weeks	no serious risk of bias	serious[3]	no serious indirectness	serious[4]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,3,4] due to inconsistency, imprecision, publication bias	50	55	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.11 standard deviations lower** (0.49 lower to 0.28 higher)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Social (change scores); Better indicated by lower values)											
105 (2 studies) 4-9 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	50	55	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.16 standard deviations lower** (0.55 lower to 0.22

											higher)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Affective (change scores); Better indicated by lower values)											
105 (2 studies) 4-9 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	50	55	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.27 standard deviations lower** (0.66 lower to 0.11 higher)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Sensory (change scores); Better indicated by lower values)											
105 (2 studies) 4-9 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [2]	⊕⊕⊖⊖ **LOW**[2,4] due to imprecision, publication bias	50	55	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.1 standard deviations lower** (0.48 lower to 0.29 higher)
Overall autistic behaviours (measured with: Ritvo-Freeman Real-life Rating Scale (RLRS): Language (change scores); Better indicated by lower values)											
105 (2 studies) 4-9 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	50	55	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.32 standard deviations lower** (0.7 lower to 0.07 higher)
Positive treatment response (assessed with: Number of participants showing much improvement on CGI-I for autistic behaviours)											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication	1/25 (4%)	7/30 (23.3%)	**RR 5.83** (0.77 to 44.28)	**Study population**	
										40 per 1000	**193 more per 1000** (from 9 fewer to

						bias					1000 more)
										Moderate	
										40 per 1000	**193 more per 1000** (from 9 fewer to 1000 more)
Positive treatment response (assessed with: Number of participants showing minimal improvement on CGI-I for autistic behaviours)											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	14/25 (56%)	20/30 (66.7%)	**RR 1.19** (0.77 to 1.83)	**Study population**	
										560 per 1000	**106 more per 1000** (from 129 fewer to 465 more)
										Moderate	
										560 per 1000	**106 more per 1000** (from 129 fewer to 465 more)
Positive treatment response for social relatedness (assessed with: Dichotomous: Positive treatment response for social relatedness - Social response (study-specific parent-reported 'better than before'))											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	5/25 (20%)	4/30 (13.3%)	**RR 0.67** (0.2 to 2.22)	**Study population**	
										200 per 1000	**66 fewer per 1000** (from 160 fewer to 244 more)
										Moderate	
										200 per 1000	**66 fewer per 1000** (from 160 fewer to 244 more)
Positive treatment response for social relatedness (assessed with: Dichotomous: Positive treatment response for social relatedness - Social initiation (study-specific											

parent-reported 'better than before'))											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	0/25 (0%)	7/30 (23.3%)	**RR 12.58** (0.75 to 209.98)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**
Positive treatment response for social relatedness (assessed with: Dichotomous: Positive treatment response for social relatedness - Eye contact (study-specific parent-reported 'better than before'))											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	4/25 (16%)	7/30 (23.3%)	**RR 1.46** (0.48 to 4.42)	**Study population**	
										160 per 1000	**74 more per 1000** (from 83 fewer to 547 more)
										Moderate	
										160 per 1000	**74 more per 1000** (from 83 fewer to 547 more)
Positive treatment response for social relatedness (assessed with: Dichotomous: Positive treatment response for social relatedness - Share (study-specific parent-reported 'better than before'))											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	1/25 (4%)	0/30 (0%)	**RR 0.28** (0.01 to 6.58)	**Study population**	
										40 per 1000	**29 fewer per 1000** (from 40 fewer to 223 more)
										Moderate	
										40 per 1000	**29 fewer per 1000** (from 40 fewer to 223 more)
Positive treatment response for social relatedness (assessed with: Dichotomous: Positive treatment response for social relatedness - Curiosity (study-specific											

parent-reported 'better than before'))											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	1/25 (4%)	0/30 (0%)	**RR 0.28** (0.01 to 6.58)	**Study population**	
										40 per 1000	**29 fewer per 1000** (from 40 fewer to 223 more)
										Moderate	
										40 per 1000	**29 fewer per 1000** (from 40 fewer to 223 more)
Positive treatment response for social relatedness (assessed with: Dichotomous: Positive treatment response for social relatedness - Patience (study-specific parent-reported 'better than before'))											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	0/25 (0%)	1/30 (3.3%)	**RR 2.52** (0.11 to 59.18)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**
Positive treatment response for non-verbal and verbal communication (assessed with: Dichotomous: Positive treatment response for non-verbal and verbal communication - Expressive language (study-specific parent-reported 'better than before'))											
54 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	7/24 (29.2%)	11/30 (36.7%)	**RR 1.26** (0.58 to 2.75)	**Study population**	
										292 per 1000	**76 more per 1000** (from 123 fewer to 510 more)
										Moderate	
										292 per 1000	**76 more per 1000** (from 123 fewer to 511 more)
Positive treatment response for non-verbal and verbal communication (assessed with: Dichotomous: Positive treatment response for non-verbal and											

verbal communication - Receptive language (study-specific parent-reported 'better than before'))											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[6]	reporting bias strongly suspected [2]	⊕⊕⊖⊖ **LOW**[2,6] due to imprecision, publication bias	5/25 (20%)	17/30 (56.7%)	**RR 2.83** (1.22 to 6.59)	**Study population**	
										200 per 1000	**366 more per 1000** (from 44 more to 1000 more)
										Moderate	
										200 per 1000	**366 more per 1000** (from 44 more to 1000 more)
Positive treatment response for non-verbal and verbal communication (assessed with: Dichotomous: Positive treatment response for non-verbal and verbal communication - Pointing (study-specific parent-reported 'better than before'))											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	0/25 (0%)	1/30 (3.3%)	**RR 2.52** (0.11 to 59.18)	**Study population**	
										0 per 1000	-
										Moderate	
										0 per 1000	-
Positive treatment response for non-verbal and verbal communication (assessed with: Dichotomous: Positive treatment response for non-verbal and verbal communication - Imitation (study-specific parent-reported 'better than before'))											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	0/25 (0%)	1/30 (3.3%)	**RR 2.52** (0.11 to 59.18)	**Study population**	
										0 per 1000	-
										Moderate	
										0 per 1000	-
Positive treatment response for stereotypy interest and behaviour (assessed with: Dichotomous: Positive treatment response for stereotypy interest and behaviour - Temper (study-specific parent-reported 'better than before'))											

55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	5/25 (20%)	8/30 (26.7%)	**RR 1.33** (0.5 to 3.56)	**Study population**	
										200 per 1000	**66 more per 1000** (from 100 fewer to 512 more)
										Moderate	
										200 per 1000	**66 more per 1000** (from 100 fewer to 512 more)
Positive treatment response for stereotypy interest and behaviour (assessed with: Dichotomous: Positive treatment response for stereotypy interest and behaviour - Compulsive behaviour (study-specific parent-reported 'better than before'))											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	1/25 (4%)	1/30 (3.3%)	**RR 0.83** (0.05 to 12.66)	**Study population**	
										40 per 1000	**7 fewer per 1000** (from 38 fewer to 466 more)
										Moderate	
										40 per 1000	**7 fewer per 1000** (from 38 fewer to 466 more)
Positive treatment response for stereotypy interest and behaviour (assessed with: Dichotomous: Positive treatment response for stereotypy interest and behaviour - Adaptation to change (study-specific parent-reported 'better than before'))											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	1/25 (4%)	0/30 (0%)	**RR 0.28** (0.01 to 6.58)	**Study population**	
										40 per 1000	**29 fewer per 1000** (from 40 fewer to 223 more)
										Moderate	
										40 per 1000	**29 fewer per 1000** (from 40 fewer to

										223 more)
Positive treatment response for cognition (assessed with: Dichotomous: Positive treatment response for cognition - Memory (study-specific parent-reported 'better than before'))										
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	2/25 (8%)	1/30 (3.3%)	**RR 0.42** (0.04 to 4.33)	**Study population**
										80 per 1000 — **46 fewer per 1000** (from 77 fewer to 266 more)
										Moderate
										80 per 1000 — **46 fewer per 1000** (from 77 fewer to 266 more)
Positive treatment response for cognition (assessed with: Dichotomous: Positive treatment response for cognition - Learning ability (study-specific parent-reported 'better than before'))										
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	2/25 (8%)	2/30 (6.7%)	**RR 0.83** (0.13 to 5.5)	**Study population**
										80 per 1000 — **14 fewer per 1000** (from 70 fewer to 360 more)
										Moderate
										80 per 1000 — **14 fewer per 1000** (from 70 fewer to 360 more)
Positive treatment response for motor abnormalities (assessed with: Dichotomous: Positive treatment response for motor abnormalities - Motor skill (study-specific parent-reported 'better than before'))										
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication	0/25 (0%)	5/30 (16.7%)	**RR 9.23** (0.53 to 159.14)	**Study population**
										0 per 1000 — N/A
										Moderate

						bias				**0 per 1000**	N/A
Positive treatment response for motor abnormalities (assessed with: Dichotomous: Positive treatment response for motor abnormalities - Coordination (study-specific parent-reported 'better than before'))											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	2/25 (8%)	8/30 (26.7%)	**RR 3.33** (0.78 to 14.29)	**Study population**	
										80 per 1000	**186 more per 1000** (from 18 fewer to 1000 more)
										Moderate	
										80 per 1000	**186 more per 1000** (from 18 fewer to 1000 more)
Positive treatment response for motor abnormalities (assessed with: Dichotomous: Positive treatment response for motor abnormalities - Drooling (study-specific parent-reported 'better than before'))											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	1/25 (4%)	2/30 (6.7%)	**RR 1.67** (0.16 to 17.32)	**Study population**	
										40 per 1000	**27 more per 1000** (from 34 fewer to 653 more)
										Moderate	
										40 per 1000	**27 more per 1000** (from 34 fewer to 653 more)
Positive treatment response for other parent-reported changes (assessed with: Dichotomous: Positive treatment response for other parent-reported changes - Appetite (study-specific parent-reported 'better than before'))											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision,	1/25 (4%)	3/30 (10%)	**RR 2.5** (0.28 to 22.56)	**Study population**	
										40 per 1000	**60 more per 1000** (from 29 fewer to

						publication bias					862 more)
										Moderate	
										40 per 1000	**60 more per 1000** (from 29 fewer to 862 more)
Positive treatment response for other parent-reported changes (assessed with: Dichotomous: Positive treatment response for other parent-reported changes - Attention span (study-specific parent-reported 'better than before'))											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	0/25 (0%)	9/30 (30%)	**RR 15.94** (0.97 to 260.91)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Positive treatment response for other parent-reported changes (assessed with: Dichotomous: Positive treatment response for other parent-reported changes - Sleeping pattern (study-specific parent-reported 'better than before'))											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[2,5] due to imprecision, publication bias	3/25 (12%)	7/30 (23.3%)	**RR 1.94** (0.56 to 6.75)	**Study population**	
										120 per 1000	**113 more per 1000** (from 53 fewer to 690 more)
										Moderate	
										120 per 1000	**113 more per 1000** (from 53 fewer to 690 more)
Positive treatment response for other parent-reported changes (assessed with: Dichotomous: Positive treatment response for other parent-reported changes - "Crafty" (study-specific parent-reported 'better than before'))											
55	no	no serious	no serious	very	reporting	⊕⊖⊖⊖	1/25	2/30	**RR 1.67**	**Study population**	

(1 study) 4 weeks	serious risk of bias	inconsistency	indirectness	serious[5]	bias strongly suspected [2]	**VERY LOW**[2,5] due to imprecision, publication bias	(4%)	(6.7%)	(0.16 to 17.32)	**40 per 1000**	**27 more per 1000** (from 34 fewer to 653 more)
										Moderate	
										40 per 1000	**27 more per 1000** (from 34 fewer to 653 more)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] High risk of selective reporting bias as trial protocol for WONG2010B states that follow-up measurements will be taken but these are not reported
[3] Moderate heterogeneity
[4] N<400
[5] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)
[6] Events<300

Qigong massage training versus waitlist for overall autistic behaviours as an indirect outcome

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 Waitlist	**With Qigong massage training**		**Risk with Waitlist**	**Risk difference with Qigong massage training** (95% CI)
Overall autistic behaviours (measured with: Teacher-rated Autism Behaviour Checklist (ABC): Total or Parent-rated Pervasive Development Disorder Behavior Inventory (PDDBI): Autism Composite; Better indicated by lower values)											
79 (2 studies) 17-22 weeks	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	39	40	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.85 standard deviations lower** (1.32 to 0.39 lower)
Overall autistic behaviours (measured with: Teacher-rated Autism Behaviour Checklist (ABC): Total; Better indicated by lower values)											
46 (1 study)	serious[3]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,3]	21	25	N/A	N/A	The mean overall autistic behaviours in the intervention

22 weeks						due to risk of bias, imprecision					groups was **0.91 standard deviations lower** (1.52 to 0.3 lower)
Overall autistic behaviours (measured with: Parent-rated Pervasive Development Disorder Behavior Inventory (PDDBI): Autism Composite; Better indicated by lower values)											
33 (1 study) 17 weeks	serious[4]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,4] due to risk of bias, imprecision	18	15	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.77 standard deviations lower** (1.49 to 0.06 lower)
Social, language, and communication abilities (measured with: Teacher-rated Pervasive Development Disorder Behavior Inventory (PDDBI): Social, language and communication abilities; Better indicated by lower values)											
46 (1 study) 22 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,3] due to risk of bias, imprecision	21	25	N/A	N/A	The mean social, language, and communication abilities in the intervention groups was **0.82 standard deviations higher** (0.22 to 1.43 higher)
Social, language, and communication abilities (measured with: Parent-rated Pervasive Development Disorder Behavior Inventory (PDDBI): Social, language and communication abilities; Better indicated by lower values)											
79 (2 studies) 17-22 weeks	very serious[1]	very serious[5]	no serious indirectness	serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,5] due to risk of bias, inconsistency, imprecision	39	40	N/A	N/A	The mean social, language, and communication abilities in the intervention groups was **0.53 standard deviations higher** (0.07 to 1 higher)
Maladaptive behaviour (measured with: Teacher-rated Pervasive Development Disorder Behavior Inventory (PDDBI): Maladaptive behaviour; Better indicated by lower values)											
46 (1 study) 22 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[6]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,6] due to risk of bias, imprecision	21	25	N/A	N/A	The mean maladaptive behaviour in the intervention groups was **0.56 standard deviations**

										lower (1.16 lower to 0.03 higher)
Maladaptive behaviour (measured with: Parent-rated Pervasive Development Disorder Behavior Inventory (PDDBI): Maladaptive behaviour; Better indicated by lower values)										
79 (2 studies) 17-22 weeks	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	39	40	N/A	N/A	The mean maladaptive behaviour in the intervention groups was **1.03 standard deviations lower** (1.5 to 0.55 lower)

[1] High risk of selection bias in SILVA2009 as groups were assigned using a random number generator but there were caveats to the randomisation (five sets of siblings were co-assigned due to parental involvement in the treatment and different geographical areas were assigned separately to meet the 'therapist to participant requirements'), groups were also not comparable at baseline for measures of parent rated social communication and autism composite and teacher rated sensory problems. There was also a high risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias was high for the parent-rated outcome measure as parents were non-blind and involved in the intervention.
[2] N<400
[3] High risk of selection bias in SILVA2009 as groups were assigned using a random number generator but there were caveats to the randomisation (five sets of siblings were co-assigned due to parental involvement in the treatment and different geographical areas were assigned separately to meet the 'therapist to participant requirements'), groups were also not comparable at baseline for measures of parent rated social communication and autism composite and teacher rated sensory problems.
[4] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as parent-rated and parents were non-blind and involved in intervention
[5] Substantial to considerable heterogeneity
[6] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.7.2 Hormones for overall autistic behaviours as a direct or indirect outcome

Secretin versus placebo for overall autistic behaviours as a direct or indirect outcome

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 Placebo	**With Secretin**		**Risk with Placebo**	**Risk difference with Secretin** (95% CI)
Positive treatment response (assessed with: Dichotomous: Positive treatment response (decrease of >4.07 points CARS))											
57 (1 study) 3 weeks	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,	7/29 (24.1%)	11/28 (39.3%)	**RR 1.63** (0.74 to 3.6)	**Study population**	
										241 per	**152 more per 1000**

						imprecision, publication bias			**1000**	(from 63 fewer to 628 more)
									Moderate	
									241 per 1000	**152 more per 1000** (from 63 fewer to 627 more)
Positive treatment response (assessed with: Dichotomous: Positive treatment response (decrease of >4.07 points CARS) or Dichotomous: Positive treatment response ('much improved/very improved' on CGI-improvement))										
109 (2 studies) 4-6 weeks	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	15/54 (27.8%)	19/55 (34.5%)	**RR 1.24** (0.71 to 2.19)	**Study population**
									278 per 1000	**67 more per 1000** (from 81 fewer to 331 more)
									Moderate	
									278 per 1000	**67 more per 1000** (from 81 fewer to 331 more)
Overall autistic behaviours (measured with: Childhood Autism Rating Scale (CARS): Total (endpoint or change scores); Better indicated by lower values)										
137 (2 studies) 3-6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	71	66	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.14 standard deviations higher** (0.2 lower to 0.48 higher)
Overall autistic behaviours (measured with: Autism Behaviour Checklist (ABC): Total (change score); Better indicated by lower values)										
145 (2 studies) 1-3 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	73	72	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.09 standard deviations lower** (0.42 lower to 0.23 higher)
Overall autistic behaviours (measured with: Autism Behaviour Checklist (ABC): Total (change score); Better indicated by lower values)										
52 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	25	27	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.46 standard deviations lower** (1.01 lower to 0.1 higher)

Sensory function (measured with: Autism Behaviour Checklist (ABC): Sensory (change score); Better indicated by lower values)										
140 (2 studies) 1-3 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	70	70	N/A	N/A	The mean sensory function in the intervention groups was **0.09 standard deviations lower** (0.42 lower to 0.25 higher)
Sensory function (measured with: Autism Behaviour Checklist (ABC): Sensory (change score); Better indicated by lower values)										
52 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	25	27	N/A	N/A	The mean sensory function in the intervention groups was **0.52 standard deviations lower** (1.08 lower to 0.03 higher)
Social relatedness (measured with: Autism Behaviour Checklist (ABC): Social relatedness (change score); Better indicated by lower values)										
143 (2 studies) 1-3 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	72	71	N/A	N/A	The mean social relatedness in the intervention groups was **0.11 standard deviations lower** (0.44 lower to 0.22 higher)
Social relatedness (measured with: Autism Behaviour Checklist (ABC): Social relatedness (change score); Better indicated by lower values)										
52 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	25	27	N/A	N/A	The mean social relatedness in the intervention groups was **0.3 standard deviations lower** (0.85 lower to 0.25 higher)
Body and object use (measured with: Autism Behavior Checklist (ABC): Body and object use (change score); Better indicated by lower values)										
145 (2 studies) 1-3 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	73	72	N/A	N/A	The mean body and object use in the intervention groups was **0.05 standard deviations lower** (0.38 lower to 0.28 higher)
Body and object use (measured with: Autism Behavior Checklist (ABC): Body and object use (change score); Better indicated by lower values)										
52 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	25	27	N/A	N/A	The mean body and object use in the intervention groups was **0.11 standard deviations lower** (0.66 lower to 0.43 higher)
Language (measured with: Autism Behaviour Checklist (ABC): Language (change score); Better indicated by lower values)										
136 (2 studies) 1-3 weeks	no serious risk of bias	serious[6]	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4,6] due to	67	69	N/A	N/A	The mean language in the intervention groups was **0.01 standard deviations lower**

						inconsistency, imprecision				(0.35 lower to 0.33 higher)
Language (measured with: Autism Behaviour Checklist (ABC): Language (change score); Better indicated by lower values)										
52 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	25	27	N/A	N/A	The mean language in the intervention groups was **0.32 standard deviations lower** (0.87 lower to 0.23 higher)
Socialization (measured with: Autism Behaviour Checklist (ABC): Socialization (change score); Better indicated by lower values)										
139 (2 studies) 1-3 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	70	69	N/A	N/A	The mean socialization in the intervention groups was **0.05 standard deviations lower** (0.39 lower to 0.28 higher)
Socialization (measured with: Autism Behaviour Checklist (ABC): Socialization (change score); Better indicated by lower values)										
52 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	25	27	N/A	N/A	The mean socialization in the intervention groups was **0.25 standard deviations lower** (0.8 lower to 0.3 higher)
Overall autistic behaviours (measured with: Gilliam Autism Rating Scale (GARS): Autism Quotient; Better indicated by lower values)										
98 (2 studies) 4-6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	51	47	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.34 standard deviations higher** (0.06 lower to 0.74 higher)
Overall autistic behaviours (measured with: Gilliam Autism Rating Scale (GARS): Social Interaction; Better indicated by lower values)										
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	28	28	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.42 standard deviations higher** (0.11 lower to 0.95 higher)
Overall autistic behaviours (measured with: Gilliam Autism Rating Scale (GARS): Stereotyped behaviours; Better indicated by lower values)										
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	28	28	N/A	N/A	The mean overall autistic behaviours in the intervention groups was

										0.17 standard deviations higher (0.36 lower to 0.69 higher)
Overall autistic behaviours (measured with: Gilliam Autism Rating Scale (GARS): Communication; Better indicated by lower values)										
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	28	28	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.38 standard deviations higher** (0.15 lower to 0.9 higher)
Overall autistic behaviours (measured with: Clinical Global Impression (CGI) scale; Better indicated by lower values)										
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	28	28	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.23 standard deviations higher** (0.29 lower to 0.76 higher)
Response to social interaction (measured with: Clinical Global Impression (CGI): Response to social interaction (change score); Better indicated by lower values)										
52 (1 study) 1 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	25	27	N/A	N/A	The mean response to social interaction in the intervention groups was **0 standard deviations higher** (0.54 lower to 0.54 higher)
Response to social interaction (measured with: Clinical Global Impression (CGI): Response to social interaction (change score); Better indicated by lower values)										
49 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	24	25	N/A	N/A	The mean response to social interaction in the intervention groups was **0.34 standard deviations lower** (0.9 lower to 0.23 higher)
Social initiation (measured with: Clinical Global Impression (CGI): Social initiation (change score); Better indicated by lower values)										
52 (1 study) 1 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	25	27	N/A	N/A	The mean social initiation in the intervention groups was **0.09 standard deviations lower** (0.64 lower to 0.45 higher)
Social initiation (measured with: Clinical Global Impression (CGI): Social initiation (change score); Better indicated by lower values)										

49 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	24	25	N/A	N/A	The mean social initiation in the intervention groups was **0 standard deviations higher** (0.56 lower to 0.56 higher)
Use of speech (measured with: Clinical Global Impression (CGI): Use of speech (change score); Better indicated by lower values)											
52 (1 study) 1 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	25	27	N/A	N/A	The mean use of speech in the intervention groups was **0.2 standard deviations lower** (0.74 lower to 0.35 higher)
Use of speech (measured with: Clinical Global Impression (CGI): Use of speech (change score); Better indicated by lower values)											
49 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	24	25	N/A	N/A	The mean use of speech in the intervention groups was **0 standard deviations higher** (0.56 lower to 0.56 higher)
Types of repetitive behaviour (measured with: Clinical Global Impression (CGI): Types of repetitive behaviour (change score); Better indicated by lower values)											
52 (1 study) 1 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	25	27	N/A	N/A	The mean types of repetitive behaviour in the intervention groups was **0.18 standard deviations lower** (0.72 lower to 0.37 higher)
Types of repetitive behaviour (measured with: Clinical Global Impression (CGI): Types of repetitive behaviour (change score); Better indicated by lower values)											
49 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	24	25	N/A	N/A	The mean types of repetitive behaviour in the intervention groups was **0.26 standard deviations lower** (0.82 lower to 0.3 higher)
Behaviour problems (measured with: Clinical Global Impression (CGI): Behaviour problems (change score); Better indicated by lower values)											
52 (1 study) 1 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	25	27	N/A	N/A	The mean behaviour problems in the intervention groups was **0.4 standard deviations higher** (0.15 lower to 0.95 higher)
Behaviour problems (measured with: Clinical Global Impression (CGI): Behaviour problems (change score); Better indicated by lower values)											
49 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	24	25	N/A	N/A	The mean behaviour problems in the intervention groups was **0.42 standard deviations**

										higher (0.14 lower to 0.99 higher)
Activity level (measured with: Clinical Global Impression (CGI): Activity level (change score); Better indicated by lower values)										
52 (1 study) 1 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	25	27	N/A	N/A	The mean activity level in the intervention groups was **0.32 standard deviations higher** (0.23 lower to 0.87 higher)
Activity level (measured with: Clinical Global Impression (CGI): Activity level (change score); Better indicated by lower values)										
49 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	24	25	N/A	N/A	The mean activity level in the intervention groups was **0.08 standard deviations higher** (0.48 lower to 0.64 higher)
Sleep problems (measured with: Clinical Global Impression (CGI): Sleep problems (change score); Better indicated by lower values)										
49 (1 study) 1 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	24	25	N/A	N/A	The mean sleep problems in the intervention groups was **0.16 standard deviations higher** (0.41 lower to 0.72 higher)
Sleep problems (measured with: Clinical Global Impression (CGI): Sleep problems (change score); Better indicated by lower values)										
48 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	24	24	N/A	N/A	The mean sleep problems in the intervention groups was **0.23 standard deviations lower** (0.79 lower to 0.34 higher)
Digestive problems (measured with: Clinical Global Impression (CGI): Digestive problems (change score); Better indicated by lower values)										
50 (1 study) 1 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	24	26	N/A	N/A	The mean digestive problems in the intervention groups was **0.18 standard deviations lower** (0.74 lower to 0.37 higher)
Digestive problems (measured with: Clinical Global Impression (CGI): Digestive problems (change score); Better indicated by lower values)										
48 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	24	24	N/A	N/A	The mean digestive problems in the intervention groups was **0 standard deviations higher**

										(0.57 lower to 0.57 higher)
Overall autistic behaviours (porcine and synthetic secretin groups combined) (measured with: Parent-rated Secretin Outcome Survey-Modified (SOS-M): Total (change score); Better indicated by lower values)										
78 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	31	47	N/A	N/A	The mean overall autistic behaviours (porcine and synthetic secretin groups combined) in the intervention groups was **0.1 standard deviations lower** (0.56 lower to 0.35 higher)
Overall autistic behaviours (porcine and synthetic secretin groups combined) (measured with: Teacher-rated Secretin Outcome Survey-Modified (SOS-M): Total (change score); Better indicated by lower values)										
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	22	34	N/A	N/A	The mean overall autistic behaviours (porcine and synthetic secretin groups combined) in the intervention groups was **0.17 standard deviations higher** (0.37 lower to 0.71 higher)
Social (porcine and synthetic secretin groups combined) (measured with: Parent-rated Secretin Outcome Survey-Modified (SOS-M): Social (change score); Better indicated by lower values)										
78 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	31	47	N/A	N/A	The mean social (porcine and synthetic secretin groups combined) in the intervention groups was **0.07 standard deviations higher** (0.38 lower to 0.53 higher)
Social (porcine and synthetic secretin groups combined) (measured with: Teacher-rated Secretin Outcome Survey-Modified (SOS-M): Social (change score); Better indicated by lower values)										
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	22	34	N/A	N/A	The mean social (porcine and synthetic secretin groups combined) in the intervention groups was **0.25 standard deviations higher** (0.28 lower to 0.79 higher)

Communication (porcine and synthetic secretin groups combined) (measured with: Parent-rated Secretin Outcome Survey-Modified (SOS-M): Communication (change score); Better indicated by lower values)										
78 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	31	47	N/A	N/A	The mean communication (porcine and synthetic secretin groups combined) in the intervention groups was **0.25 standard deviations higher** (0.2 lower to 0.71 higher)
Communication (porcine and synthetic secretin groups combined) (measured with: Teacher-rated Secretin Outcome Survey-Modified (SOS-M): Communication (change score); Better indicated by lower values)										
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	22	34	N/A	N/A	The mean communication (porcine and synthetic secretin groups combined) in the intervention groups was **0.5 standard deviations higher** (0.05 lower to 1.04 higher)
Repetitive behaviour (porcine and synthetic secretin groups combined) (measured with: Parent-rated Secretin Outcome Survey-Modified (SOS-M): Repetitive behaviour (change score); Better indicated by lower values)										
78 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	31	47	N/A	N/A	The mean repetitive behaviour (porcine and synthetic secretin groups combined) in the intervention groups was **0.2 standard deviations lower** (0.65 lower to 0.25 higher)
Repetitive behaviour (porcine and synthetic secretin groups combined) (measured with: Teacher-rated Secretin Outcome Survey-Modified (SOS-M): Repetitive behaviour (change score); Better indicated by lower values)										
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	22	34	N/A	N/A	The mean repetitive behaviour (porcine and synthetic secretin groups combined) in the intervention groups was **0.18 standard deviations higher** (0.36 lower to 0.72 higher)
Digestive (porcine and synthetic secretin groups combined) (measured with: Parent-rated Secretin Outcome Survey-Modified (SOS-M): Digestive (change score); Better indicated by lower values)										
78 (1 study)	no serious risk of	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5]	31	47	N/A	N/A	The mean digestive (porcine and synthetic secretin groups

4 weeks	bias					due to imprecision				combined) in the intervention groups was **0.08 standard deviations higher** (0.37 lower to 0.54 higher)
Digestive (porcine and synthetic secretin groups combined) (measured with: Teacher-rated Secretin Outcome Survey-Modified (SOS-M): Digestive (change score); Better indicated by lower values)										
35 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	15	20	N/A	N/A	The mean digestive (porcine and synthetic secretin groups combined) in the intervention groups was **0.28 standard deviations higher** (0.39 lower to 0.96 higher)
Mood (porcine and synthetic secretin groups combined) (measured with: Parent-rated Secretin Outcome Survey-Modified (SOS-M): Mood (change score); Better indicated by lower values)										
77 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	31	46	N/A	N/A	The mean mood (porcine and synthetic secretin groups combined) in the intervention groups was **0.06 standard deviations lower** (0.51 lower to 0.4 higher)
Mood (porcine and synthetic secretin groups combined) (measured with: Teacher-rated Secretin Outcome Survey-Modified (SOS-M): Mood (change score); Better indicated by lower values)										
47 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	18	29	N/A	N/A	The mean mood (porcine and synthetic secretin groups combined) in the intervention groups was **0.33 standard deviations higher** (0.26 lower to 0.93 higher)
Sensory (porcine and synthetic secretin groups combined) (measured with: Parent-rated Secretin Outcome Survey-Modified (SOS-M): Sensory (change score); Better indicated by lower values)										
77 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	31	46	N/A	N/A	The mean sensory (porcine and synthetic secretin groups combined) in the intervention groups was **0.39 standard deviations lower**

										(0.85 lower to 0.07 higher)
Sensory (porcine and synthetic secretin groups combined) (measured with: Teacher-rated Secretin Outcome Survey-Modified (SOS-M): Sensory (change score); Better indicated by lower values)										
46 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	18	28	N/A	N/A	The mean sensory (porcine and synthetic secretin groups combined) in the intervention groups was **0 standard deviations higher** (0.59 lower to 0.59 higher)
Hyperactivity (porcine and synthetic secretin groups combined) (measured with: Parent-rated Secretin Outcome Survey-Modified (SOS-M): Hyperactivity (change score); Better indicated by lower values)										
77 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	31	46	N/A	N/A	The mean hyperactivity (porcine and synthetic secretin groups combined) in the intervention groups was **0.05 standard deviations lower** (0.51 lower to 0.4 higher)
Hyperactivity (porcine and synthetic secretin groups combined) (measured with: Teacher-rated Secretin Outcome Survey-Modified (SOS-M): Hyperactivity (change score); Better indicated by lower values)										
43 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	16	27	N/A	N/A	The mean hyperactivity (porcine and synthetic secretin groups combined) in the intervention groups was **0.14 standard deviations higher** (0.48 lower to 0.76 higher)
Lethargy (porcine and synthetic secretin groups combined) (measured with: Parent-rated Secretin Outcome Survey-Modified (SOS-M): Lethargy (change score); Better indicated by lower values)										
76 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	30	46	N/A	N/A	The mean lethargy (porcine and synthetic secretin groups combined) in the intervention groups was **0.09 standard deviations higher** (0.37 lower to 0.55 higher)
Lethargy (porcine and synthetic secretin groups combined) (measured with: Teacher-rated Secretin Outcome Survey-Modified (SOS-M): Lethargy (change score); Better indicated by lower values)										

41 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	15	26	N/A	N/A	The mean lethargy (porcine and synthetic secretin groups combined) in the intervention groups was **0.31 standard deviations higher** (0.33 lower to 0.95 higher)
Sleep (porcine and synthetic secretin groups combined) (measured with: Parent-rated Secretin Outcome Survey-Modified (SOS-M): Lethargy (change score); Better indicated by lower values)											
76 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	31	45	N/A	N/A	The mean sleep (porcine and synthetic secretin groups combined) in the intervention groups was **0.02 standard deviations higher** (0.44 lower to 0.48 higher)

[1] Risk of detection bias is unclear/unknown in CONIGLIO2001 as the paper reports that it was 'double-blind study' but it is not clear whether outcome assessors were blinded
[2] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)
[3] High risk of selective reporting bias in CONIGLIO2001 as data could not be extracted for the CARS (continuous measure), GARS or PLS
[4] N<400
[5] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[6] Moderate heterogeneity

1.7.3 Medical procedures for overall autistic behaviours as a direct or indirect outcome

Long-term chelation (7-rounds of DMSA therapy) versus short-term chelation (1-round of DMSA therapy and 6-rounds of placebo) for overall autistic behaviours as a direct outcome

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 Short-term chelation (1-round of DMSA	**With Long-term chelation (7-rounds of Dimercaptosuccinic Acid [DMSA] therapy)**		**Risk with Short-term chelation (1-round of DMSA**	**Risk difference with Long-term chelation (7-rounds of Dimercaptosuccinic Acid [DMSA] therapy)** (95% CI)

							therapy and 6-rounds of placebo)			**therapy and 6-rounds of placebo)**	
Overall autistic behaviours (measured with: Autism Evaluation Treatment Checklist (ATEC): Total; Better indicated by lower values)											
24 (1 study) 17 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	10	14	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.25 standard deviations higher** (0.57 lower to 1.06 higher)
Speech/Language/Communication (measured with: Autism Evaluation Treatment Checklist (ATEC): Speech/Language/Communication; Better indicated by lower values)											
40 (1 study) 17 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	15	25	N/A	N/A	The mean speech/language/communication in the intervention groups was **0.01 standard deviations higher** (0.63 lower to 0.65 higher)
Sociability (measured with: Autism Evaluation Treatment Checklist (ATEC): Sociability; Better indicated by lower values)											
40 (1 study) 17 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	15	25	N/A	N/A	The mean sociability in the intervention groups was **0.14 standard deviations higher** (0.51 lower to 0.78 higher)

Sensory/Cognitive Awareness (measured with: Autism Evaluation Treatment Checklist (ATEC): Sensory/Cognitive Awareness; Better indicated by lower values)											
40 (1 study) 17 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	15	25	N/A	N/A	The mean sensory/cognitive awareness in the intervention groups was **0.28 standard deviations higher** (0.36 lower to 0.93 higher)
Health/Physical/Behavior (measured with: Autism Evaluation Treatment Checklist (ATEC): Health/Physical/Behavior; Better indicated by lower values)											
24 (1 study) 17 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	10	14	N/A	N/A	The mean health/physical/behavior in the intervention groups was **0.33 standard deviations higher** (0.49 lower to 1.14 higher)
Overall autistic behaviours (measured with: Pervasive Development Disorder Behavior Inventory (PDDBI): Autism Composite; Better indicated by lower values)											
40 (1 study) 17 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	15	25	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.24 standard deviations higher** (0.41 lower to 0.88 higher)
Overall autistic behaviours (measured with: Severity of Autism Scale (SAS): Total; Better indicated by lower values)											
36 (1 study) 17 weeks	no serious risk of	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly suspected [2]	⊕⊖⊖⊖ **VERY LOW**[1,2] due to	14	22	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.13 standard deviations lower**

	bias					imprecision, publication bias				(0.8 lower to 0.54 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] High risk of selective reporting bias as efficacy data cannot be extracted for the Parent Global Impressions scale as no measure of variability reported

HBOT versus placebo for overall autistic behaviours as a direct or indirect outcome

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 Attention-placebo control	**With Hyperbaric oxygen treatment (HBOT)**		**Risk with Attention-placebo control**	**Risk difference with Hyperbaric oxygen treatment (HBOT)** (95% CI)
Positive treatment response (assessed with: Number of participants who showed improvement in ADOS Total score)											
34 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	4/16 (25%)	5/18 (27.8%)	**RR 1.11** (0.36 to 3.44)	**Study population**	
										250 per 1000	**28 more per 1000** (from 160 fewer to 610 more)
										Moderate	
										250 per 1000	**28 more per 1000** (from 160 fewer to 610 more)
Overall autistic behaviours (measured with: Autism Diagnostic Observation Schedule (ADOS): Total score; Better indicated by lower values)											
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	26	30	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.16 standard deviations lower** (0.69 lower to 0.37 higher)
Overall autistic behaviours (parent-rated) (measured with: Autism Evaluation Treatment Checklist (ATEC): Total; Better indicated by lower values)											
114	no	no serious	no serious	serious[3]	undetected	⊕⊕⊕⊖	55	59	N/A	N/A	The mean overall autistic behaviours

(2 studies) 4 weeks	serious risk of bias	inconsistency	indirectness			**MODERATE**[3] due to imprecision				(parent-rated) in the intervention groups was **0.05 standard deviations lower** (0.42 lower to 0.32 higher)
Speech/Language/Communication (parent-rated) (measured with: Autism Evaluation Treatment Checklist (ATEC): Speech/Language/Communication; Better indicated by lower values)										
114 (2 studies) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊕⊖ **MODERATE**[3] due to imprecision	55	59	N/A	N/A	The mean speech/language/communication (parent-rated) in the intervention groups was **0.10 standard deviations higher** (0.27 lower to 0.47 higher)
Sociability (parent-rated) (measured with: Autism Evaluation Treatment Checklist (ATEC): Sociability; Better indicated by lower values)										
114 (2 studies) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊕⊖ **MODERATE**[3] due to imprecision	55	59	N/A	N/A	The mean sociability (parent-rated) in the intervention groups was **0.02 standard deviations lower** (0.39 lower to 0.35 higher)
Sensory/Cognitive Awareness (parent-rated) (measured with: Autism Evaluation Treatment Checklist (ATEC): Sensory/Cognitive Awareness; Better indicated by lower values)										
114 (2 studies) 4 weeks	no serious risk of bias	very serious[4]	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,4] due to inconsistency, imprecision	55	59	N/A	N/A	The mean sensory/cognitive awareness (parent-rated) in the intervention groups was **0.25 standard deviations lower** (0.62 lower to 0.13 higher)
Health/Physical/Behavior (parent-rated) (measured with: Autism Evaluation Treatment Checklist (ATEC): Health/Physical/Behavior; Better indicated by lower values)										
114 (2 studies) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊕⊖ **MODERATE**[3] due to imprecision	55	59	N/A	N/A	The mean health/physical/behavior (parent-rated) in the intervention groups was **0.02 standard deviations higher** (0.35 lower to 0.39 higher)
Overall autistic behaviours (clinician-rated) (measured with: Autism Evaluation Treatment Checklist (ATEC): Total; Better indicated by lower values)										
58	no	no serious	no serious	very	undetected	⊕⊕⊖⊖	29	29	N/A	N/A	The mean overall autistic behaviours

(1 study) 4 weeks	serious risk of bias	inconsistency	indirectness	serious[2]		**LOW**[2] due to imprecision					(clinician-rated) in the intervention groups was **0.03 standard deviations lower** (0.54 lower to 0.49 higher)
Speech/Language/Communication (clinician-rated) (measured with: Autism Evaluation Treatment Checklist (ATEC): Speech/Language/Communication; Better indicated by lower values)											
58 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	29	29	N/A	N/A	The mean speech/language/communication (clinician-rated) in the intervention groups was **0.04 standard deviations lower** (0.55 lower to 0.48 higher)
Sociability (clinician-rated) (measured with: Autism Evaluation Treatment Checklist (ATEC): Sociability; Better indicated by lower values)											
58 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	29	29	N/A	N/A	The mean sociability (clinician-rated) in the intervention groups was **0.27 standard deviations higher** (0.25 lower to 0.79 higher)
Sensory/Cognitive Awareness (clinician-rated) (measured with: Autism Evaluation Treatment Checklist (ATEC): Sensory/Cognitive Awareness; Better indicated by lower values)											
58 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	29	29	N/A	N/A	The mean sensory/cognitive awareness (clinician-rated) in the intervention groups was **0.07 standard deviations lower** (0.59 lower to 0.44 higher)
Health/Physical/Behaviour (clinician-rated) (measured with: Autism Evaluation Treatment Checklist (ATEC): Health/Physical/Behavior; Better indicated by lower values)											
58 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	29	29	N/A	N/A	The mean health/physical/behaviour (clinician-rated) in the intervention groups was **0.2 standard deviations lower** (0.72 lower to 0.31 higher)
Global severity (parent-rated) (measured with: Clinical Global Impression Scale (CGI-S): Severity; Better indicated by lower values)											
58	no	no serious	no serious	very	undetected	⊕⊕⊖⊖	29	29	N/A	N/A	The mean global severity (parent-

(1 study) 4 weeks	serious risk of bias	inconsistency	indirectness	serious[2]		**LOW**[2] due to imprecision					rated) in the intervention groups was **0.03 standard deviations higher** (0.48 lower to 0.55 higher)
Global severity (clinician-rated) (measured with: Clinical Global Impression Scale (CGI-S): Severity; Better indicated by lower values)											
58 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	29	29	N/A	N/A	The mean global severity (clinician-rated) in the intervention groups was **0.34 standard deviations lower** (0.86 lower to 0.18 higher)
Global improvement (parent-rated) (measured with: Clinical Global Impression Scale (CGI-I): Improvement; Better indicated by lower values)											
58 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	29	29	N/A	N/A	The mean global improvement (parent-rated) in the intervention groups was **0.28 standard deviations lower** (0.8 lower to 0.23 higher)
Global improvement (cinician-rated) (measured with: Clinical Global Impression Scale (CGI-I): Improvement; Better indicated by lower values)											
58 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊕⊖ **MODERATE**[3] due to imprecision	29	29	N/A	N/A	The mean global improvement (cinician-rated) in the intervention groups was **0.57 standard deviations lower** (1.1 to 0.05 lower)

[1] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] N<400
[4] I-squared value indicates substantial to considerable heterogeneity

1.7.4 Nutritional interventions for overall autistic behaviours as a direct or indirect outcome

Multivitamin/mineral supplement versus placebo for overall autistic behaviours as a direct outcome

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 Placebo	With Multivitamin and mineral supplement		Risk with Placebo	Risk difference with Multivitamin and mineral supplement (95% CI)
Average improvement (measured with: Parent Global Impressions-Revised (PGI-R): Average improvement (average of all subscales); Better indicated by lower values)											
104 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	51	53	N/A	N/A	The mean average improvement in the intervention groups was **0.55 standard deviations higher** (0.16 to 0.94 higher)
Overall improvement (measured with: Parent Global Impressions-Revised (PGI-R): Overall improvement; Better indicated by lower values)											
104 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	51	53	N/A	N/A	The mean overall improvement in the intervention groups was **0.49 standard deviations higher** (0.1 to 0.88 higher)
Overall autistic behaviours (measured with: Autism Evaluation Treatment Checklist (ATEC): Total; Better indicated by lower values)											
104 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	51	53	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.04 standard deviations higher** (0.34 lower to 0.43 higher)
Overall autistic behaviours (measured with: Severity of Autism Scale (SAS): Total; Better indicated by lower values)											
104 (1 study)	no serious	no serious	no serious	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1]	51	53	N/A	N/A	The mean overall autistic behaviours in the

13 weeks	risk of bias	inconsistency	indirectness			due to imprecision				intervention groups was **0.04 standard deviations lower** (0.43 lower to 0.34 higher)
Overall autistic behaviours (measured with: Pervasive Development Disorder Behavior Inventory (PDDBI): Total; Better indicated by lower values)										
104 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	51	53	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.02 standard deviations higher** (0.37 lower to 0.4 higher)
[1] N<400										

L-carnosine or L-carnitine supplement versus placebo for overall autistic behaviours as a direct outcome

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 Placebo	**With L-carnosine/L-carnitine supplement**		**Risk with Placebo**	**Risk difference with L-carnosine/L-carnitine supplement** (95% CI)
Global improvement (measured with: Parent Global Impressions-Improvement (PGI-I): Overall improvement across subscales; Better indicated by lower values)											
31 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	14	N/A	N/A	The mean global improvement in the intervention groups was **0.47 standard deviations higher** (0.25 lower to 1.19 higher)
Overall autistic behaviours (measured with: Childhood Autism Rating Scale (CARS): Total; Better indicated by lower values)											
56 (2 studies)	no serious risk of bias	very serious[2]	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2]	28	28	N/A	N/A	The mean overall autistic behaviours in the

8-26 weeks						due to inconsistency, imprecision					intervention groups was **0.12 standard deviations lower** (0.65 lower to 0.42 higher)
Overall autistic behaviours (measured with: Gilliam Autism Rating Scale (GARS): Total; Better indicated by lower values)											
31 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	14	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.34 standard deviations lower** (1.05 lower to 0.38 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] I-squared value indicates substantial heterogeneity

Omega-3 fatty acids versus placebo for overall autistic behaviours as an indirect outcome

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 Healthy diet control	**With Omega-3 fatty acids**		**Risk with Healthy diet control**	**Risk difference with Omega-3 fatty acids** (95% CI)
Pervasive Developmental Disorder (PDD) symptoms (measured with: Child Behavior Checklist 1.5 - 5 (CBCL/1.5-5): PDD; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A	The mean pervasive developmental disorder (pdd) symptoms in the intervention groups was **0.98 standard deviations lower**

										(1.86 to 0.1 lower)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as the outcome assessor for this outcome measure was not blinded.
[2] N<400

Gluten-free and casein-free diet versus treatment-as-usual for overall autistic behaviours as a direct outcome

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 Treatment-as-usual	**With Gluten-free and casein-free diet**		**Risk with Treatment-as-usual**	**Risk difference with Gluten-free and casein-free diet** (95% CI)
Overall autistic behaviours (measured with: Diagnose of Psykotisk Adferd hos Børn (Diagnosis of Psychotic Behaviour in Children; DIPAB): Total; Better indicated by lower values)											
20 (1 study) 52 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	10	10	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **1.37 standard deviations lower** (2.36 to 0.37 lower)

[1] High risk of performance and response bias as intervention administrators (parents) and participants were non-blind. There was also a high risk of detection bias for the DIPAB as although the investigator was blinded to group assignment, this outcome measure was based on parental interview and parents were non-blind to group assignment and other potentially confounding factors
[2] N<400

1.7.5 Sensory interventions for overall autistic behaviours as a direct or indirect outcome

Neurofeedback versus treatment-as-usual for overall autistic behaviours as a direct outcome

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 Treatment-as-usual	With Neurofeedback		Risk with Treatment-as-usual	Risk difference with Neurofeedback (95% CI)
Parent-rated overall autistic behaviours (measured with: Social Communication Questionnaire (SCQ): Total; Better indicated by lower values)											
20 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	10	10	N/A	N/A	The mean parent-rated overall autistic behaviours in the intervention groups was **1.85 standard deviations lower** (2.94 to 0.77 lower)
Teacher-rated overall autistic behaviours (measured with: Social Communication Questionnaire (SCQ): Total; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated overall autistic behaviours in the intervention groups was **0.29 standard deviations lower** (1.18 lower to 0.59 higher)

[1] High risk of performance, response and detection bias as intervention administrators, participants and outcome assessors were non-blind. The risk of other bias due to potential conflict of interest is also high as neurofeedback equipment provided by manufacturer for trial.
[2] N<400
[3] High risk of selective reporting bias as data cannot be extracted for 6-month follow-up
[4] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Auditory integration training versus attention-placebo (structured listening) for overall autistic behaviours as an indirect outcome

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 Attention-placebo (structured listening) control	**With Auditory integration training**		**Risk with Attention-placebo (structured listening) control**	**Risk difference with Auditory integration training** (95% CI)
Overall autistic behaviours (measured with: Autism Behaviour Checklist (ABC): Total; Better indicated by lower values)											
80 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.1 standard deviations higher** (0.34 lower to 0.54 higher)
Overall autistic behaviours (measured with: Autism Behaviour Checklist (ABC): Total; Better indicated by lower values)											
80 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.22 standard deviations higher** (0.22 lower to 0.66 higher)
Overall autistic behaviours (measured with: Autism Behaviour Checklist (ABC): Total; Better indicated by lower values)											

80 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.25 standard deviations higher** (0.19 lower to 0.69 higher)
Overall autistic behaviours (measured with: Autism Behaviour Checklist (ABC): Total; Better indicated by lower values)											
80 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean overall autistic behaviours in the intervention groups was **0.27 standard deviations higher** (0.17 lower to 0.71 higher)
[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)											

1.8 BIOMEDICAL INTERVENTIONS AIMED AT THE CORE AUTISM FEATURE OF IMPAIRED RECIPROCAL SOCIAL COMMUNICATION AND INTERACTION

1.8.1 Complementary therapies for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

Electro-acupuncture and conventional educational programme versus conventional educational programme only for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

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 Acupuncture/electro-acupuncture and conventional educational programme versus conventional educational programme only for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome**		**Risk with Control**	**Risk difference with Acupuncture/electro-acupuncture and conventional educational programme versus conventional educational programme only for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome** (95% CI)
Communication (measured with: Autism Diagnostic Observation Schedule (ADOS/ADOS-G): Communication (change score); Better indicated by lower values)											
36 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	18	18	N/A	N/A	The mean communication in the intervention groups was **0.19 standard deviations lower** (0.85 lower to 0.46 higher)
Social interaction (measured with: Autism Diagnostic Observation Schedule (ADOS/ADOS-G): Social Interaction (change score); Better indicated by lower values)											
36 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	18	18	N/A	N/A	The mean social interaction in the intervention groups was **0 standard deviations higher** (0.65 lower to 0.65 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.8.2 Hormones for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

Secretin versus placebo for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

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 Secretin versus placebo for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome		Risk with Control	Risk difference with Secretin versus placebo for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome (95% CI)
Communication (measured with: Autism Diagnostic Observation Schedule (ADOS/ADOS-G): Communication (endpoint and change scores); Better indicated by lower values)											
141 (2 studies) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	61	80	N/A	N/A	The mean communication in the intervention groups was **0.1 standard deviations lower** (0.44 lower to 0.24 higher)
Communication (measured with: Gilliam Autism Rating Scale (GARS): Communication; Better indicated by lower values)											
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	28	28	N/A	N/A	The mean communication in the intervention groups was **0.38 standard deviations higher** (0.15 lower to 0.9 higher)
Social interaction (measured with: Autism Diagnostic Observation Schedule (ADOS/ADOS-G): Social interaction (endpoint and change scores); Better indicated by lower values)											

141 (2 studies) 4 weeks	no serious risk of bias	very serious[3]	no serious indirectness	serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to inconsistency, imprecision	61	80	N/A	N/A	The mean social interaction in the intervention groups was **0.46 standard deviations higher** (0.12 to 0.8 higher)
Social interaction (measured with: Gilliam Autism Rating Scale (GARS): Social Interaction; Better indicated by lower values)										
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	28	28	N/A	N/A	The mean social interaction in the intervention groups was **0.42 standard deviations higher** (0.11 lower to 0.95 higher)
Communication and Social interaction (measured with: Autism Diagnostic Observation Schedule (ADOS/ADOS-G): Communication & Social Interaction (change score); Better indicated by lower values)										
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	28	28	N/A	N/A	The mean communication and social interaction in the intervention groups was **0.55 standard deviations higher** (0.02 to 1.09 higher)

[1] N<400
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] Moderate to substantial heterogeneity

1.8.3 Medical procedures for the core autism feature of impaired reciprocal social communication and interaction as a direct or indirect outcome

Hyperbaric oxygen treatment (HBOT) versus attention-placebo for the core autism feature of impaired reciprocal social communication and interaction as a direct outcome

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 Attention-placebo control	With Hyperbaric oxygen treatment (HBOT)		Risk with Attention-placebo control	Risk difference with Hyperbaric oxygen treatment (HBOT) (95% CI)
Positive treatment response (assessed with: Number of participants who showed improvement in ADOS Communication)											
34 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	2/16 (12.5%)	3/18 (16.7%)	**RR 1.33** (0.25 to 7)	**Study population**	
										125 per 1000	**41 more per 1000** (from 94 fewer to 750 more)
										Moderate	
										125 per 1000	**41 more per 1000** (from 94 fewer to 750 more)
Positive treatment response (assessed with: Number of participants who showed improvement in ADOS Socialization)											
34 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	2/16 (12.5%)	3/18 (16.7%)	**OR 1.4** (0.2 to 9.66)	**Study population**	
										125 per 1000	**42 more per 1000** (from 97 fewer to 455 more)
										Moderate	
										125 per 1000	**42 more per 1000** (from 97 fewer to 455 more)
Social Awareness (measured with: Social Responsiveness Scale (SRS): Social Awareness (change score); Better indicated by lower values)											
29 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	15	14	N/A	N/A	The mean social awareness in the intervention groups was **0.11 standard**

											deviations lower (0.84 lower to 0.62 higher)
Social Cognition (measured with: Social Responsiveness Scale (SRS): Social Cognition (change score); Better indicated by lower values)											
29 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	15	14	N/A	N/A	The mean social cognition in the intervention groups was **0.53 standard deviations higher** (0.21 lower to 1.27 higher)
Social Communication (measured with: Social Responsiveness Scale (SRS): Social Communication (change score); Better indicated by lower values)											
29 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	15	14	N/A	N/A	The mean social communication in the intervention groups was **0.32 standard deviations lower** (1.05 lower to 0.41 higher)
Social Motivation (measured with: Social Responsiveness Scale (SRS): Social Motivation (change score); Better indicated by lower values)											
29 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	15	14	N/A	N/A	The mean social motivation in the intervention groups was **0.06 standard deviations higher** (0.67 lower to 0.79 higher)
Autistic Mannerisms (measured with: Social Responsiveness Scale (SRS): Autistic Mannerisms (change score); Better indicated by lower values)											
29 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	15	14	N/A	N/A	The mean autistic mannerisms in the intervention groups was **0.36 standard deviations higher** (0.38 lower to 1.09

										higher)
Appropriate vocalization (measured with: Behavioural observation: Appropriate vocalization (change score); Better indicated by lower values)										
34 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	16	18	N/A	N/A	The mean appropriate vocalization in the intervention groups was **0.17 standard deviations higher** (0.51 lower to 0.84 higher)

[1] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Long-term chelation (7-rounds of DMSA therapy) versus short-term chelation (1-round of DMSA therapy and 6-rounds of placebo) for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

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 Short-term chelation (1-round of DMSA therapy and 6-rounds of placebo)	**With Long-term chelation (7-rounds of Dimercaptosuccinic Acid [DMSA] therapy)**		**Risk with Short-term chelation (1-round of DMSA therapy and 6-rounds of placebo)**	**Risk difference with Long-term chelation (7-rounds of Dimercaptosuccinic Acid [DMSA] therapy)** (95% CI)
Social Pragmatic Problems (measured with: Pervasive Development Disorder Behavior Inventory (PDDBI): Social Pragmatic; Better indicated by lower values)											
40 (1 study) 17 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	15	25	N/A	N/A	The mean social pragmatic problems in the intervention groups was **0.52 standard deviations higher** (0.13 lower to 1.17

										higher)
Social Approach Behaviours (measured with: Pervasive Development Disorder Behavior Inventory (PDDBI): Social Approach; Better indicated by lower values)										
40 (1 study) 17 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	15	25	N/A	N/A	The mean social approach behaviours in the intervention groups was **0.08 standard deviations lower** (0.72 lower to 0.56 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] High risk of selective reporting bias as efficacy data cannot be extracted for the ADOS Communication, Sociability, and Communication+Sociability or the Parent Global Impressions scale as no measure of variability reported

1.8.4 Nutritional interventions for the core autism feature of impaired reciprocal social communication and interaction as a direct or indirect outcome

Gluten-free and casein-free diet versus treatment-as-usual for the core autism feature of impaired reciprocal social communication and interaction as a direct or indirect outcome

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 Treatment-as-usual	**With Gluten-free and casein-free diet**		**Risk with Treatment-as-usual**	**Risk difference with Gluten-free and casein-free diet** (95% CI)
Communication (measured with: Autism Diagnostic Observation Schedule (ADOS): Communication (change score); Better indicated by lower values)											
55 (1 study) 35 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	29	26	N/A	N/A	The mean communication in the intervention groups was **0.42 standard deviations lower** (0.95 lower to 0.12 higher)

Communication (measured with: Gilliam Autism Rating Scale (GARS): Communication (change score); Better indicated by lower values)											
55 (1 study) 35 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	29	26	N/A	N/A	The mean communication in the intervention groups was **0.34 standard deviations lower** (0.87 lower to 0.19 higher)
Social Interaction (measured with: Autism Diagnostic Observation Schedule (ADOS): Social Interaction (change score); Better indicated by lower values)											
55 (1 study) 35 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	29	26	N/A	N/A	The mean social interaction in the intervention groups was **0.01 standard deviations lower** (0.54 lower to 0.52 higher)
Social Interaction (measured with: Gilliam Autism Rating Scale (GARS): Social Interaction (change score); Better indicated by lower values)											
55 (1 study) 35 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[3,4] due to risk of bias, imprecision	29	26	N/A	N/A	The mean social interaction in the intervention groups was **0.67 standard deviations lower** (1.22 to 0.13 lower)
Communication and interaction (measured with: Diagnose of Psykotisk Adferd hos Børn (Diagnosis of Psychotic Behaviour in Children; DIPAB): Communication and interaction (K-scores); Better indicated by lower values)											
20 (1 study) 52 weeks	serious[5]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4,5] due to risk of bias, imprecision	10	10	N/A	N/A	The mean communication and interaction in the intervention groups was **1.19 standard deviations higher** (0.22 to 2.15 higher)
Resistance to communication and interaction (measured with: Diagnose of Psykotisk Adferd hos Børn (Diagnosis of Psychotic Behaviour in Children; DIPAB): Resistance to communication and interaction (M-scores); Better indicated by lower values)											
20 (1 study) 52 weeks	serious[5]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4,5] due to risk of	10	10	N/A	N/A	The mean resistance to communication and interaction in the intervention

						bias, imprecision					groups was **1.58 standard deviations lower** (2.61 to 0.55 lower)
Social isolation (measured with: Diagnose of Psykotisk Adferd hos Børn (Diagnosis of Psychotic Behaviour in Children; DIPAB): Social interaction or isolation (I-scores); Better indicated by lower values)											
20 (1 study) 52 weeks	serious[5]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4,5] due to risk of bias, imprecision	10	10	N/A	N/A	The mean social isolation in the intervention groups was **1.35 standard deviations lower** (2.34 to 0.35 lower)

[1] High risk of attrition bias as over twice as many dropouts in the experimental group relative to the controls (32% in experimental group and 15% in the control group)
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] High risk of performance and response bias as intervention administrators (parents) and participants were non-blind, and unclear/unknown risk of detection bias as the identity and blinding of outcome assessors not reported. Also high risk of attrition bias as over twice as many dropouts in the experimental group relative to the controls (32% in experimental group and 15% in the control group)
[4] N<400
[5] High risk of performance and response bias as intervention administrators (parents) and participants were non-blind. There was also a high risk of detection bias for the DIPAB as although the investigator was blinded to group assignment, this outcome measure was based on parental interview and parents were non-blind to group assignment and other potentially confounding factors

Omega-3 fatty acids versus placebo for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

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 Placebo	**With Omega-3 fatty acids**		**Risk with Placebo**	**Risk difference with Omega-3 fatty acids** (95% CI)
Social skills (measured with: Social Responsiveness Scale (SRS): Total; Better indicated by lower values)											
22 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	11	11	N/A	N/A	The mean social skills in the intervention groups was **0.06 standard deviations higher**

										(0.77 lower to 0.9 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Omega-3 fatty acids versus healthy diet control for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

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 Healthy diet control	**With Omega-3 fatty acids**		**Risk with Healthy diet control**	**Risk difference with Omega-3 fatty acids** (95% CI)
Frequency of positive vocalizations (measured with: Behavioural observation; Better indicated by lower values)											
23 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	13	10	N/A	N/A	The mean frequency of positive vocalizations in the intervention groups was **0.21 standard deviations higher** (0.62 lower to 1.03 higher)
Frequency of social initiations (measured with: Behavioural observation; Better indicated by lower values)											
23 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	13	10	N/A	N/A	The mean frequency of social initiations in the intervention groups was **0.44 standard deviations higher** (0.4 lower to 1.27 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm

Multivitamin/mineral supplement versus placebo for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

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 Placebo	With Multivitamin and mineral supplement		Risk with Placebo	Risk difference with Multivitamin and mineral supplement (95% CI)
Sociability improvement (measured with: Parent Global Impressions-Revised (PGI-R): Sociability improvement; Better indicated by lower values)											
104 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	51	53	N/A	N/A	The mean sociability improvement in the intervention groups was **0.14 standard deviations higher** (0.24 lower to 0.53 higher)
Eye contact improvement (measured with: Parent Global Impressions-Revised (PGI-R): Eye contact improvement; Better indicated by lower values)											
104 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	51	53	N/A	N/A	The mean eye contact improvement in the intervention groups was **0.28 standard deviations higher** (0.11 lower to 0.67 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

L-carnosine supplement versus placebo for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

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 Placebo	With L-carnosine supplement		Risk with Placebo	Risk difference with L-carnosine supplement (95% CI)
Communication (measured with: Gilliam Autism Rating Scale (GARS): Communication; Better indicated by lower values)											
31 (1 study)	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1]	17	14	N/A	N/A	The mean communication in the intervention groups

8 weeks						due to imprecision				was **0.19 standard deviations higher** (0.52 lower to 0.9 higher)
Social interaction (measured with: Gilliam Autism Rating Scale (GARS): Social Interaction; Better indicated by lower values)										
31 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	14	N/A	N/A	The mean social interaction in the intervention groups was **0.51 standard deviations lower** (1.23 lower to 0.21 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.8.5 Sensory interventions for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

Neurofeedback versus treatment-as-usual for the core autism feature of impaired reciprocal social communication and interaction as an indirect outcome

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 Treatment-as-usual	**With Neurofeedback**		**Risk with Treatment-as-usual**	**Risk difference with Neurofeedback** (95% CI)
Parent-rated reciprocal social interaction (measured with: Social Communication Questionnaire (SCQ): Reciprocal social interactions; Better indicated by lower values)											
20 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	10	10	N/A	N/A	The mean parent-rated reciprocal social interaction in the intervention groups was **1.54 standard deviations lower** (2.57 to 0.52 lower)

Teacher-rated reciprocal social interaction (measured with: Social Communication Questionnaire (SCQ): Reciprocal social interactions; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated reciprocal social interaction in the intervention groups was **0.39 standard deviations lower** (1.28 lower to 0.49 higher)
Parent-rated communication (measured with: Social Communication Questionnaire (SCQ): Communication; Better indicated by lower values)											
20 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	10	10	N/A	N/A	The mean parent-rated communication in the intervention groups was **1.14 standard deviations lower** (2.1 to 0.18 lower)
Teacher-rated communication (measured with: Social Communication Questionnaire (SCQ): Communication; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated communication in the intervention groups was **0.19 standard deviations lower** (1.07 lower to 0.69 higher)
Parent-rated communication (measured with: Children's Communication Checklist (CCC-2): Total; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean parent-rated communication in the intervention groups was **0.88 standard deviations lower** (1.81 lower to 0.04 higher)

Teacher-rated communication (measured with: Children's Communication Checklist (CCC-2): Total; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated communication in the intervention groups was **0.05 standard deviations lower** (0.93 lower to 0.83 higher)
Parent-rated social impairment (measured with: Social Responsiveness Scale (SRS): Total; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean parent-rated social impairment in the intervention groups was **0.92 standard deviations lower** (1.85 lower to 0.02 higher)
Teacher-rated social impairment (measured with: Social Responsiveness Scale (SRS): Total; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated social impairment in the intervention groups was **0.01 standard deviations higher** (0.87 lower to 0.88 higher)
Parent-rated social awareness (measured with: Social Responsiveness Scale (SRS): Social Awareness ; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean parent-rated social awareness in the intervention groups was **0.64 standard deviations lower** (1.55 lower to 0.26 higher)
Teacher-rated social awareness (measured with: Social Responsiveness Scale (SRS): Social Awareness ; Better indicated by lower values)											

20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated social awareness in the intervention groups was **0.22 standard deviations higher** (0.66 lower to 1.1 higher)
Parent-rated social cognition (measured with: Social Responsiveness Scale (SRS): Social Cognition ; Better indicated by lower values)											
20 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	10	10	N/A	N/A	The mean parent-rated social cognition in the intervention groups was **1.38 standard deviations lower** (2.38 to 0.38 lower)
Teacher-rated social cognition (measured with: Social Responsiveness Scale (SRS): Social Cognition ; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated social cognition in the intervention groups was **0.35 standard deviations higher** (0.53 lower to 1.24 higher)
Parent-rated social communication (measured with: Social Responsiveness Scale (SRS): Social Communication ; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean parent-rated social communication in the intervention groups was **0.78 standard deviations lower** (1.7 lower to 0.14 higher)
Teacher-rated social communication (measured with: Social Responsiveness Scale (SRS): Social Communication ; Better indicated by lower values)											
20 (1 study) 20 weeks	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,	10	10	N/A	N/A	The mean teacher-rated social communication in the intervention groups was

						publication bias					**0.49 standard deviations higher** (0.4 lower to 1.38 higher)
Parent-rated social motivation (measured with: Social Responsiveness Scale (SRS): Social Motivation; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean parent-rated social motivation in the intervention groups was **0.54 standard deviations lower** (1.43 lower to 0.36 higher)
Teacher-rated social motivation (measured with: Social Responsiveness Scale (SRS): Social Motivation; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated social motivation in the intervention groups was **0.45 standard deviations higher** (0.44 lower to 1.34 higher)
Parent-rated autistic mannerisms (measured with: Social Responsiveness Scale (SRS): Autistic Mannerisms ; Better indicated by lower values)											
20 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	10	10	N/A	N/A	The mean parent-rated autistic mannerisms in the intervention groups was **0.98 standard deviations lower** (1.92 to 0.04 lower)
Teacher-rated atistic mannerisms (measured with: Social Responsiveness Scale (SRS): Autistic Mannerisms ; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated atistic mannerisms in the intervention groups was **0.41 standard deviations lower** (1.3 lower to 0.48 higher)

Parent-rated social relations (measured with: Children's Communication Checklist (CCC-2): Social relations; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean parent-rated social relations in the intervention groups was **0.37 standard deviations lower** (1.26 lower to 0.51 higher)
Teacher-rated social relations (measured with: Children's Communication Checklist (CCC-2): Social relations; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated social relations in the intervention groups was **0 standard deviations higher** (0.88 lower to 0.88 higher)
Parent-rated interests (measured with: Children's Communication Checklist (CCC-2): Interests; Better indicated by lower values)											
20 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	10	10	N/A	N/A	The mean parent-rated interests in the intervention groups was **1.18 standard deviations lower** (2.15 to 0.21 lower)
Teacher-rated interests (measured with: Children's Communication Checklist (CCC-2): Interests; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated interests in the intervention groups was **0 standard deviations higher** (0.88 lower to 0.88 higher)
Parent-rated inappropriate initialization (measured with: Children's Communication Checklist (CCC-2): Inappropriate initialization; Better indicated by lower values)											
20	serious[1]	no serious	no serious	serious[2]	reporting bias	⊕⊖⊖⊖	10	10	N/A	N/A	The mean parent-rated

(1 study) 20 weeks		inconsistency	indirectness		strongly suspected [3]	**VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias					inappropriate initialization in the intervention groups was **1.08 standard deviations lower** (2.03 to 0.13 lower)
Teacher-rated inappropriate initialization (measured with: Children's Communication Checklist (CCC-2): Inappropriate initialization; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated inappropriate initialization in the intervention groups was **0.15 standard deviations lower** (1.03 lower to 0.73 higher)
Parent-rated stereotyped conversation (measured with: Children's Communication Checklist (CCC-2): Stereotyped conversation; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean parent-rated stereotyped conversation in the intervention groups was **0.56 standard deviations lower** (1.45 lower to 0.34 higher)
Teacher-rated stereotyped conversation (measured with: Children's Communication Checklist (CCC-2): Stereotyped conversation; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated stereotyped conversation in the intervention groups was **0.31 standard deviations higher** (0.58 lower to 1.19 higher)
Parent-rated context use (measured with: Children's Communication Checklist (CCC-2): Context use; Better indicated by lower values)											

20 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	10	10	N/A	N/A	The mean parent-rated context use in the intervention groups was **1 standard deviations lower** (1.94 to 0.06 lower)
Teacer-rated context use (measured with: Children's Communication Checklist (CCC-2): Context use; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacer-rated context use in the intervention groups was **0.29 standard deviations higher** (0.6 lower to 1.17 higher)
Parent-rated non-verbal communication (measured with: Children's Communication Checklist (CCC-2): Non-verbal communication; Better indicated by lower values)											
20 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	10	10	N/A	N/A	The mean parent-rated non-verbal communication in the intervention groups was **1.05 standard deviations lower** (2 to 0.1 lower)
Teacher-rated non-verbal communication (measured with: Children's Communication Checklist (CCC-2): Non-verbal communication; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated non-verbal communication in the intervention groups was **0.33 standard deviations higher** (0.55 lower to 1.22 higher)
Parent-rated pragmatics (measured with: Children's Communication Checklist (CCC-2): Pragmatics; Better indicated by lower values)											
20 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of	10	10	N/A	N/A	The mean parent-rated pragmatics in the intervention groups was

						bias, imprecision, publication bias				**0.98 standard deviations lower** (1.92 to 0.04 lower)
Teacher-rated pragmatics (measured with: Children's Communication Checklist (CCC-2): Pragmatics; Better indicated by lower values)										
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated pragmatics in the intervention groups was **0.24 standard deviations higher** (0.64 lower to 1.13 higher)

[1] High risk of performance, response and detection bias as intervention administrators, participants and outcome assessors were non-blind. The risk of other bias due to potential conflict of interest is also high as neurofeedback equipment provided by manufacturer for trial.
[2] N<400
[3] High risk of selective reporting bias as data cannot be extracted for 6-month follow-up
[4] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.9 BIOMEDICAL INTERVENTIONS AIMED AT THE CORE AUTISM FEATURE OF RESTRICTED INTERESTS AND RIGID AND REPETITIVE BEHAVIOURS

1.9.1 Hormones for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

Secretin versus placebo for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

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 Secretin versus placebo for the core		Risk with	Risk difference with Secretin versus placebo for the core autism feature of

								autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome		Control	restricted interests and rigid and repetitive behaviours as an indirect outcome (95% CI)
Stereotyped behaviour/interests (measured with: Autism Diagnostic Observation Schedule (ADOS/ADOS-G): Stereotyped behaviour/interests; Better indicated by lower values)											
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	28	28	N/A	N/A	The mean stereotyped behaviour/interests in the intervention groups was **0.36 standard deviations higher** (0.17 lower to 0.89 higher)
Stereotyped behaviours (measured with: Gilliam Autism Rating Scale (GARS): Stereotyped behaviours; Better indicated by lower values)											
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	28	28	N/A	N/A	The mean stereotyped behaviours in the intervention groups was **0.17 standard deviations higher** (0.36 lower to 0.69 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.9.2 Medical procedures for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

Long-term chelation (7-rounds of DMSA therapy) versus short-term chelation (1-round of DMSA therapy and 6-rounds of placebo) for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

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 Short-term chelation (1-round of	With Long-term chelation (7-rounds of Dimercaptosuccinic		Risk with Short-term chelation (1-	Risk difference with Long-term chelation (7-rounds of Dimercaptosuccinic Acid

							DMSA therapy and 6-rounds of placebo)	**Acid [DMSA] therapy)**		**round of DMSA therapy and 6-rounds of placebo)**	**[DMSA] therapy)** (95% CI)
Sensory/Perceptual Approach Behaviours (measured with: Pervasive Development Disorder Behavior Inventory (PDDBI): Sensory/Perceptual Approach Behaviours; Better indicated by lower values)											
40 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	25	N/A	N/A	The mean sensory/perceptual approach behaviours in the intervention groups was **0.29 standard deviations higher** (0.35 lower to 0.94 higher)
Ritualisms/Resistance to Change (measured with: Pervasive Development Disorder Behavior Inventory (PDDBI): Ritualisms/Resistance to Change; Better indicated by lower values)											
40 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	25	N/A	N/A	The mean ritualisms/resistance to change in the intervention groups was **0.18 standard deviations lower** (0.83 lower to 0.46 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

HBOT versus attention-placebo for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

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 Attention-placebo control	**With Hyperbaric oxygen treatment (HBOT)**		**Risk with Attention-placebo control**	**Risk difference with Hyperbaric oxygen treatment (HBOT)** (95% CI)

Vocal stereotypy (measured with: Behavioural observation: Vocal stereotypy (change score); Better indicated by lower values)										
34 (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	16	18	N/A	N/A	The mean vocal stereotypy in the intervention groups was **0.29 standard deviations lower** (0.97 lower to 0.39 higher)
Physical stereotypy (measured with: Behavioural observation: Physical stereotypy (change score); Better indicated by lower values)										
34 (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	16	18	N/A	N/A	The mean physical stereotypy in the intervention groups was **0.42 standard deviations lower** (1.1 lower to 0.26 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] High risk of selective reporting bias as data cannot be extracted for the Repetitive Behavior Scale (RBS)

1.9.3 Motor intervention for the core autism feature of restricted interests and rigid and repetitive behaviours as a direct outcome

Kata exercise training versus treatment-as-usual for the core autism feature of restricted interests and rigid and repetitive behaviours as a direct outcome

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 Treatment-as-usual	**With Kata exercise training**		**Risk with Treatment-as-usual**	**Risk difference with Kata exercise training** (95% CI)

Stereotyped behaviour (measured with: Gilliam Autism Rating Scale (GARS): Stereotyped behaviour; Better indicated by lower values)										
30 (1 study) 15 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	15	15	N/A	N/A	The mean stereotyped behaviour in the intervention groups was **0.9 standard deviations lower** (1.66 to 0.15 lower)
Stereotyped behaviour (measured with: Gilliam Autism Rating Scale (GARS): Stereotyped behaviour; Better indicated by lower values)										
30 (1 study) 19 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	15	15	N/A	N/A	The mean stereotyped behaviour in the intervention groups was **0.76 standard deviations lower** (1.51 to 0.02 lower)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind. The risk of detection bias was also high as the outcome measure was based on interview with carers and teachers who were non-blind and blinding of examiner not reported.
[2] N<400

1.9.4 Nutritional interventions for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

Gluten-free and casein-free diet versus treatment-as-usual for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

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 Treatment-as-usual	**With Gluten-free and casein-free diet**		**Risk with Treatment-as-usual**	**Risk difference with Gluten-free and casein-free diet** (95% CI)
Unusual or bizarre behaviour (measured with: Diagnose of Psykotisk Adferd hos Børn (Diagnosis of Psychotic Behaviour in Children; DIPAB): Unusual or bizarre behaviour (B-scores); Better indicated by lower values)											

20 (1 study) 52 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	10	10	N/A	N/A	The mean unusual or bizarre behaviour in the intervention groups was **0.96 standard deviations lower** (1.9 to 0.02 lower)
Repetitive behaviours (measured with: Autism Diagnostic Observation Schedule (ADOS): Repetitive Behaviours (change score); Better indicated by lower values)										
55 (1 study) 35 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to risk of bias, imprecision	29	26	N/A	N/A	The mean repetitive behaviours in the intervention groups was **0.33 standard deviations lower** (0.86 lower to 0.2 higher)
Stereotyped behaviour (measured with: Gilliam Autism Rating Scale (GARS): Stereotyped behaviour (change score); Better indicated by lower values)										
55 (1 study) 35 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	29	26	N/A	N/A	The mean stereotyped behaviour in the intervention groups was **0.08 standard deviations lower** (0.61 lower to 0.45 higher)

[1] High risk of performance and response bias as intervention administrators (parents) and participants were non-blind. There was also a high risk of detection bias for the DIPAB as although the investigator was blinded to group assignment, this outcome measure was based on parental interview and parents were non-blind to group assignment and other potentially confounding factors
[2] N<400
[3] High risk of attrition bias as over twice as many dropouts in the experimental group relative to the controls (32% in experimental group and 15% in the control group)
[4] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[5] High risk of performance and response bias as intervention administrators (parents) and participants were non-blind, and unclear/unknown risk of detection bias as the identity and blinding of outcome assessors not reported. Also high risk of attrition bias as over twice as many dropouts in the experimental group relative to the controls (32% in experimental group and 15% in the control group)

L-carnosine supplement versus placebo for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

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 Placebo	With L-carnosine supplement		Risk with Placebo	Risk difference with L-carnosine supplement (95% CI)
Stereotyped behaviours (measured with: Gilliam Autism Rating Scale (GARS): Stereotyped behaviour; Better indicated by lower values)											
31 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	14	N/A	N/A	The mean stereotyped behaviours in the intervention groups was **0.41 standard deviations lower** (1.13 lower to 0.3 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.9.5 Sensory intervention for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

Neurofeedback versus treatment-as-usual for the core autism feature of restricted interests and rigid and repetitive behaviours as an indirect outcome

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 Treatment-as-usual	With Neurofeedback		Risk with Treatment-as-usual	Risk difference with Neurofeedback (95% CI)
Parent-rated stereotyped behaviour (measured with: Social Communication Questionnaire (SCQ): Stereotyped behaviour; Better indicated by lower values)											
20 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	10	10	N/A	N/A	The mean parent-rated stereotyped behaviour in the intervention groups was **1.41 standard deviations lower**

										(2.41 to 0.4 lower)
Teacher-rated stereotyped behaviour (measured with: Social Communication Questionnaire (SCQ): Stereotyped behaviour; Better indicated by lower values)										
20 (1 study) 20 weeks	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	10 10	N/A	N/A	The mean teacher-rated stereotyped behaviour in the intervention groups was **0.56 standard deviations higher** (0.33 lower to 1.46 higher)

[1] High risk of performance, response and detection bias as intervention administrators, participants and outcome assessors were non-blind. The risk of other bias due to potential conflict of interest is also high as neurofeedback equipment provided by manufacturer for trial.
[2] N<400
[3] High risk of selective reporting bias as data cannot be extracted for 6-month follow-up
[4] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.10 PSYCHOSOCIAL INTERVENTIONS AIMED AT BEHAVIOUR THAT CHALLENGES

1.10.1 Animal-based intervention for behaviour that challenges as an indirect outcome

Horseback riding versus waitlist control for behaviour that challenges as an indirect outcome

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 Waitlist control	**With Horseback riding**		**Risk with Waitlist control**	**Risk difference with Horseback riding** (95% CI)

Inattention/distractability (measured with: Sensory Profile: Inattention/distractability; Better indicated by lower values)										
34 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	15 19	N/A	N/A	The mean inattention/distractability in the intervention groups was **1.2 standard deviations higher** (0.46 to 1.94 higher)
Sedentary (measured with: Sensory Profile: Sedentary; Better indicated by lower values)										
34 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	15 19	N/A	N/A	The mean sedentary in the intervention groups was **1.14 standard deviations higher** (0.4 to 1.88 higher)

[1] High risk of performance and response bias as intervention administrators and participants non-blind. There is also a high risk of detection bias as outcome measures are parent-rated and parents non-blind
[2] N<400
[3] High risk of selective reporting bias as not all subscales that measure behaviour that challenges are reported, for instance, data are missing for the emotionally reactive subscale

1.10.2 Behavioural interventions for behaviour that challenges as a direct or indirect outcome

Behavioural and medical intervention versus medical intervention only for behaviour that challenges as a direct outcome

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 Behaviour-focused intervention versus treatment-as-usual for behaviour that challenges**		**Risk with Control**	**Risk difference with Behaviour-focused intervention versus treatment-as-usual for behaviour that challenges as a direct**

							as a direct outcome				**outcome** (95% CI)
Illness-related problem behaviour (measured with: Study-specific questionnaire; Better indicated by lower values)											
21 (1 study) 43 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to risk of bias, imprecision	11	10	N/A	N/A	The mean illness-related problem behaviour in the intervention groups was **1.65 standard deviations lower** (2.64 to 0.66 lower)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome assessors were non-blind intervention administrators and the outcome measure was designed specifically for the study and as such lacked formal assessments of reliability and validity
[2] N<400

EIBI versus parent training for behaviour that challenges as an indirect outcome

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 EIBI versus parent training for behaviour that challenges as an indirect outcome**		**Risk with Control**	**Risk difference with EIBI versus parent training for behaviour that challenges as an indirect outcome** (95% CI)
Aggression (parent-rated) (measured with: Achenbach Child Behavior Checklist (Parent report): Aggression; Better indicated by lower values)											
28 (1 study) 260 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⨁⨀⨀⨀ **VERY LOW**[1,2] due to risk of bias, imprecision	13	15	N/A	N/A	The mean aggression (parent-rated) in the intervention groups was **0.36 standard deviations lower** (1.1 lower to 0.39 higher)

Aggression (teacher-rated) (measured with: Achenbach Child Behavior Checklist (Teacher report): Aggression; Better indicated by lower values)										
28 (1 study) 260 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13 15	N/A	N/A	The mean aggression (teacher-rated) in the intervention groups was **0.47 standard deviations higher** (0.28 lower to 1.23 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome measure was non-blind parent- or teacher- completed checklist and checklist was not validated in autism population
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.10.3 Cognitive-behavioural interventions for behaviour that challenges as a direct or indirect outcome

CBT versus waitlist control for behaviour that challenges as a direct outcome

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 CBT for anger management versus waiting-list control for behaviour that challenges**		**Risk with Control**	**Risk difference with CBT for anger management versus waiting-list control for behaviour that challenges** (95% CI)
Parent-reported instances of child anger (measured with: Study-specific parent monitoring of anger: Parent-reported instances of child anger over a week; Better indicated by lower values)											
45 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	21	24	N/A	N/A	The mean parent-reported instances of child anger in the intervention groups was **0.92 standard deviations lower**

										(1.54 to 0.3 lower)
Parent-reported instances of child anger (measured with: Study-specific parent monitoring of anger: Parent-reported instances of child anger over a week; Better indicated by lower values)										
45 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	21	24	N/A	N/A	The mean parent-reported instances of child anger in the intervention groups was **1.03 standard deviations lower** (1.65 to 0.4 lower)
Parent confidence in child managing own anger (measured with: Study-specific parent monitoring of anger: Parent-reported confidence in their child managing their own anger; Better indicated by lower values)										
45 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	21	24	N/A	N/A	The mean parent confidence in child managing own anger in the intervention groups was **0.61 standard deviations higher** (0 to 1.21 higher)
Parent confidence in child managing own anger (measured with: Study-specific parent monitoring of anger: Parent-reported confidence in their child managing their own anger; Better indicated by lower values)										
45 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	21	24	N/A	N/A	The mean parent confidence in child managing own anger in the intervention groups was **1.1 standard deviations higher** (0.47 to 1.74 higher)
[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome measure parent-rated and parents were non-blind										

[2] N<400
[3] High risk of selective reporting bias as data cannot be extracted for the Children's Inventory of Anger (ChIA-P) as no measure of variability is reported

CBT versus waitlist control for behaviour that challenges as an indirect outcome

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 Waitilist control	**With CBT for anxiety**		**Risk with Waitilist control**	**Risk difference with CBT for anxiety** (95% CI)
Hyperactivity and conduct problems (parent-rated) (measured with: Strengths and Difficulties Questionnaire: Externalising; Better indicated by lower values)											
47 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	19	28	N/A	N/A	The mean hyperactivity and conduct problems (parent-rated) in the intervention groups was **0.62 standard deviations lower** (1.22 to 0.03 lower)
Hyperactivity and conduct problems (teacher-rated) (measured with: Strengths and Difficulties Questionnaire: Externalising; Better indicated by lower values)											
47 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,3] due to risk of bias, imprecision	19	28	N/A	N/A	The mean hyperactivity and conduct problems (teacher-rated) in the intervention groups was **0.62 standard deviations lower** (1.21 to 0.02 lower)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome meausre parent- rated and parents non-blind
[2] N<400
[3] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear/unknown as teacher-rated and blinding of teachers is not reported

1.10.4 Parent training for behaviour that challenges as a direct or indirect outcome

Parent training versus treatment-as-usual for behaviour that challenges as a direct or indirect outcome

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 Parent training versus treatment-as-usual for behaviour that challenges		Risk with Control	Risk difference with Parent training versus treatment-as-usual for behaviour that challenges (95% CI)
Number of problem behaviours (combined workshop + individual sessions) (measured with: Eyberg Child Behaviour Inventory (ECBI): Number of problem behaviours; Better indicated by lower values)											
51 (1 study) 4-10 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	15	36	N/A	N/A	The mean number of problem behaviours (combined workshop + individual sessions) in the intervention groups was **1.26 standard deviations lower** (1.91 to 0.61 lower)
Number of problem behaviours (combined workshop + individual sessions) (measured with: Eyberg Child Behaviour Inventory (ECBI): Number of problem behaviours; Better indicated by lower values)											
51 (1 study) 13-19 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	15	36	N/A	N/A	The mean number of problem behaviours (combined workshop + individual sessions) in the intervention groups was **1.23 standard deviations lower** (1.88 to 0.58 lower)

Intensity of problem behaviours (individual sessions) (measured with: Eyberg Child Behaviour Inventory (ECBI): Intensity of problem behaviours; Better indicated by lower values)											
33 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	15	18	N/A	N/A	The mean intensity of problem behaviours (individual sessions) in the intervention groups was **1.41 standard deviations lower** (2.18 to 0.63 lower)
Intensity of problem behaviours (individual sessions) (measured with: Eyberg Child Behaviour Inventory (ECBI): Intensity of problem behaviours; Better indicated by lower values)											
33 (1 study) 19 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	15	18	N/A	N/A	The mean intensity of problem behaviours (individual sessions) in the intervention groups was **1.35 standard deviations lower** (2.12 to 0.59 lower)
Intensity of problem behaviours (workshop) (measured with: Eyberg Child Behaviour Inventory (ECBI): Intensity of problem behaviours; Better indicated by lower values)											
33 (1 study) 4 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	18	N/A	N/A	The mean intensity of problem behaviours (workshop) in the intervention groups was **0.60 standard deviations lower** (1.30 lower to 0.10 higher)
Intensity of problem behaviours (workshop) (measured with: Eyberg Child Behaviour Inventory (ECBI): Intensity of problem behaviours; Better indicated by lower values)											
33 (1 study)	serious[1]	no serious	no serious	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3]	15	18	N/A	N/A	The mean intensity of problem behaviours (workshop) in the

13 weeks		inconsistency	indirectness			due to risk of bias, imprecision				intervention groups was **0.59 standard deviations lower** (1.30 lower to 0.11 higher)
Problem behaviour (PEC+PEBM combined) (measured with: Developmental Behaviour Checklist (DBC): Total Behaviour Problem Score (TBPS); Better indicated by lower values)										
103 (1 study) 46 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	35	68	N/A	N/A	The mean problem behaviour (pec+pebm combined) in the intervention groups was **0.35 standard deviations lower** (0.76 lower to 0.06 higher)

[1] High risk of performance bias as intervention administrators were non-blind, and high risk of detection bias as outcome assessors were non-blind parents who were involved in the intervention
[2] N<400
[3] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Combined parent training and antipsychotic versus antipsychotic-only for behaviour that challenges as a direct outcome

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 Combined antipsychotic and parent training versus antipsychotic only for behaviour that challenges as a direct outcome**		**Risk with Control**	**Risk difference with Combined antipsychotic and parent training versus antipsychotic only for behaviour that challenges as a direct outcome** (95% CI)
Noncompliant behaviour in everyday circumstances (measured with: Home Situations Questionnaire (HSQ): Severity; Better indicated by lower values)											
95 (1 study) 24 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias,	40	55	N/A	N/A	The mean noncompliant behaviour in everyday circumstances in the intervention groups was

						imprecision					**0.33 standard deviations lower** (0.74 lower to 0.08 higher)
Noncompliant behaviour in everyday circumstances (measured with: Home Situations Questionnaire (HSQ): Severity; Better indicated by lower values)											
87 (1 study) 80 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	36	51	N/A	N/A	The mean noncompliant behaviour in everyday circumstances in the intervention groups was **0.17 standard deviations lower** (0.6 lower to 0.26 higher)
Noncompliant behaviour in everyday circumstances (measured with: Study-specific noncompliance index based on the Vineland Adaptive Behaviour Scale (VABS) Daily living skills subscale; Better indicated by lower values)											
124 (1 study) 24 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	49	75	N/A	N/A	The mean noncompliant behaviour in everyday circumstances in the intervention groups was **0.46 standard deviations lower** (0.83 to 0.1 lower)
Irritability (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation; Better indicated by lower values)											
95 (1 study) 24 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	40	55	N/A	N/A	The mean irritability in the intervention groups was **0.43 standard deviations lower** (0.85 to 0.02 lower)
Irritability (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation; Better indicated by lower values)											
87 (1 study) 80 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	36	51	N/A	N/A	The mean irritability in the intervention groups was **0.33 standard deviations lower** (0.75 lower to 0.1 higher)

Lethargy/Social withdrawal (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal; Better indicated by lower values)											
95 (1 study) 24 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	40	55	N/A	N/A	The mean lethargy/social withdrawal in the intervention groups was **0.36 standard deviations lower** (0.77 lower to 0.06 higher)
Lethargy/Social withdrawal (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal; Better indicated by lower values)											
87 (1 study) 80 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	36	51	N/A	N/A	The mean lethargy/social withdrawal in the intervention groups was **0.46 standard deviations lower** (0.89 to 0.03 lower)
Stereotypic behaviour (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic Behaviour; Better indicated by lower values)											
95 (1 study) 24 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	40	55	N/A	N/A	The mean stereotypic behaviour in the intervention groups was **0.63 standard deviations lower** (1.04 to 0.21 lower)
Stereotypic behaviour (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic Behaviour; Better indicated by lower values)											
87 (1 study) 80 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	36	51	N/A	N/A	The mean stereotypic behaviour in the intervention groups was **0.35 standard deviations lower** (0.78 lower to 0.08 higher)
Hyperactivity (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance; Better indicated by lower values)											
95 (1 study) 24 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias,	40	55	N/A	N/A	The mean hyperactivity in the intervention groups was **0.48 standard deviations lower**

						imprecision					(0.89 to 0.07 lower)
Hyperactivity (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance; Better indicated by lower values)											
87 (1 study) 80 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	36	51	N/A	N/A	The mean hyperactivity in the intervention groups was **0.13 standard deviations lower** (0.56 lower to 0.29 higher)
Inappropriate speech (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech; Better indicated by lower values)											
95 (1 study) 24 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	40	55	N/A	N/A	The mean inappropriate speech in the intervention groups was **0.23 standard deviations lower** (0.63 lower to 0.18 higher)
Inappropriate speech (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech; Better indicated by lower values)											
87 (1 study) 80 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	36	51	N/A	N/A	The mean inappropriate speech in the intervention groups was **0.02 standard deviations higher** (0.41 lower to 0.44 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome measure based on interview with parents who were non-blind. Also high risk of attrition bias due to higher dropout rates in the experimental (combined risperidone and parent training) group (N=20; 27% attrition) than the control (risperidone only) group (N=9; 18% attrition)
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] N<400

Combined parent training and early intervention centre programme versus early intervention centre programme only for behaviour that challenges as an indirect outcome

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 Combined parent training and early intervention centre programme versus early intervention centre programme only for behaviour that challenges as an indirect outcome		Risk with Control	Risk difference with Combined parent training and early intervention centre programme versus early intervention centre programme only for behaviour that challenges as an indirect outcome (95% CI)
Parent-reported behaviour that challenges (mixed ASD & DD sample) (measured with: Behavior Screening Questionnaire (BSQ): Total; Better indicated by lower values)											
58 (1 study) 40 weeks	serious[1]	no serious inconsistency	serious[2]	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, indirectness, imprecision	28	30	N/A	N/A	The mean parent-reported behaviour that challenges (mixed asd & dd sample) in the intervention groups was **0.02 standard deviations lower** (0.54 lower to 0.49 higher)
Parent-reported behaviour that challenges (mixed ASD & DD sample) (measured with: Behavior Screening Questionnaire (BSQ): Total; Better indicated by lower values)											
50 (1 study) 108 weeks	serious[1]	no serious inconsistency	serious[2]	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, indirectness, imprecision	23	27	N/A	N/A	The mean parent-reported behaviour that challenges (mixed asd & dd sample) in the intervention groups was **0.16 standard deviations lower** (0.71 lower to 0.4 higher)
Teacher-rated behaviour that challenges (mixed ASD & DD sample) (measured with: Preschool Behavior Checklist (PBCL): Total; Better indicated by lower values)											
53 (1 study) 40 weeks	serious[4]	no serious inconsistency	serious[2]	serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,4,5] due to risk of bias, indirectness, imprecision	26	27	N/A	N/A	The mean teacher-rated behaviour that challenges (mixed asd & dd sample) in the intervention groups was **0.67 standard deviations lower** (1.23 to 0.12 lower)
Teacher-rated behaviour that challenges (ASD-only sample) (measured with: Preschool Behavior Checklist (PBCL): Total; Better indicated by lower values)											

34 (1 study) 40 weeks	serious[4]	no serious inconsistency	no serious indirectness	serious[5]	undetected	⊕⊕⊖⊖ **LOW**[4,5] due to risk of bias, imprecision	18	16	N/A	N/A	The mean teacher-rated behaviour that challenges (asd-only sample) in the intervention groups was **0.98 standard deviations lower** (1.69 to 0.26 lower)
Teacher-rated behaviour that challenges (mixed ASD & DD sample) (measured with: Preschool Behavior Checklist (PBCL): Total; Better indicated by lower values)											
46 (1 study) 108 weeks	serious[4]	no serious inconsistency	serious[2]	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3,4] due to risk of bias, indirectness, imprecision	23	23	N/A	N/A	The mean teacher-rated behaviour that challenges (mixed asd & dd sample) in the intervention groups was **0.11 standard deviations lower** (0.68 lower to 0.47 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear/unknown as although there was a blinded psychologist outcome assessor this outcome measure relied on non-blind parental report
[2] Population was indirect (as the sample included participants with developmental delay or language delay without autism)
[3] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[4] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome assessors were non-blind teachers
[5] N<400

1.10.5 Social-communication interventions for behaviour that challenges as an indirect outcome

Social skills group versus treatment-as-usual for behaviour that challenges as an indirect outcome

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 Social skills groups versus treatment-as-usual for behaviour that challenges as an indirect outcome**		**Risk with Control**	**Risk difference with Social skills groups versus treatment-as-usual for behaviour that challenges as an indirect outcome** (95% CI)
Conflict (parent-rated) (measured with: Quality of Play Questionnaire (QPQ): Conflict; Better indicated by lower values)											

95 (2 studies) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	43	52	N/A	N/A	The mean conflict (parent-rated) in the intervention groups was **0.6 standard deviations lower** (1.01 to 0.18 lower)
Conflict (self-rated) (measured with: Quality of Play Questionnaire (QPQ): Conflict; Better indicated by lower values)											
33 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to risk of bias, imprecision	16	17	N/A	N/A	The mean conflict (self-rated) in the intervention groups was **0.09 standard deviations lower** (0.77 lower to 0.59 higher)
Intrusive/aggressive behaviour (parent-rated) (measured with: Social Skills Rating System (SSRS): Externalizing or Social Skills Rating System (SSRS): Problem behaviours; Better indicated by lower values)											
101 (2 studies) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	49	52	N/A	N/A	The mean intrusive/aggressive behaviour (parent-rated) in the intervention groups was **0.78 standard deviations lower** (1.19 to 0.37 lower)
Intrusive/aggressive behaviour (teacher-rated) (measured with: Pupil Evaluation Inventory (PEI): Aggression; Better indicated by lower values)											
59 (1 study) 12 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias, imprecision	28	31	N/A	N/A	The mean intrusive/aggressive behaviour (teacher-rated) in the intervention groups was **0.24 standard deviations lower** (0.75 lower to 0.28 higher)
Social withdrawal (parent-rated) (measured with: Social Skills Rating System (SSRS): Internalizing or Behavior Assessment System for Children, 2nd ed., parent rated (BASC-2-PRS): Withdrawal; Better indicated by lower values)											
104 (2 studies) 6-12 weeks	serious[1]	very serious[6]	no serious indirectness	serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,6] due to risk of bias, inconsistency, imprecision	51	53	N/A	N/A	The mean social withdrawal (parent-rated) in the intervention groups was **0.68 standard deviations lower** (1.08 to 0.28 lower)
Social withdrawal (teacher-rated) (measured with: Pupil Evaluation Inventory (PEI): Withdrawal; Better indicated by lower values)											
59 (1 study) 12 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,5] due to risk of bias,	28	31	N/A	N/A	The mean social withdrawal (teacher-rated) in the intervention groups was

						imprecision				**0.04 standard deviations lower** (0.55 lower to 0.47 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as parent-rated and parents were non-blind and involved in the intervention.
[2] N<400
[3] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as self-rated
[4] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[5] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as teacher-rated and teachers were non-blind
[6] Substantial to considerable heterogeneity

LEGO therapy versus SULP for behaviour that challenges as an indirect outcome

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 LEGO therapy versus Social Use of Language Programme (SULP) for behaviour that challenges as an indirect outcome**		**Risk with Control**	**Risk difference with LEGO therapy versus Social Use of Language Programme (SULP) for behaviour that challenges as an indirect outcome** (95% CI)
Maladaptive behaviour (measured with: Vineland Adaptive Behaviour Scale (VABS): Maladaptive Behaviour Index; Better indicated by lower values)											
31 (1 study) 18 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1] due to risk of bias, imprecision	15	16	N/A	N/A	The mean maladaptive behaviour in the intervention groups was **0.51 standard deviations lower** (1.23 lower to 0.21 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.11 PHARMACOLOGICAL INTERVENTIONS AIMED AT BEHAVIOUR THAT CHALLENGES

1.11.1 Anticonvulsants for behaviour that challenges as a direct outcome

Divalproex versus placebo for behaviour that challenges as a direct outcome

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 Anticonvulsants versus placebo for behaviour that challenges as a direct outcome		Risk with Control	Risk difference with Anticonvulsants versus placebo for behaviour that challenges as a direct outcome (95% CI)
Irritability (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation; Better indicated by lower values)											
57 (2 studies) 8-12 weeks	no serious risk of bias	serious[1]	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to inconsistency, imprecision	25	32	N/A	N/A	The mean irritability in the intervention groups was **0.05 standard deviations lower** (0.58 lower to 0.48 higher)
Irritability (measured with: Overt Aggression Scale (OAS): Irritability; Better indicated by lower values)											
27 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency[1]	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to imprecision	11	16	N/A	N/A	The mean irritability in the intervention groups was **0.43 standard deviations lower** (1.21 lower to 0.35 higher)
Aggression (measured with: Overt Aggression Scale (OAS): Total; Better indicated by lower values)											
30 (1 study) 8 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	14	16	N/A	N/A	The mean aggression in the intervention groups was **0.03 standard deviations higher**

										(0.69 lower to 0.75 higher)
Global severity (measured with: Clinical Global Impression Scale (CGI-S): Severity; Better indicated by lower values)										
30 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	14	16	N/A	N/A	The mean global severity in the intervention groups was **0 standard deviations higher** (0.72 lower to 0.72 higher)
Global improvement (measured with: Clinical Global Impression Scale (CGI-I): Improvement; Better indicated by lower values)										
30 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	14	16	N/A	N/A	The mean global improvement in the intervention groups was **0.43 standard deviations lower** (1.16 lower to 0.29 higher)
Global improvement (assessed with: Dichotomous: Positive treatment response ('much improved/very improved' on CGI-improvement))										
27 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	1/11 (9.1%)	10/16 (62.5%)	**RR 6.88** (1.02 to 46.28)	**Study population** **91 per 1000** **Moderate** **91 per 1000**	**535 more per 1000** (from 2 more to 1000 more) **535 more per 1000** (from 2 more to 1000 more)

[1] Moderate heterogeneity
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] High risk of selective reporting bias as results for the teacher-rated OAS are not reported
[4] Events<300

Topiramate and risperidone versus placebo and risperidone for behaviour that challenges as a direct outcome

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 Combined anticonvulsants and antipsychotics versus combined placebo and antipsychotics for behaviour that challenges as a direct outcome		Risk with Control	Risk difference with Combined anticonvulsants and antipsychotics versus combined placebo and antipsychotics for behaviour that challenges as a direct outcome (95% CI)
Irritability (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation; Better indicated by lower values)											
40 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	20	20	N/A	N/A	The mean irritability in the intervention groups was **1.88 standard deviations lower** (2.63 to 1.12 lower)
Lethargy/Social withdrawal (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal; Better indicated by lower values)											
40 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	20	20	N/A	N/A	The mean lethargy/social withdrawal in the intervention groups was **0.25 standard deviations lower** (0.88 lower to 0.37 higher)
Stereotypic behaviour (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic Behaviour; Better indicated by lower values)											
40 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	20	20	N/A	N/A	The mean stereotypic behaviour in the intervention groups was **2.02 standard deviations lower** (2.8 to 1.25 lower)
Hyperactivity (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance; Better indicated by lower values)											
40 (1 study) 8 weeks	no serious risk of	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to	20	20	N/A	N/A	The mean hyperactivity in the intervention groups was **1.87 standard deviations**

	bias					imprecision				**lower** (2.63 to 1.12 lower)
Inappropriate speech (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech; Better indicated by lower values)										
40 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	20	20	N/A	N/A	The mean inappropriate speech in the intervention groups was **0.16 standard deviations lower** (0.78 lower to 0.46 higher)

[1] N<400
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.11.2 Antidepressants for behaviour that challenges as an indirect outcome

Citalopram versus placebo for behaviour that challenges as an indirect outcome

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 Antidepressants versus placebo for behaviour that challenges as an indirect outcome**		**Risk with Control**	**Risk difference with Antidepressants versus placebo for behaviour that challenges as an indirect outcome** (95% CI)
Irritability (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation; Better indicated by lower values)											
149 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	76	73	N/A	N/A	The mean irritability in the intervention groups was **0.01 standard deviations lower** (0.33 lower to 0.31 higher)
Lethargy /Social withdrawal (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal; Better indicated by lower values)											
149	no	no serious	no serious	serious[1]	undetected	⊕⊕⊕⊖	76	73	N/A	N/A	The mean lethargy /social

(1 study) 12 weeks	serious risk of bias	inconsistency	indirectness			**MODERATE**[1] due to imprecision				withdrawal in the intervention groups was **0.01 standard deviations lower** (0.33 lower to 0.31 higher)
Stereotypic behaviour (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic Behaviour; Better indicated by lower values)										
149 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	76	73	N/A	N/A	The mean stereotypic behaviour in the intervention groups was **0.05 standard deviations higher** (0.27 lower to 0.37 higher)
Hyperactivity (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance; Better indicated by lower values)										
149 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	76	73	N/A	N/A	The mean hyperactivity in the intervention groups was **0.09 standard deviations higher** (0.23 lower to 0.41 higher)
Inappropriate speech (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech; Better indicated by lower values)										
149 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	76	73	N/A	N/A	The mean inappropriate speech in the intervention groups was **0.06 standard deviations higher** (0.26 lower to 0.38 higher)
[1] N<400										

1.11.3 Antihistamines for behaviour that challenges as a direct outcome

Cyproheptadine and haloperidol versus placebo and haloperidol for behaviour that challenges as a direct outcome

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 Combined antihistamine and antipsychotic versus combined antipsychotic and placebo for behaviour that challenges as a direct outcome		Risk with Control	Risk difference with Combined antihistamine and antipsychotic versus combined antipsychotic and placebo for behaviour that challenges as a direct outcome (95% CI)
Behaviour that challenges (measured with: Aberrant Behaviour Checklist (ABC): Total (Change Score); Better indicated by lower values)											
40 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	20	20	N/A	N/A	The mean behaviour that challenges in the intervention groups was **0.98 standard deviations lower** (1.64 to 0.32 lower)
[1] N<400											

1.11.4 Antioxidants for behaviour that challenges as a direct outcome

N-acetylcysteine versus placebo for behaviour that challenges as a direct outcome

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 Placebo	With Antioxidants		Risk with Placebo	Risk difference with Antioxidants (95% CI)
Irritability (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation; Better indicated by lower values)											
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	14	N/A	N/A	The mean irritability in the intervention groups was **0.7 standard deviations lower** (1.46 lower to 0.05 higher)

Lethargy/Social Withdrawal (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal; Better indicated by lower values)											
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	14	N/A	N/A	The mean lethargy/social withdrawal in the intervention groups was **0.31 standard deviations higher** (0.43 lower to 1.04 higher)
Stereotypic behaviour (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic Behaviour; Better indicated by lower values)											
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	14	N/A	N/A	The mean stereotypic behaviour in the intervention groups was **0.36 standard deviations lower** (1.1 lower to 0.37 higher)
Hyperactivity (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance; Better indicated by lower values)											
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	14	N/A	N/A	The mean hyperactivity in the intervention groups was **0.73 standard deviations lower** (1.49 lower to 0.03 higher)
Inappropriate speech (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech; Better indicated by lower values)											
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	14	N/A	N/A	The mean inappropriate speech in the intervention groups was **0.34 standard deviations lower** (1.07 lower to 0.4 higher)
Global severity (measured with: Clinical Global Impression Scale (CGI-S): Severity; Better indicated by lower values)											
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	14	N/A	N/A	The mean global severity in the intervention groups was **0.46 standard deviations lower**

										(1.19 lower to 0.28 higher)
Global improvement (measured with: Clinical Global Impression Scale (CGI-I): Improvement; Better indicated by lower values)										
29 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	14	N/A	N/A	The mean global improvement in the intervention groups was **0.29 standard deviations lower** (1.02 lower to 0.44 higher)
[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)										

1.11.5 Antipsychotics for behaviour that challenges as a direct or indirect outcome

Antipsychotic (risperidone or aripiprazole) versus placebo for behaviour that challenges as a direct outcome

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 Antipsychotics versus placebo for behaviour that challenges**		**Risk with Control**	**Risk difference with Antipsychotics versus placebo for behaviour that challenges** (95% CI)
Positive treatment response (risperidone or aripiprazole) (assessed with: Positive treatment response (clinician-rated: >25% improvement on ABC-Irritability with or without 'much improved/very improved' on CGI-improvement))											
501 (4 studies) 6-8 weeks	no serious risk of bias	very serious[1]	no serious indirectness	no serious imprecision	undetected	⊕⊕⊖⊖ **LOW**[1] due to inconsistency	44/184 (23.9%)	183/317 (57.7%)	**RR 2.27** (1.75 to 2.94)	**Study population**	
										239 per 1000	**304 more per 1000** (from 179 more to 464 more)
										Moderate	
										245 per 1000	**311 more per 1000** (from 184 more to 475 more)

Positive treatment response (risperidone) (assessed with: Positive treatment response (clinician-rated: >25% improvement on ABC-Irritability with or without 'much improved/very improved' on CGI-improvement))										
193 (2 studies) 6-8 weeks	no serious risk of bias	very serious[1]	no serious indirectness	serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to inconsistency, imprecision	20/86 (23.3%) 73/107 (68.2%)	**RR 2.72** (1.85 to 3.99)	**Study population**	
									233 per 1000	**400 more per 1000** (from 198 more to 695 more)
									Moderate	
									245 per 1000	**421 more per 1000** (from 208 more to 733 more)
Positive treatment response (aripiprazole) (assessed with: Positive treatment response (clinician-rated: >25% improvement on ABC-Irritability with or without 'much improved/very improved' on CGI-improvement))										
308 (2 studies) 8 weeks	no serious risk of bias	very serious[1]	no serious indirectness	serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to inconsistency, imprecision	24/98 (24.5%) 110/210 (52.4%)	**RR 1.95** (1.37 to 2.78)	**Study population**	
									245 per 1000	**233 more per 1000** (from 91 more to 436 more)
									Moderate	
									245 per 1000	**233 more per 1000** (from 91 more to 436 more)
Positive treatment response (risperidone) (assessed with: Dichotomous: Positive treatment response (<3 "definitely improved" or better on 9-point parent-defined target symptom scale))										
87 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	9/43 (20.9%) 31/44 (70.5%)	**RR 3.37** (1.83 to 6.21)	**Study population**	
									209 per 1000	**496 more per 1000** (from 174 more to 1000 more)
									Moderate	

									209 per 1000	**495 more per 1000** (from 173 more to 1000 more)
Maladaptive behaviour (measured with: Vineland Adaptive Behaviour Scale (VABS): Maladaptive Behaviour Index; Better indicated by lower values)										
101 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊕⊖ **MODERATE**[3] due to imprecision	52	49	N/A	N/A	The mean maladaptive behaviour in the intervention groups was **1.17 standard deviations lower** (1.59 to 0.75 lower)
Irritability (risperidone or aripiprazole) (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation (Endpoint or Change score); Better indicated by lower values)										
363 (4 studies) 6-8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊕⊖ **MODERATE**[3] due to imprecision	173	190	N/A	N/A	The mean irritability (risperidone or aripiprazole) in the intervention groups was **0.92 standard deviations lower** (1.14 to 0.7 lower)
Irritability (risperidone) (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation (Endpoint or Change score); Better indicated by lower values)										
268 (3 studies) 6-8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊕⊖ **MODERATE**[3] due to imprecision	124	144	N/A	N/A	The mean irritability (risperidone) in the intervention groups was **0.96 standard deviations lower** (1.22 to 0.71 lower)
Irritability (aripiprazole) (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation (Endpoint or Change score); Better indicated by lower values)										
95 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊕⊖ **MODERATE**[3] due to imprecision	49	46	N/A	N/A	The mean irritability (aripiprazole) in the intervention groups was **0.81 standard deviations lower** (1.23 to 0.39 lower)

Lethargy/Social withdrawal (risperidone or aripiprazole) (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal (Endpoint and Change scores); Better indicated by lower values)										
486 (4 studies) 8 weeks	serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to risk of bias	188	298	N/A	N/A	The mean lethargy/social withdrawal (risperidone or aripiprazole) in the intervention groups was **0.28 standard deviations lower** (0.47 to 0.08 lower)
Lethargy/Social withdrawal (risperidone) (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal (Endpoint and Change scores); Better indicated by lower values)										
178 (2 studies) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊕⊖ **MODERATE**[3] due to imprecision	90	88	N/A	N/A	The mean lethargy/social withdrawal (risperidone) in the intervention groups was **0.45 standard deviations lower** (0.75 to 0.15 lower)
Lethargy/Social Withdrawal (aripiprazole) (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal (Endpoint and Change scores); Better indicated by lower values)										
308 (2 studies) 8 weeks	serious[4]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3,4] due to risk of bias, imprecision	98	210	N/A	N/A	The mean lethargy/social withdrawal (aripiprazole) in the intervention groups was **0.15 standard deviations lower** (0.40 lower to 0.10 higher)
Stereotypic behaviour (risperidone or aripiprazole) (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic Behaviour (Endpoint and Change scores); Better indicated by lower values)										
485 (4 studies) 8 weeks	serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to risk of bias	188	297	N/A	N/A	The mean stereotypic behaviour (risperidone or aripiprazole) in the intervention groups was **0.48 standard deviations lower**

										(0.68 to 0.29 lower)
Stereotypic behaviour (risperidone) (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic Behaviour (Endpoint and Change scores); Better indicated by lower values)										
177 (2 studies) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊕⊖ **MODERATE**[3] due to imprecision	90	87	N/A	N/A	The mean stereotypic behaviour (risperidone) in the intervention groups was **0.34 standard deviations lower** (0.64 to 0.05 lower)
Stereotypic behaviour (aripiprazole) (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic Behaviour (Endpoint and Change scores); Better indicated by lower values)										
308 (2 studies) 8 weeks	serious[4]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3,4] due to risk of bias, imprecision	98	210	N/A	N/A	The mean stereotypic behaviour (aripiprazole) in the intervention groups was **0.59 standard deviations lower** (0.84 to 0.33 lower)
Hyperactivity (risperidone or aripiprazole) (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance (Endpoint or Change score); Better indicated by lower values)										
484 (4 studies) 8 weeks	serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to risk of bias	187	297	N/A	N/A	The mean hyperactivity (risperidone or aripiprazole) in the intervention groups was **0.84 standard deviations lower** (1.04 to 0.64 lower)
Hyperactivity (risperidone) (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance (Endpoint or Change score); Better indicated by lower values)										
176 (2 studies) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊕⊖ **MODERATE**[3] due to imprecision	89	87	N/A	N/A	The mean hyperactivity (risperidone) in the intervention groups was **1.03 standard deviations lower**

										(1.34 to 0.71 lower)
Hyperactivity (aripiprazole) (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance (Endpoint or Change score); Better indicated by lower values)										
308 (2 studies) 8 weeks	serious[4]	serious[5]	no serious indirectness	serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4,5] due to risk of bias, inconsistency, imprecision	98	210	N/A	N/A	The mean hyperactivity (aripiprazole) in the intervention groups was **0.72 standard deviations lower** (0.97 to 0.46 lower)
Inappropriate speech (risperidone or aripiprazole) (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech (Endpoint and Change scores); Better indicated by lower values)										
485 (4 studies) 8 weeks	serious[4]	serious[5]	no serious indirectness	no serious imprecision	undetected	⊕⊕⊖⊖ **LOW**[4,5] due to risk of bias, inconsistency	187	298	N/A	N/A	The mean inappropriate speech (risperidone or aripiprazole) in the intervention groups was **0.54 standard deviations lower** (0.74 to 0.35 lower)
Inappropriate Speech (risperidone) (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech (Endpoint and Change scores); Better indicated by lower values)										
178 (2 studies) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊕⊖ **MODERATE**[3] due to imprecision	90	88	N/A	N/A	The mean inappropriate speech (risperidone) in the intervention groups was **0.66 standard deviations lower** (0.96 to 0.36 lower)
Inappropriate Speech (aripiprazole) (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech (Endpoint and Change scores); Better indicated by lower values)										
307 (2 studies) 8 weeks	serious[4]	very serious[6]	no serious indirectness	serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4,6] due to risk of bias, inconsistency, imprecision	97	210	N/A	N/A	The mean inappropriate speech (aripiprazole) in the intervention groups was **0.46 standard deviations lower**

										(0.72 to 0.20 lower)
Parent-defined target symptoms (measured with: Parent-defined target symptom scale (9-point) or Visual Analog Scale for the most troublesome symptom (VAS-MS); Better indicated by lower values)										
163 (2 studies) 8 weeks	serious[7]	very serious[8]	no serious indirectness	serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,7,8] due to risk of bias, inconsistency, imprecision	80	83	N/A	N/A	The mean parent-defined target symptoms in the intervention groups was **0.96 standard deviations lower** (1.29 to 0.63 lower)
Global state: Positive treatment response (risperidone) (assessed with: Dichotomous: Positive treatment response ('much improved/very improved' on CGI-improvement))										
171 (2 studies) 6-8 weeks	serious[9]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,9] due to risk of bias, imprecision	12/72 (16.7%)	45/99 (45.5%)	**RR 2.83** (1.61 to 4.95)	**Study population**	
										167 per 1000	**305 more per 1000** (from 102 more to 658 more)
										Moderate	
										166 per 1000	**304 more per 1000** (from 101 more to 656 more)
Global state: Symptom severity (risperidone or aripiprazole) (measured with: Clinical Global Impression Scale (CGI-S): Severity (Endpoint or Change Scores); Better indicated by lower values)										
273 (2 studies) 6-8 weeks	serious[10]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3,10] due to risk of bias, imprecision	75	198	N/A	N/A	The mean global state: symptom severity (risperidone or aripiprazole) in the intervention groups was **0.32 standard deviations lower** (0.59 to 0.05 lower)
Global state: Symptom severity (risperdione) (measured with: Clinical Global Impression Scale (CGI-S): Severity (change score); Better indicated by lower values)										

92 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3,11]	undetected	⊕⊕⊖⊖ **LOW**[3,11] due to imprecision	34	58	N/A	N/A	The mean global state: symptom severity (risperdione) in the intervention groups was **0.28 standard deviations lower** (0.71 lower to 0.14 higher)
Global state: Symptom severity (aripiprazole) (measured with: Clinical Global Impression Scale (CGI-S): Severity; Better indicated by lower values)											
181 (1 study) 8 weeks	serious[10]	no serious inconsistency	no serious indirectness	very serious[11]	undetected	⊕⊖⊖⊖ **VERY LOW**[10,11] due to risk of bias, imprecision	41	140	N/A	N/A	The mean global state: symptom severity (aripiprazole) in the intervention groups was **0.34 standard deviations lower** (0.69 lower to 0.01 higher)
Global state: Improvement (risperidone) (measured with: Clinical Global Impression Scale (CGI-I): Improvement; Better indicated by lower values)											
77 (1 study) 8 weeks	serious[9]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3,9] due to risk of bias, imprecision	38	39	N/A	N/A	The mean global state: improvement (risperidone) in the intervention groups was **0.98 standard deviations lower** (1.45 to 0.51 lower)

[1] Substantial to considerable heterogeneity
[2] Events<300
[3] N<400
[4] With the exception of RUPPRISPERIDONE2001, the blinding is unclear for the trials as the papers state 'double-blind' but give no further detail with regards to who is blinded, i.e. participant, parent, investigator, intervention administrator, outcome assessor.
[5] Moderate heterogeneity
[6] Substantial heterogeneity
[7] In RUPPRISPERIDONE2001 a study-specific outcome measure without indpendent reliability and validity data was used and in SHEA2004/PANDINA2007 the blinding is unclear as the paper states 'double-blind' but gives no further detail with regards to who is blinded, i.e. participant, parent, investigator, intervention administrator, outcome assessor
[8] Substantial to considerable heterogeneity
[9] Blinding is unclear in SHEA2004/PANDINA2007 as paper states 'double-blind' but gives no further detail with regards to who is blinded, i.e. participant, parent, investigator, intervention administrator, outcome assessor
[10] Blinding is unclear in MARCUS2009 as paper states 'double-blind' but gives no further detail with regards to who is blinded, i.e. participant, parent, investigator, intervention administrator, outcome assessor
[11] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Low dose antipsychotic (risperidone or aripiprazole) versus placebo for behaviour that challenges as a direct outcome

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 Low dose antipsychotics versus placebo for behaviour that challenges		Risk with Control	Risk difference with Low dose antipsychotics versus placebo for behaviour that challenges (95% CI)
Positive treatment response (risperidone or aripiprazole) (assessed with: Dichotomous: Positive treatment response (>25% improvement on ABC-Irritability) or Dichotomous: Positive treatment response (>25% improvement on ABC-Irritability & 'much improved/very improved' on CGI-improvement))											
164 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to risk of bias, imprecision	31/83 (37.3%)	44/81 (54.3%)	**RR 1.46** (1.03 to 2.06)	**Study population**	
										373 per 1000	**172 more per 1000** (from 11 more to 396 more)
										Moderate	
										379 per 1000	**174 more per 1000** (from 11 more to 402 more)
Positive treatment response (risperidone) (assessed with: Dichotomous: Positive treatment response (>25% improvement on ABC-Irritability))											
63 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	undetected	⨁⨁⨀⨀ **LOW**[2,3] due to imprecision	14/34 (41.2%)	15/29 (51.7%)	**RR 1.26** (0.74 to 2.14)	**Study population**	
										412 per 1000	**107 more per 1000** (from 107 fewer to 469 more)
										Moderate	
										379 per 1000	**99 more per 1000** (from 99 fewer to 432 more)
Positive treatment response (aripiprazole) (assessed with: Dichotomous: Positive treatment response (>25% improvement on ABC-Irritability & 'much improved/very improved' on CGI-improvement))											

101 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	17/49 (34.7%)	29/52 (55.8%)	**RR 1.61** (1.02 to 2.53)	**Study population**	
										347 per 1000	**212 more per 1000** (from 7 more to 531 more)
										Moderate	
										379 per 1000	**231 more per 1000** (from 8 more to 580 more)
Irritability (risperidone) (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation; Better indicated by lower values)											
63 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	34	29	N/A	N/A	The mean irritability (risperidone) in the intervention groups was **0.52 standard deviations lower** (1.02 to 0.01 lower)
Lethargy/Social withdrawal (aripiprazole) (measured with: Aberrant Behaviour Checklist (ABC): Lethargy (Change Score); Better indicated by lower values)											
101 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[1,4] due to risk of bias, imprecision	49	52	N/A	N/A	The mean lethargy/social withdrawal (aripiprazole) in the intervention groups was **0.07 standard deviations lower** (0.46 lower to 0.32 higher)
Stereotypic behaviour (aripiprazole) (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic behaviour (Change Score); Better indicated by lower values)											
101 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[1,4] due to risk of bias, imprecision	49	52	N/A	N/A	The mean stereotypic behaviour (aripiprazole) in the intervention groups was **0.55 standard deviations lower** (0.95 to 0.15 lower)
Hyperactivity (aripiprazole) (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity (Change Score); Better indicated by lower values)											
101 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[1,4] due to risk of bias, imprecision	49	52	N/A	N/A	The mean hyperactivity (aripiprazole) in the intervention groups was **0.53 standard deviations**

											lower (0.93 to 0.14 lower)
Inappropriate Speech (aripiprazole) (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech (Change Score); Better indicated by lower values)											
100 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	48	52	N/A	N/A	The mean inappropriate speech (aripiprazole) in the intervention groups was **0.25 standard deviations lower** (0.65 lower to 0.14 higher)
Global state: Positive treatment response (assessed with: Dichotomous: Positive treatment response ('much improved/very improved' on CGI-improvement))											
64 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3] due to imprecision	5/34 (14.7%)	5/30 (16.7%)	**RR 1.13** (0.36 to 3.54)	**Study population**	
										147 per 1000	**19 more per 1000** (from 94 fewer to 374 more)
										Moderate	
										147 per 1000	**19 more per 1000** (from 94 fewer to 373 more)
Global severity (risperidone or aripiprazole) (measured with: Clinical Global Impression Scale (CGI-S): Severity (Endpoint or Change Scores); Better indicated by lower values)											
148 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	75	73	N/A	N/A	The mean global severity (risperidone or aripiprazole) in the intervention groups was **0.09 standard deviations lower** (0.41 lower to 0.24 higher)
Global severity (risperidone) (measured with: Clinical Global Impression Scale (CGI-S): Severity; Better indicated by lower values)											
63 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	34	29	N/A	N/A	The mean global severity (risperidone) in the intervention groups was **0.1 standard deviations higher** (0.39 lower to 0.6 higher)

Global severity (aripiprazole) (measured with: Clinical Global Impression Scale (CGI-S): Severity (Change Scores); Better indicated by lower values)										
85 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	41 44	N/A	N/A	The mean global severity (aripiprazole) in the intervention groups was **0.23 standard deviations lower** (0.65 lower to 0.2 higher)

[1] Blinding is unclear in MARCUS2009 as paper states 'double-blind' but gives no further detail with regards to who is blinded, i.e. participant, parent, investigator, intervention administrator, outcome assessor.
[2] Events<300
[3] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)
[4] N<400
[5] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Continued risperidone versus switch to placebo for behaviour that challenges as a direct outcome

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 Switch to placebo	**With Continued antipscyhotic**		**Risk with Switch to placebo**	**Risk difference with Continued antipscyhotic** (95% CI)
Relapse rate after discontinuation (assessed with: Number of participants showing >25% worsening in ABC-Irritability and rated as 'worse/very much worse' on CGI-I)											
56 (2 studies) 32-33 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	18/28 (64.3%)	5/28 (17.9%)	**RR 0.28** (0.12 to 0.64)	**Study population**	
										643 per 1000	**463 fewer per 1000** (from 231 fewer to 566 fewer)
										Moderate	
										646 per 1000	**465 fewer per 1000** (from 233 fewer to 568 fewer)
Time to relapse after discontinuation (in weeks) (measured with: Time to relapse (in weeks); Better indicated by lower values)											
24 (1 study) 32 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	12	12	N/A	N/A	The mean time to relapse after discontinuation (in weeks) in the intervention groups was **0.97 standard deviations**

										higher (0.11 to 1.82 higher)
Irritability (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation; Better indicated by lower values)										
24 (1 study) 32 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3] due to imprecision	12	12	N/A	N/A	The mean irritability in the intervention groups was **0.74 standard deviations lower** (1.58 lower to 0.09 higher)
Lethargy/Social withdrawal (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal; Better indicated by lower values)										
24 (1 study) 32 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3] due to imprecision	12	12	N/A	N/A	The mean lethargy/social withdrawal in the intervention groups was **0.58 standard deviations lower** (1.4 lower to 0.24 higher)
Stereotypic behaviour (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic Behaviour; Better indicated by lower values)										
24 (1 study) 32 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3] due to imprecision	12	12	N/A	N/A	The mean stereotypic behaviour in the intervention groups was **0.02 standard deviations lower** (0.82 lower to 0.78 higher)
Hyperactivity (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance; Better indicated by lower values)										
24 (1 study) 32 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3] due to imprecision	12	12	N/A	N/A	The mean hyperactivity in the intervention groups was **0.23 standard deviations lower** (1.03 lower to 0.58 higher)
Inappropriate speech (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech; Better indicated by lower values)										
24 (1 study) 32 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3] due to imprecision	12	12	N/A	N/A	The mean inappropriate speech in the intervention groups was **0 standard deviations higher**

										(0.8 lower to 0.8 higher)

[1] Events<300
[2] N<400
[3] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Risperidone versus haloperidol for behaviour that challenges as an indirect outcome

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 Risperidone versus haloperidol for behaviour that challenges as an indirect outcome**		**Risk with Control**	**Risk difference with Risperidone versus haloperidol for behaviour that challenges as an indirect outcome** (95% CI)
Behaviour that challenges (measured with: Aberrant Behaviour Checklist (ABC): Total; Better indicated by lower values)											
28 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	15	13	N/A	N/A	The mean behaviour that challenges in the intervention groups was **0.5 standard deviations lower** (1.25 lower to 0.26 higher)

[1] Paper states 'Double-blind' but gives no further detail with regards to who is blinded, i.e. participant, parent, investigator, intervention administrator, outcome assessor
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.11.6 Antivirals for behaviour that challenges as a direct outcome

Amantadine hydrochloride versus placebo for behaviour that challenges as a direct outcome

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 Placebo	**With Antivirals**		**Risk with Placebo**	**Risk difference with Antivirals** (95% CI)

Positive parent-rated treatment response (assessed with: >25% improvement on ABC-Irritability and/or hyperactivity)											
38 (1 study) 5 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	7/19 (36.8%)	9/19 (47.4%)	**RR 1.29** (0.6 to 2.74)	**Study population**	
										368 per 1000	**107 more per 1000** (from 147 fewer to 641 more)
										Moderate	
										368 per 1000	**107 more per 1000** (from 147 fewer to 640 more)
Positive investigator-rated treatment response (assessed with: 'much improved/very improved' on CGI-improvement)											
39 (1 study) 5 weeks	serious[2]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	5/20 (25%)	10/19 (52.6%)	**RR 2.11** (0.88 to 5.03)	**Study population**	
										250 per 1000	**277 more per 1000** (from 30 fewer to 1000 more)
										Moderate	
										250 per 1000	**277 more per 1000** (from 30 fewer to 1000 more)

[1] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)
[2] Blinding of outcome assessor is not clear and trial funded by pharmaceutical company

1.11.7 Cognitive enhancers for behaviour that challenges as a direct outcome

Piracetam and risperidone versus placebo and risperidone for behaviour that challenges as a direct outcome

Quality assessment							Summary of Findings			
Participants (studies)	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**		**Relative effect**	**Anticipated absolute effects**
							With	**With Combined piracetam**		**Risk** **Risk difference with Combined**

Follow up							**Control**	**and risperidone versus combined placebo and risperidone for behaviour that challenges as a direct outcome**	(95% CI)	**with Control**	**piracetam and risperidone versus combined placebo and risperidone for behaviour that challenges as a direct outcome** (95% CI)
Behaviour that challenges (measured with: Aberrant Behaviour Checklist (ABC): Total (Change Score); Better indicated by lower values)											
40 (1 study) 10 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	20	20	N/A	N/A	The mean behaviour that challenges in the intervention groups was **1.93 standard deviations lower** (2.69 to 1.16 lower)
[1] N<400											

1.11.8 Methylxanthines for behaviour that challenges as a direct outcome

Pentoxifylline and risperidone versus placebo and risperidone for behaviour that challenges as a direct outcome

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 Combined methylxanthine and antipsychotic versus combined antipsychotic and placebo for behaviour that challenges as a direct outcome**		**Risk with Control**	**Risk difference with Combined methylxanthine and antipsychotic versus combined antipsychotic and placebo for behaviour that challenges as a direct outcome** (95% CI)
Irritability (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation; Better indicated by lower values)											
40 (1 study) 10 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	20	20	N/A	N/A	The mean irritability in the intervention groups was **1.71 standard deviations lower**

											(2.44 to 0.97 lower)
Lethargy & Social Withdrawal (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal; Better indicated by lower values)											
40 (1 study) 10 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	20	20	N/A	N/A	The mean lethargy & social withdrawal in the intervention groups was **1.69 standard deviations lower** (2.42 to 0.96 lower)
Stereotypic Behaviour (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic Behaviour; Better indicated by lower values)											
40 (1 study) 10 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	20	20	N/A	N/A	The mean stereotypic behaviour in the intervention groups was **1.55 standard deviations lower** (2.27 to 0.83 lower)
Hyperactivity (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance; Better indicated by lower values)											
40 (1 study) 10 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	20	20	N/A	N/A	The mean hyperactivity in the intervention groups was **1.14 standard deviations lower** (1.81 to 0.47 lower)
Inappropriate Speech (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech; Better indicated by lower values)											
40 (1 study) 10 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	20	20	N/A	N/A	The mean inappropriate speech in the intervention groups was **2.1 standard deviations lower** (2.89 to 1.31 lower)

[1] N<400

1.11.9 Opioid antagonists for behaviour that challenges as a direct outcome

Naltrexone versus placebo for behaviour that challenges as a direct outcome

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 Placebo	With Opioid antagonists		Risk with Placebo / Risk difference with Opioid antagonists (95% CI)
Global positive treatment response for behaviour that challenges (assessed with: Dichotomous measure of 'much improved/very improved' on Clinical Global Impression-Improvement [CGI-I])										
41 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	7/18 (38.9%)	13/23 (56.5%)	**RR 1.45** (0.74 to 2.87)	**Study population**
										389 per 1000 / **175 more per 1000** (from 101 fewer to 727 more)
										Moderate
										389 per 1000 / **175 more per 1000** (from 101 fewer to 727 more)

[1] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)

1.11.10 Selective noradrenaline reuptake inhibitors (SNRIs) for behaviour that challenges as an indirect outcome

Atomoxetine versus placebo for behaviour that challenges as an indirect outcome

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 Selective noradrenaline reuptake		Risk with / Risk difference with Selective noradrenaline reuptake inhibitors

								inhibitors versus placebo for behaviour that challenges as an indirect outcome		**Control**	**versus placebo for behaviour that challenges as an indirect outcome** (95% CI)
Irritability (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation; Better indicated by lower values)											
89 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	46	43	N/A	N/A	The mean irritability in the intervention groups was **0.09 standard deviations lower** (0.51 lower to 0.32 higher)
Lethargy/Social withdrawal (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal; Better indicated by lower values)											
89 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	46	43	N/A	N/A	The mean lethargy/social withdrawal in the intervention groups was **0.05 standard deviations lower** (0.46 lower to 0.37 higher)
Stereotypic behaviour (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic Behaviour; Better indicated by lower values)											
89 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	46	43	N/A	N/A	The mean stereotypic behaviour in the intervention groups was **0 standard deviations higher** (0.42 lower to 0.42 higher)
Hyperactivity (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance; Better indicated by lower values)											
88 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	45	43	N/A	N/A	The mean hyperactivity in the intervention groups was **0.19 standard deviations lower** (0.61 lower to 0.22 higher)
Inappropriate speech (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech; Better indicated by lower values)											
89	no	no serious	no serious	very	undetected	⊕⊕⊖⊖	46	43	N/A	N/A	The mean inappropriate

(1 study) 8 weeks	serious risk of bias	inconsistency	indirectness	serious[1]		**LOW**[1] due to imprecision				speech in the intervention groups was **0.22 standard deviations lower** (0.64 lower to 0.19 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] N<400

1.12 BIOMEDICAL INTERVENTIONS AIMED AT BEHAVIOUR THAT CHALLENGES

1.12.1 Complementary therapies for behaviour that challenges as a direct or indirect outcome

Thai massage and sensory integration therapy versus sensory integration therapy only for behaviour that challenges as a direct outcome

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 Thai massage and sensory integration therapy versus sensory integration therapy only for behaviour that challenges as a direct outcome**		**Risk with Control**	**Risk difference with Thai massage and sensory integration therapy versus sensory integration therapy only for behaviour that challenges as a direct outcome** (95% CI)
Teacher-rated behaviour that challenges (measured with: Conners Teacher Rating Scales (CTRS): Conduct problem; Better indicated by lower values)											
60 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	30	30	N/A	N/A	The mean teacher-rated behaviour that challenges in the intervention groups was **0.22 standard deviations lower** (0.73 lower to 0.28 higher)
Teacher-rated behaviour that challenges (measured with: Conners Teacher Rating Scales (CTRS): Hyperactivity; Better indicated by lower values)											
60	no	no serious	no serious	serious[2]	undetected	⊕⊕⊕⊖	30	30	N/A	N/A	The mean teacher-rated

(1 study) 8 weeks	serious risk of bias	inconsistency	indirectness			**MODERATE**[2] due to imprecision				behaviour that challenges in the intervention groups was **0.56 standard deviations lower** (1.08 to 0.04 lower)
Teacher-rated behaviour that challenges (measured with: Conners Teacher Rating Scales (CTRS): Inattention-passivity; Better indicated by lower values)										
60 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	30	30	N/A	N/A	The mean teacher-rated behaviour that challenges in the intervention groups was **0.36 standard deviations lower** (0.87 lower to 0.15 higher)
Teacher-rated behaviour that challenges (measured with: Conners Teacher Rating Scales (CTRS): Hyperactivity index; Better indicated by lower values)										
60 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	30	30	N/A	N/A	The mean teacher-rated behaviour that challenges in the intervention groups was **0.4 standard deviations lower** (0.91 lower to 0.11 higher)
Parent-rated behaviour that challenges (measured with: Conners Parent Rating Scales (CPRS): Conduct problem; Better indicated by lower values)										
60 (1 study) 8 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	30	30	N/A	N/A	The mean parent-rated behaviour that challenges in the intervention groups was **0.1 standard deviations lower** (0.61 lower to 0.41 higher)
Parent-rated behaviour that challenges (measured with: Conners Parent Rating Scales (CPRS): Learning Problem; Better indicated by lower values)										
60 (1 study) 8 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	30	30	N/A	N/A	The mean parent-rated behaviour that challenges in the intervention groups was **0.21 standard deviations lower** (0.72 lower to 0.29 higher)

Parent-rated behaviour that challenges (measured with: Conners Parent Rating Scales (CPRS): Psychosomatic; Better indicated by lower values)											
60 (1 study) 8 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	30	30	N/A	N/A	The mean parent-rated behaviour that challenges in the intervention groups was **0.07 standard deviations higher** (0.44 lower to 0.57 higher)
Parent-rated behaviour that challenges (measured with: Conners Parent Rating Scales (CPRS): Impulsivity-hyperactivity; Better indicated by lower values)											
60 (1 study) 8 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	30	30	N/A	N/A	The mean parent-rated behaviour that challenges in the intervention groups was **0.5 standard deviations lower** (1.02 lower to 0.01 higher)
Parent-rated behaviour that challenges (measured with: Conners Parent Rating Scales (CPRS): Anxiety; Better indicated by lower values)											
60 (1 study) 8 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	30	30	N/A	N/A	The mean parent-rated behaviour that challenges in the intervention groups was **0.2 standard deviations lower** (0.71 lower to 0.3 higher)
Parent-rated behaviour that challenges (measured with: Conners Parent Rating Scales (CPRS): Hyperactivity; Better indicated by lower values)											
60 (1 study) 8 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	30	30	N/A	N/A	The mean parent-rated behaviour that challenges in the intervention groups was **0.24 standard deviations lower** (0.75 lower to 0.27 higher)
Parent-rated sleep-related problems (measured with: Sleep Diary (SD): Sleep behaviour; Better indicated by lower values)											
60 (1 study) 8 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,3] due to risk of bias,	30	30	N/A	N/A	The mean parent-rated sleep-related problems in the intervention groups was **0.53 standard deviations**

						imprecision				**lower** (1.04 to 0.01 lower)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] N<400
[3] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome measure parent-rated and parents were non-blind

Electro-acupuncture versus sham electro-acupuncture for behaviour that challenges as an indirect outcome

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 Acupuncture/Electro-acupuncture versus sham acupuncture/electro-acupuncture for behaviour that challenges as an indirect outcome**		**Risk with Control**	**Risk difference with Acupuncture/Electro-acupuncture versus sham acupuncture/electro-acupuncture for behaviour that challenges as an indirect outcome** (95% CI)
Irritability (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation; Better indicated by lower values)											
55 (1 study) 4 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	25	30	N/A	N/A	The mean irritability in the intervention groups was **0.18 standard deviations higher** (0.36 lower to 0.71 higher)
Lethargy/Social withdrawal (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal; Better indicated by lower values)											
55 (1 study) 4 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	25	30	N/A	N/A	The mean lethargy/social withdrawal in the intervention groups was **0.02 standard deviations lower** (0.56 lower to 0.51 higher)
Stereotypic behaviour (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic Behaviour; Better indicated by lower values)											
55 (1 study)	no serious	no serious inconsistency	no serious indirectness	very serious[1]	reporting bias strongly	⊕⊖⊖⊖ **VERY LOW**[1,2]	25	30	N/A	N/A	The mean stereotypic behaviour in the intervention

4 weeks	risk of bias				suspected [2]	due to imprecision, publication bias				groups was **0.05 standard deviations higher** (0.48 lower to 0.58 higher)
Hyperactivity (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance; Better indicated by lower values)										
55 (1 study) 4 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	25	30	N/A	N/A	The mean hyperactivity in the intervention groups was **0.01 standard deviations lower** (0.54 lower to 0.52 higher)
Inappropriate speech (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech; Better indicated by lower values)										
55 (1 study) 4 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	25	30	N/A	N/A	The mean inappropriate speech in the intervention groups was **0.14 standard deviations lower** (0.68 lower to 0.39 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] High risk of selective reporting bias as trial protocol for WONG2010B states that follow-up measurements will be taken but these are not reported

Electro-acupuncture and conventional educational programme versus conventional educational programme only for behaviour that challenges as an indirect outcome

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 Acupuncture/electro-acupuncture and conventional educational programme versus conventional educational programme only for behaviour that challenges as an indirect outcome**		**Risk with Control**	**Risk difference with Acupuncture/electro-acupuncture and conventional educational programme versus conventional educational programme only for behaviour that challenges as an indirect outcome** (95% CI)
Behaviour that challenges (measured with: Aberrant Behaviour Checklist (ABC): Total; Better indicated by lower values)											

36 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	18	18	N/A	N/A	The mean behaviour that challenges in the intervention groups was **0.3 standard deviations higher** (0.36 lower to 0.95 higher)
Irritability (measured with: Aberrant Behaviour Checklist (ABC): Irritability (Change Score); Better indicated by lower values)											
36 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	18	18	N/A	N/A	The mean irritability in the intervention groups was **0.42 standard deviations higher** (0.24 lower to 1.08 higher)
Lethargy/Social withdrawal (measured with: Aberrant Behaviour Checklist (ABC): Lethargy (Change Score); Better indicated by lower values)											
36 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	18	18	N/A	N/A	The mean lethargy/social withdrawal in the intervention groups was **0.23 standard deviations higher** (0.42 lower to 0.89 higher)
Stereotypic behaviour (measured with: Aberrant Behaviour Checklist (ABC): Stereotypy (Change Score); Better indicated by lower values)											
36 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	18	18	N/A	N/A	The mean stereotypic behaviour in the intervention groups was **0.29 standard deviations higher** (0.37 lower to 0.94 higher)
Hyperactivity (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity (Change Score); Better indicated by lower values)											
36 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	18	18	N/A	N/A	The mean hyperactivity in the intervention groups was **0.06 standard deviations lower** (0.72 lower to 0.59 higher)
Inappropriate speech (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech (Change Score); Better indicated by lower values)											
36	serious[1]	no serious	no serious	very	undetected	⊕⊖⊖⊖	18	18	N/A	N/A	The mean inappropriate speech

(1 study) 8 weeks		inconsistency	indirectness	serious[2]		**VERY LOW**[1,2] due to risk of bias, imprecision				in the intervention groups was **0.58 standard deviations higher** (0.09 lower to 1.25 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind and potential for care confounds as the conventional education programme differed for each participant which may introduce bias. There was also an unclear risk of detection bias as although all outcomes were measured by blinded assessors, some outcomes involved input from parents who were not blind to treatment allocation or confounding variables and systematic review from which data was extracted does not report which outcome measures relied on non-blind parental report
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.12.2 Hormones for behaviour that challenges as an indirect outcome

Secretin versus placebo for behaviour that challenges as an indirect outcome

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 Secretin versus placebo for behaviour that challenges as an indirect outcome**		**Risk with Control**	**Risk difference with Secretin versus placebo for behaviour that challenges as an indirect outcome** (95% CI)
Behaviour that challenges (Parent-rated) (measured with: Aberrant Behaviour Checklist (ABC): Total (change score); Better indicated by lower values)											
77 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	29	48	N/A	N/A	The mean behaviour that challenges (parent-rated) in the intervention groups was **0.13 standard deviations lower** (0.59 lower to 0.33 higher)
Behaviour that challenges (Teacher-rated) (measured with: Aberrant Behaviour Checklist (ABC): Total (change score); Better indicated by lower values)											
65 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	26	39	N/A	N/A	The mean behaviour that challenges (teacher-rated) in the intervention groups was **0.51 standard deviations higher**

										(0 to 1.01 higher)
Irritability (Parent-rated) (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation (endpoint and change scores); Better indicated by lower values)										
133 (2 studies) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	57	76	N/A	N/A	The mean irritability (parent-rated) in the intervention groups was **0.11 standard deviations lower** (0.45 lower to 0.24 higher)
Irritability (Teacher-rated) (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation (change scores); Better indicated by lower values)										
65 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	26	39	N/A	N/A	The mean irritability (teacher-rated) in the intervention groups was **0.2 standard deviations higher** (0.3 lower to 0.69 higher)
Lethargy (Parent-rated) (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal (endpoint and change scores); Better indicated by lower values)										
133 (2 studies) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	57	76	N/A	N/A	The mean lethargy (parent-rated) in the intervention groups was **0.11 standard deviations higher** (0.24 lower to 0.46 higher)
Lethargy (Teacher-rated porcine secretin) (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal (change scores); Better indicated by lower values)										
48 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	26	22	N/A	N/A	The mean lethargy (teacher-rated porcine secretin) in the intervention groups was **0.74 standard deviations higher** (0.15 to 1.33 higher)
Lethargy (Teacher-rated synthetic porcine secretin) (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal (change scores); Better indicated by lower values)										

43 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	26	17	N/A	N/A	The mean lethargy (teacher-rated synthetic porcine secretin) in the intervention groups was **0.05 standard deviations higher** (0.56 lower to 0.67 higher)
Stereotypic behaviour (Parent-rated) (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic Behaviour (endpoint and change scores); Better indicated by lower values)											
133 (2 studies) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	57	76	N/A	N/A	The mean stereotypic behaviour (parent-rated) in the intervention groups was **0.1 standard deviations higher** (0.25 lower to 0.45 higher)
Stereotypic behaviour (Teacher-rated) (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic Behaviour (change scores); Better indicated by lower values)											
65 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	26	39	N/A	N/A	The mean stereotypic behaviour (teacher-rated) in the intervention groups was **0.33 standard deviations higher** (0.17 lower to 0.82 higher)
Hyperactivity (Parent-rated) (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance (endpoint and change scores); Better indicated by lower values)											
133 (2 studies) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	57	76	N/A	N/A	The mean hyperactivity (parent-rated) in the intervention groups was **0.01 standard deviations lower** (0.36 lower to 0.34 higher)
Hyperactivity (Teacher-rated) (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance (change scores); Better indicated by lower values)											
65 (1 study) 4 weeks	no serious risk of	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to	26	39	N/A	N/A	The mean hyperactivity (teacher-rated) in the intervention groups was

	bias					imprecision				**0.53 standard deviations higher** (0.03 to 1.04 higher)
Inappropriate speech (Parent-rated) (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech (endpoint and change scores); Better indicated by lower values)										
131 (2 studies) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	55	76	N/A	N/A	The mean inappropriate speech (parent-rated) in the intervention groups was **0.39 standard deviations lower** (0.75 to 0.04 lower)
Inappropriate speech (Teacher-rated) (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech (change scores); Better indicated by lower values)										
65 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	26	39	N/A	N/A	The mean inappropriate speech (teacher-rated) in the intervention groups was **0.28 standard deviations higher** (0.22 lower to 0.78 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] N<400

1.12.3 Medical procedures for behaviour that challenges as a direct or indirect outcome

HBOT versus attention-placebo for behaviour that challenges as a direct or indirect outcome

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 Attention-placebo	**With Hyperbaric oxygen treatment**		**Risk with Attention-placebo**	**Risk difference with Hyperbaric oxygen treatment (HBOT)** (95% CI)

							control	(HBOT)		control	
Behaviour that challenges (measured with: Aberrant Behaviour Checklist (ABC): Total or Behavioural observation: Challenging behaviours (change score); Better indicated by lower values)											
90 (2 studies) 4-15 weeks	no serious risk of bias	serious[1]	no serious indirectness	very serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to inconsistency, imprecision, publication bias	42	48	N/A	N/A	The mean behaviour that challenges in the intervention groups was **0.17 standard deviations lower** (0.59 lower to 0.24 higher)
Behaviour that challenges (direct outcome) (measured with: Aberrant Behaviour Checklist (ABC): Total; Better indicated by lower values)											
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	26	30	N/A	N/A	The mean behaviour that challenges (direct outcome) in the intervention groups was **0.04 standard deviations higher** (0.48 lower to 0.57 higher)
Behaviour that challenges (indirect outcome) (measured with: Behavioural observation: Challenging behaviours (change score); Better indicated by lower values)											
34 (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	16	18	N/A	N/A	The mean behaviour that challenges (indirect outcome) in the intervention groups was **0.54 standard deviations lower** (1.23 lower to 0.15 higher)

Irritability (measured with: Aberrant Behaviour Checklist (ABC): Irritability; Better indicated by lower values)											
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	26	30	N/A	N/A	The mean irritability in the intervention groups was **0.11 standard deviations lower** (0.64 lower to 0.41 higher)
Lethargy/Social withdrawal (measured with: Aberrant Behaviour Checklist (ABC): Lethargy; Better indicated by lower values)											
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	26	30	N/A	N/A	The mean lethargy/social withdrawal in the intervention groups was **0.06 standard deviations higher** (0.46 lower to 0.59 higher)
Stereotypy (measured with: Aberrant Behaviour Checklist (ABC): Stereotypy; Better indicated by lower values)											
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	26	30	N/A	N/A	The mean stereotypy in the intervention groups was **0.17 standard deviations higher** (0.36 lower to 0.7 higher)
Hyperactivity (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity or Behavioural observation: Hyperactivity (change score); Better indicated by lower values)											
90 (2 studies) 4-15 weeks	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	42	48	N/A	N/A	The mean hyperactivity in the intervention groups was **0.06 standard deviations higher**

											(0.36 lower to 0.47 higher)
Hyperactivity (direct outcome) (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity; Better indicated by lower values)											
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	26	30	N/A	N/A	The mean hyperactivity (direct outcome) in the intervention groups was **0.12 standard deviations higher** (0.41 lower to 0.64 higher)
Hyperactivity (indirect outcome) (measured with: Behavioural observation: Hyperactivity (change score); Better indicated by lower values)											
34 (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	16	18	N/A	N/A	The mean hyperactivity (indirect outcome) in the intervention groups was **0.04 standard deviations lower** (0.72 lower to 0.63 higher)
Inappropriate speech (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech; Better indicated by lower values)											
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	26	30	N/A	N/A	The mean inappropriate speech in the intervention groups was **0.24 standard deviations lower** (0.77 lower to 0.28 higher)

[1] Evidence for moderate inconsistency with an I-squared value of 43% but this is not statistically significant (p=0.19)
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] High risk of selective reporting bias for GRANPEESHEH2010 as data cannot be extracted for the Aberrant Behavior Checklist (ABC)
[4] N<400

Long-term chelation (7-rounds of DMSA therapy) versus short-term chelation (1-round of DMSA therapy and 6-rounds of placebo) for behaviour that challenges as an indirect outcome

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 Short-term chelation (1-round of DMSA therapy and 6-rounds of placebo)	**With Long-term chelation (7-rounds of Dimercaptosuccinic Acid [DMSA] therapy)**		**Risk with Short-term chelation (1-round of DMSA therapy and 6-rounds of placebo)**	**Risk difference with Long-term chelation (7-rounds of Dimercaptosuccinic Acid [DMSA] therapy)** (95% CI)
Maladaptive Behaviours Composite (measured with: Pervasive Development Disorder Behavior Inventory (PDDBI): Maladaptive Behaviours Composite; Better indicated by lower values)											
40 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	25	N/A	N/A	The mean maladaptive behaviours composite in the intervention groups was **0.17 standard deviations higher** (0.47 lower to 0.81 higher)
Arousal Regulation Problems (measured with: Pervasive Development Disorder Behavior Inventory (PDDBI): Arousal Regulation Problems; Better indicated by lower values)											
40 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	25	N/A	N/A	The mean arousal regulation problems in the intervention groups was **0.2 standard deviations higher** (0.44 lower to 0.85

										higher)
Aggressiveness (measured with: Pervasive Development Disorder Behavior Inventory (PDDBI): Aggressiveness; Better indicated by lower values)										
40 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15 25	N/A	N/A	The mean aggressiveness in the intervention groups was **0.2 standard deviations higher** (0.44 lower to 0.84 higher)
[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)										

1.12.4 Nutritional interventions for behaviour that challenges as a direct or indirect outcome

Omega-3 fatty acids versus placebo for behaviour that challenges as a direct outcome

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 Placebo	**With Omega-3 fatty acids**		**Risk with Placebo**	**Risk difference with Omega-3 fatty acids** (95% CI)
Irritability (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation; Better indicated by lower values)											
24 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	12	12	N/A	N/A	The mean irritability in the intervention groups was **0.09 standard deviations lower** (0.89 lower to 0.71 higher)
Lethargy/Social withdrawal (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal; Better indicated by lower values)											
24 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to	12	12	N/A	N/A	The mean lethargy/social withdrawal in the intervention groups was

						imprecision				**0.28 standard deviations lower** (1.09 lower to 0.52 higher)
Stereotypic behaviour (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic Behaviour; Better indicated by lower values)										
24 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	12	12	N/A	N/A	The mean stereotypic behaviour in the intervention groups was **0.81 standard deviations lower** (1.65 lower to 0.03 higher)
Hyperactivity (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance; Better indicated by lower values)										
24 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	12	12	N/A	N/A	The mean hyperactivity in the intervention groups was **0.42 standard deviations lower** (1.23 lower to 0.39 higher)
Inappropriate speech (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech; Better indicated by lower values)										
24 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	12	12	N/A	N/A	The mean inappropriate speech in the intervention groups was **0.68 standard deviations lower** (1.51 lower to 0.14 higher)
Externalizing (measured with: Behavior Assessment System for Children (BASC): Externalizing; Better indicated by lower values)										
24 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	12	12	N/A	N/A	The mean externalizing in the intervention groups was **0.44 standard deviations lower** (1.25 lower to 0.37 higher)
Behavioural symptoms (measured with: Behavior Assessment System for Children (BASC): Behavioral symptoms; Better indicated by lower values)										
23 (1 study)	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1]	11	12	N/A	N/A	The mean behavioural symptoms in the intervention

12 weeks						due to imprecision				groups was **0.24 standard deviations lower** (1.06 lower to 0.58 higher)
Hyperactivity (measured with: Behavior Assessment System for Children (BASC): Hyperactivity; Better indicated by lower values)										
24 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	12	12	N/A	N/A	The mean hyperactivity in the intervention groups was **0.19 standard deviations lower** (0.99 lower to 0.61 higher)
[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefot or harm (SMD -0.5/0.5)										

Omega-3 fatty acids versus healthy diet control for behaviour that challenges as a direct outcome

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 Healthy diet control	**With Omega-3 fatty acids**		**Risk with Healthy diet control**	**Risk difference with Omega-3 fatty acids** (95% CI)
Total problem score (measured with: Child Behavior Checklist 1.5 - 5 (CBCL/1.5-5): Total problem score; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A	The mean total problem score in the intervention groups was **0.17 standard deviations lower** (0.99 lower to 0.66 higher)
Externalizing (measured with: Child Behavior Checklist 1.5 - 5 (CBCL/1.5-5): Externalizing; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A	The mean externalizing in the intervention groups was **0.1 standard deviations lower** (0.92 lower to 0.73 higher)

Emotional regulation (measured with: Child Behavior Checklist 1.5 - 5 (CBCL/1.5-5): Emotional regulation; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A	The mean emotional regulation in the intervention groups was **0.09 standard deviations lower** (0.92 lower to 0.73 higher)
Withdrawn (measured with: Child Behavior Checklist 1.5 - 5 (CBCL/1.5-5): Withdrawn; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A	The mean withdrawn in the intervention groups was **0.81 standard deviations lower** (1.67 lower to 0.05 higher)
Attention problems (measured with: Child Behavior Checklist 1.5 - 5 (CBCL/1.5-5): Attention problems; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A	The mean attention problems in the intervention groups was **0.53 standard deviations lower** (1.37 lower to 0.31 higher)
Aggressive behaviours (measured with: Child Behavior Checklist 1.5 - 5 (CBCL/1.5-5): Aggressive behaviours; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A	The mean aggressive behaviours in the intervention groups was **0 standard deviations higher** (0.83 lower to 0.82 higher)
Oppositional Defiant Disorder (ODD) symptoms (measured with: Child Behavior Checklist 1.5 - 5 (CBCL/1.5-5): ODD; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A	The mean oppositional defiant disorder (odd) symptoms in the intervention groups was **0.04 standard deviations lower** (0.87 lower to 0.78 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as the outcome assessor for this outcome

measure was not blinded.
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Ginkgo biloba and risperidone versus placebo and risperidone for behaviour that challenges as a direct outcome

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 Combined placebo and risperidone	**With Combined ginkgo biloba and risperidone**		**Risk with Combined placebo and risperidone**	**Risk difference with Combined ginkgo biloba and risperidone** (95% CI)
Irritability (measured with: Aberrant Behaviour Checklist (ABC): Irritability & Agitation; Better indicated by lower values)											
47 (1 study) 10 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	24	23	N/A	N/A	The mean irritability in the intervention groups was **0.1 standard deviations higher** (0.47 lower to 0.67 higher)
Lethargy/Social Withdrawal (measured with: Aberrant Behaviour Checklist (ABC): Lethargy & Social Withdrawal; Better indicated by lower values)											
47 (1 study) 10 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	24	23	N/A	N/A	The mean lethargy/social withdrawal in the intervention groups was **0.08 standard deviations lower** (0.65 lower to 0.49 higher)
Stereotypic Behaviour (measured with: Aberrant Behaviour Checklist (ABC): Stereotypic Behaviour; Better indicated by lower values)											
47 (1 study) 10 weeks	no serious risk of	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to	24	23	N/A	N/A	The mean stereotypic behaviour in the intervention groups

	bias					imprecision				was **0.02 standard deviations lower** (0.59 lower to 0.55 higher)
Hyperactivity (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance; Better indicated by lower values)										
47 (1 study) 10 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	24	23	N/A	N/A	The mean hyperactivity in the intervention groups was **0.22 standard deviations higher** (0.35 lower to 0.8 higher)
Inappropriate Speech (measured with: Aberrant Behaviour Checklist (ABC): Inappropriate Speech; Better indicated by lower values)										
47 (1 study) 10 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	24	23	N/A	N/A	The mean inappropriate speech in the intervention groups was **0.21 standard deviations lower** (0.79 lower to 0.36 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Dimethylglycine supplement versus placebo for behaviour that challenges as a direct outcome

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 Placebo	**With Dimethylglycine**		**Risk with Placebo**	**Risk difference with Dimethylglycine** (95% CI)
Positive treatment response (assessed with: Parental report)											

38 (1 study) 4 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	10/19 (52.6%)	11/19 (57.9%)	**RR 1.1** (0.62 to 1.95)	**Study population**	
										526 per 1000	**53 more per 1000** (from 200 fewer to 500 more)
										Moderate	
										526 per 1000	**53 more per 1000** (from 200 fewer to 500 more)

[1] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)
[2] High risk of selective reporting bias as data could not be extracted for the Aberrant Behavior Checklist (Irritability, Lethargy/Social Withdrawal, Stereotypic Behavior, Hyperactivity and Inappropriate Speech subscales) or the Maladaptive Behavior Domain of the Vineland Adaptive Behavior Scale and potential conflict of interest as trial funded by manufacturer of supplement

Multivitamin/mineral supplement versus placebo for behaviour that challenges as an indirect outcome

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 Placebo	**With Multivitamin and mineral supplement**		**Risk with Placebo**	**Risk difference with Multivitamin and mineral supplement** (95% CI)
Hyperactivity improvement (measured with: Parent Global Impressions-Revised (PGI-R): Hyperactivity improvement; Better indicated by lower values)											
104 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	51	53	N/A	N/A	The mean hyperactivity improvement in the intervention groups was **0.6 standard deviations higher** (0.2 to 0.99 higher)
Tantrumming improvement (measured with: Parent Global Impressions-Revised (PGI-R): Tantrumming improvement; Better indicated by lower values)											
104 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	51	53	N/A	N/A	The mean tantrumming improvement in the intervention groups was **0.52 standard deviations higher**

									(0.13 to 0.91 higher)

[1] N<400

Immunoglobulin (dosages combined) versus placebo for behaviour that challenges as an indirect outcome

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 Placebo	**With Immunoglobulin (dosages combined)**		**Risk with Placebo**	**Risk difference with Immunoglobulin (dosages combined)** (95% CI)
Positive clinician-rated treatment response (assessed with: Dichotomous measure of 'much improved/very improved' on Clinical Global Impression-Improvement (CGI-I))											
111 (1 study) 12 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	11/28 (39.3%)	17/83 (20.5%)	**RR 0.52** (0.28 to 0.97)	**Study population**	
										393 per 1000	**189 fewer per 1000** (from 12 fewer to 283 fewer)
										Moderate	
										393 per 1000	**189 fewer per 1000** (from 12 fewer to 283 fewer)
Positive parent-rated treatment response (assessed with: Dichotomous measure of 'much improved/very improved' on Parent Global Impression-Improvement (PGI-I))											
112 (1 study) 12 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	16/29 (55.2%)	25/83 (30.1%)	**RR 0.55** (0.34 to 0.87)	**Study population**	
										552 per 1000	**248 fewer per 1000** (from 72 fewer to 364 fewer)
										Moderate	
										552 per 1000	**248 fewer per 1000** (from 72 fewer to 364

									fewer)

[1] Events<300
[2] High risk of selective reporting bias as continuous data could not be extracted for the CGI-I or PGI-I scale

1.12.5 Sensory interventions for behaviour that challenges as an indirect outcome

Auditory integration training versus attention-placebo (structured listening) for behaviour that challenges as an indirect outcome

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 Attention-placebo (structured listening) control	**With Auditory integration training**		**Risk with Attention-placebo (structured listening) control**	**Risk difference with Auditory integration training** (95% CI)
Parent-rated behaviour that challenges (measured with: Developmental Behaviour Checklist (DBC): Total; Better indicated by lower values)											
80 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean parent-rated behaviour that challenges in the intervention groups was **0.06 standard deviations higher** (0.38 lower to 0.5 higher)
Parent-rated behaviour that challenges (measured with: Developmental Behaviour Checklist (DBC): Total; Better indicated by lower values)											
80 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean parent-rated behaviour that challenges in the intervention groups was **0.2 standard deviations higher**

											(0.24 lower to 0.64 higher)
Parent-rated behaviour that challenges (measured with: Developmental Behaviour Checklist (DBC): Total; Better indicated by lower values)											
80 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean parent-rated behaviour that challenges in the intervention groups was **0.26 standard deviations higher** (0.18 lower to 0.7 higher)
Parent-rated behaviour that challenges (measured with: Developmental Behaviour Checklist (DBC): Total; Better indicated by lower values)											
80 (1 study) 56 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean parent-rated behaviour that challenges in the intervention groups was **0.24 standard deviations higher** (0.2 lower to 0.68 higher)
Teacher-rated behaviour that challenges (measured with: Developmental Behaviour Checklist (DBC): Total; Better indicated by lower values)											
80 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean teacher-rated behaviour that challenges in the intervention groups was **0.16 standard deviations lower** (0.6 lower to 0.28 higher)
Teacher-rated behaviour that challenges (measured with: Developmental Behaviour Checklist (DBC): Total; Better indicated by lower values)											
80	no	no serious	no serious	very	undetected	⊕⊕⊖⊖	40	40	N/A	N/A	The mean teacher-

(1 study) 13 weeks	serious risk of bias	inconsistency	indirectness	serious[1]		**LOW**[1] due to imprecision					rated behaviour that challenges in the intervention groups was **0.15 standard deviations lower** (0.59 lower to 0.29 higher)
Teacher-rated behaviour that challenges (measured with: Developmental Behaviour Checklist (DBC): Total; Better indicated by lower values)											
80 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	40	40	N/A	N/A	The mean teacher-rated behaviour that challenges in the intervention groups was **0.04 standard deviations lower** (0.48 lower to 0.39 higher)
Teacher-rated behaviour that challenges (measured with: Developmental Behaviour Checklist (DBC): Total; Better indicated by lower values)											
80 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean teacher-rated behaviour that challenges in the intervention groups was **0.09 standard deviations higher** (0.35 lower to 0.53 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] N<400

1.13 PSYCHOSOCIAL INTERVENTIONS AIMED AT ADAPTIVE BEHAVIOUR

1.13.1 Behavioural interventions for adaptive behaviour as a direct or indirect outcome

EIBI or EBI (ESDM or P-ESDM) versus treatment-as-usual for adaptive behaviour as a direct or indirect outcome

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 ESDM or P-ESDM versus treatment-as-usual for adaptive behaviour as a direct or indirect outcome		Risk with Control	Risk difference with ESDM or P-ESDM versus treatment-as-usual for adaptive behaviour as a direct or indirect outcome (95% CI)
Adaptive behaviour (measured with: Vineland Adaptive Behaviour Scale (VABS/VABS II): Adaptive behaviour composite score; Better indicated by lower values)											
143 (2 studies) 12-104 weeks	serious[1]	very serious[2]	no serious indirectness	serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, inconsistency, imprecision	70	73	N/A	N/A	The mean adaptive behaviour in the intervention groups was **0.03 standard deviations higher** (0.31 lower to 0.36 higher)
Daily living skills (measured with: Vineland Adaptive Behaviour Scale (VABS/VABS II): Daily living skills; Better indicated by lower values)											
143 (2 studies) 12-104 weeks	serious[1]	very serious[2]	no serious indirectness	serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, inconsistency, imprecision	70	73	N/A	N/A	The mean daily living skills in the intervention groups was **0.1 standard deviations higher** (0.23 lower to 0.43 higher)
Socialization (measured with: Vineland Adaptive Behaviour Scale (VABS/VABS II): Socialization; Better indicated by lower values)											
143	serious[1]	very serious[2]	no serious	serious[3]	undetected	⊕⊖⊖⊖	70	73	N/A	N/A	The mean socialization in

(2 studies) 12-104 weeks			indirectness			**VERY LOW**[1,2,3] due to risk of bias, inconsistency, imprecision				the intervention groups was **0.08 standard deviations higher** (0.25 lower to 0.41 higher)
Communication (measured with: Vineland Adaptive Behaviour Scale (VABS/VABS II): Communication; Better indicated by lower values)										
143 (2 studies) 12-104 weeks	serious[1]	very serious[2]	no serious indirectness	serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, inconsistency, imprecision	70	73	N/A	N/A	The mean communication in the intervention groups was **0.11 standard deviations higher** (0.23 lower to 0.44 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind and high risk of detection bias as the outcome measure was based on interview with (non-blind) parent rather than direct observation
[2] I-squared value indicates substantial to considerable heterogeneity
[3] N<400

EIBI versus parent training for adaptive behaviour as a direct outcome

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 EIBI versus parent training for adaptive behaviour as a direct outcome**		**Risk with Control**	**Risk difference with EIBI versus parent training for adaptive behaviour as a direct outcome** (95% CI)
Adaptive behaviour (measured with: Vineland Adaptive Behaviour Scale (VABS): Total; Better indicated by lower values)											
28 (1 study) 260 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	15	N/A	N/A	The mean adaptive behaviour in the intervention groups was **0.11 standard deviations higher** (0.64 lower to 0.85 higher)

Daily living skills (measured with: Vineland Adaptive Behaviour Scale (VABS): Daily Living Skills; Better indicated by lower values)										
28 (1 study) 260 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	15	N/A	N/A	The mean daily living skills in the intervention groups was **0.03 standard deviations lower** (0.77 lower to 0.71 higher)
Socialization (measured with: Vineland Adaptive Behaviour Scale (VABS): Socialization; Better indicated by lower values)										
28 (1 study) 260 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	15	N/A	N/A	The mean socialization in the intervention groups was **0.12 standard deviations lower** (0.86 lower to 0.63 higher)
Communication (measured with: Vineland Adaptive Behaviour Scale (VABS): Communication; Better indicated by lower values)										
28 (1 study) 260 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	15	N/A	N/A	The mean communication in the intervention groups was **0.28 standard deviations higher** (0.47 lower to 1.02 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind and risk of detection bias is unclear/unknown as although outcome assessors were blinded the outcome measure was based on interview with (non-blind) parent rather than direct observation
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Home-based EBI versus centre-based EBI for adaptive behaviour as a direct outcome

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 Home-based versus Centre-based EBI for adaptive behaviour as a direct outcome**		**Risk with Control**	**Risk difference with Home-based versus Centre-based EBI for adaptive behaviour as a direct outcome** (95% CI)
Socialization (measured with: Vineland Adaptive Behaviour Scale (VABS): Socialization; Better indicated by lower values)											

56 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	29	27	N/A	N/A	The mean socialization in the intervention groups was **0.63 standard deviations lower** (1.17 to 0.09 lower)
Communication (measured with: Vineland Adaptive Behaviour Scale (VABS): Communication; Better indicated by lower values)											
55 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	29	26	N/A	N/A	The mean communication in the intervention groups was **0.46 standard deviations lower** (1 lower to 0.07 higher)
Adaptive functioning and psychopathology (measured with: Developmental Behaviour Checklist (DBC): Total; Better indicated by lower values)											
44 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	22	22	N/A	N/A	The mean adaptive functioning and psychopathology in the intervention groups was **0.11 standard deviations lower** (0.7 lower to 0.48 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as, despite blinding outcome assessors, the outcome measure relies on interview with parent and parents were non-blind to group assignment and other potentially confounding factors and were also part of the intervention so problems with self-assessment
[2] N<400
[3] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.13.2 Cognitive-behavioural interventions for adaptive behaviour as an indirect outcome

CBT versus waitlist for adaptive behaviour as an indirect outcome

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 CBT for anxiety versus waitlist control for adaptive behaviour as an**		**Risk with Control**	**Risk difference with CBT for anxiety versus waitlist control for adaptive behaviour as an indirect**

							indirect outcome			outcome (95% CI)	
Adaptive behaviour (self-care) (measured with: Vineland Adaptive Behaviour Scale (VABS): Daily Living Skills; Better indicated by lower values)											
40 (1 study) 16 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	20	20	N/A	N/A	The mean adaptive behaviour (slef-care) in the intervention groups was **0.63 standard deviations higher** (0.01 lower to 1.26 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear/unknown as outcome measure based on interview with non-blind parent rather than direct behavioural observation
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.13.3 Parent training for adaptive behaviour as a direct or indirect outcome

Parent training versus treatment-as-usual for adaptive behaviour as a direct or indirect outcome

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 Parent training versus treatment as usual for adaptive behaviour as a direct or indirect outcome**		**Risk with Control**	**Risk difference with Parent training versus treatment as usual for adaptive behaviour as a direct or indirect outcome** (95% CI)
Functional emotional development (clinician-rated) (measured with: Functional Emotional Assessment Scale (FEAS): Total; Better indicated by lower values)											
32 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	16	16	N/A	N/A	The mean functional emotional development (clinician-rated) in the intervention groups was **0.25 standard deviations lower** (0.95 lower to 0.45 higher)
Functional emotional development (parent-rated) (measured with: Functional Emotional Developmental Questionnaires (FEDQ): Total; Better indicated by lower values)											

32 (1 study) 13 weeks	serious[2]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	16	16	N/A	N/A	The mean functional emotional development (parent-rated) in the intervention groups was **0.2 standard deviations lower** (0.9 lower to 0.49 higher)
Daily living skills (PEBM) (measured with: Vineland Adaptive Behaviour Scale (VABS): Daily Living Skills; Better indicated by lower values)											
70 (1 study) 46 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	35	35	N/A	N/A	The mean daily living skills (pebm) in the intervention groups was **0.46 standard deviations higher** (0.01 lower to 0.94 higher)
Daily living skills (PEC) (measured with: Vineland Adaptive Behaviour Scale (VABS): Daily Living Skills; Better indicated by lower values)											
68 (1 study) 46 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	35	33	N/A	N/A	The mean daily living skills (pec) in the intervention groups was **0.14 standard deviations lower** (0.61 lower to 0.34 higher)
Socialization (PEBM) (measured with: Vineland Adaptive Behaviour Scale (VABS): Socialization; Better indicated by lower values)											
70 (1 study) 46 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	35	35	N/A	N/A	The mean socialization (pebm) in the intervention groups was **0.35 standard deviations higher** (0.12 lower to 0.83 higher)
Socialization (PEC) (measured with: Vineland Adaptive Behaviour Scale (VABS): Socialization; Better indicated by lower values)											
68 (1 study) 46 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	35	33	N/A	N/A	The mean socialization (pec) in the intervention groups was **0.26 standard deviations lower** (0.74 lower to 0.21 higher)

Communication (PEBM) (measured with: Vineland Adaptive Behaviour Scale (VABS): Communication; Better indicated by lower values)										
70 (1 study) 46 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	35 35	N/A	N/A	The mean communication (pebm) in the intervention groups was **0.1 standard deviations higher** (0.37 lower to 0.57 higher)
Communication (PEC) (measured with: Vineland Adaptive Behaviour Scale (VABS): Communication; Better indicated by lower values)										
68 (1 study) 46 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[3,4] due to risk of bias, imprecision	35 33	N/A	N/A	The mean communication (pec) in the intervention groups was **0.56 standard deviations lower** (1.04 to 0.07 lower)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as parent-rated and parents were non-blind and involved in the intervention so problems with self-assessment. There was also no independent reliability and validity data for the Thai-version of this outcome measure which was used in the study.
[3] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear/unknown as although the outcome assessor was a blinded clinician the measure is based on parental interview and simultaneous child observation and parents non-blind and involved in intervention
[4] N<400

Combined parent training and early intervention centre programme versus early intervention centre programme only for adaptive behaviour as an indirect outcome

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 Combined parent training and early intervention centre programme versus early intervention centre programme only for adaptive behaviour as a direct outcome**		**Risk with Control**	**Risk difference with Combined parent training and early intervention centre programme versus early intervention centre programme only for adaptive behaviour as a direct outcome** (95% CI)

Parent-reported adaptive behaviour (mixed ASD & DD sample) (measured with: Vineland Adaptive Behaviour Scale (VABS): Total; Better indicated by lower values)										
58 (1 study) 40 weeks	serious[1]	no serious inconsistency	serious[2]	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, indirectness, imprecision	28	30	N/A	N/A	The mean parent-reported adaptive behaviour (mixed asd & dd sample) in the intervention groups was **0.25 standard deviations higher** (0.27 lower to 0.77 higher)
Parent-reported adaptive behaviour (mixed ASD & DD sample) (measured with: Vineland Adaptive Behaviour Scale (VABS): Total; Better indicated by lower values)										
51 (1 study) 108 weeks	serious[1]	no serious inconsistency	serious[2]	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, indirectness, imprecision	23	28	N/A	N/A	The mean parent-reported adaptive behaviour (mixed asd & dd sample) in the intervention groups was **0.31 standard deviations higher** (0.24 lower to 0.87 higher)
Clinician-rated adaptive behaviour (mixed ASD & DD sample) (measured with: Bayley Behavior Rating Scale (BRS): Total; Better indicated by lower values)										
57 (1 study) 40 weeks	no serious risk of bias	no serious inconsistency	serious[2]	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to indirectness, imprecision	28	29	N/A	N/A	The mean clinician-rated adaptive behaviour (mixed asd & dd sample) in the intervention groups was **0.4 standard deviations higher** (0.12 lower to 0.93 higher)
Clinician-rated adaptive behaviour (mixed ASD & DD sample) (measured with: Bayley Behavior Rating Scale (BRS): Total; Better indicated by lower values)										
47 (1 study) 108 weeks	no serious risk of bias	no serious inconsistency	serious[2]	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[2,4] due to indirectness, imprecision	23	24	N/A	N/A	The mean clinician-rated adaptive behaviour (mixed asd & dd sample) in the intervention groups was **0.62 standard deviations higher** (0.04 to 1.21 higher)

[1] High risk of performance and response bias as intervention administrator and participants were non-blind, and risk of detection bias was unclear/unknown as, although the interviewer was a blinded research assistant, the outcome measure was based on non-blind parent report and parents were involved in the intervention
[2] Population was indirect (as the sample included participants with developmental delay or language delay without autism)
[3] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[4] N<400

Parent and day-care staff training versus standard day-care for adaptive behaviour as an indirect outcome

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 Parent and day-care staff training versus standard day-care for adaptive behaviour as an indirect outcome**		**Risk with Control**	**Risk difference with Parent and day-care staff training versus standard day-care for adaptive behaviour as an indirect outcome** (95% CI)
Self-care (measured with: Early Intervention Developmental Profile (EIDP)/Preschool Developmental Profile (PSDP): Self-Care; Better indicated by lower values)											
35 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⨁⨁⨀⨀ **LOW**[1] due to imprecision	19	16	N/A	N/A	The mean self-care in the intervention groups was **0.04 standard deviations lower** (0.7 lower to 0.63 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Combined parent training and antipsychotic versus antipsychotic-only for adaptive behaviour as an indirect outcome

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 Combined antipsychotic and parent training versus antipsychotic only for adaptive behaviour as an**		**Risk with Control**	**Risk difference with Combined antipsychotic and parent training versus antipsychotic only for adaptive behaviour as an indirect outcome** (95% CI)

							indirect outcome				
Adaptive behaviour (measured with: Vineland Adaptive Behaviour Scale (VABS): Adaptive Composite; Better indicated by lower values)											
124 (1 study) 24 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	49	75	N/A	N/A	The mean adaptive behaviour in the intervention groups was **0.56 standard deviations higher** (0.19 to 0.93 higher)
Daily living skills (measured with: Vineland Adaptive Behaviour Scale (VABS): Daily Living Skills; Better indicated by lower values)											
124 (1 study) 24 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	49	75	N/A	N/A	The mean daily living skills in the intervention groups was **0.48 standard deviations higher** (0.12 to 0.85 higher)
Socialization (measured with: Vineland Adaptive Behaviour Scale (VABS): Socialization; Better indicated by lower values)											
124 (1 study) 24 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	49	75	N/A	N/A	The mean socialization in the intervention groups was **0.6 standard deviations higher** (0.23 to 0.96 higher)
Communication (measured with: Vineland Adaptive Behaviour Scale (VABS): Communication; Better indicated by lower values)											
124 (1 study) 24 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	49	75	N/A	N/A	The mean communication in the intervention groups was **0.47 standard deviations higher** (0.11 to 0.84 higher)

[1] High risk of selection bias as significant group differences at baseline on this outcome measure. High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome measure based on interview with parents who were non-blind. Also high risk of attrition bias due to higher dropout rates in the experimental (combined risperidone and parent training) group (N=20; 27% attrition) than the control (risperidone only) group (N=9; 18% attrition)
[2] N<400

1.13.4 Social-communication interventions for adaptive behaviour as an indirect outcome

Caregiver-mediated social communication intervention versus treatment-as-usual for adaptive behaviour as an indirect outcome

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 Caregiver-mediated social-communication interventions versus treatment-as-usual for adaptive behaviour as an indirect outcome		Risk with Control	Risk difference with Caregiver-mediated social-communication interventions versus treatment-as-usual for adaptive behaviour as an indirect outcome (95% CI)
Adaptive behaviour (measured with: Vineland Adaptive Behaviour Scale (VABS): Total; Better indicated by lower values)											
152 (1 study) 56 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	75	77	N/A	N/A	The mean adaptive behaviour in the intervention groups was **0.17 standard deviations lower** (0.48 lower to 0.15 higher)
Daily Living Skills (measured with: Vineland Adaptive Behaviour Scale (VABS): Daily Living Skills; Better indicated by lower values)											
39 (1 study) 39 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to risk of bias, imprecision	20	19	N/A	N/A	The mean daily living skills in the intervention groups was **0.55 standard deviations higher** (0.09 lower to 1.19 higher)
Socialization (measured with: Vineland Adaptive Behaviour Scale (VABS): Socialization; Better indicated by lower values)											
39 (1 study) 39 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to risk of bias, imprecision	20	19	N/A	N/A	The mean socialization in the intervention groups was **0.1 standard deviations higher** (0.53 lower to 0.73 higher)

Communication (measured with: Vineland Adaptive Behaviour Scale (VABS): Communication; Better indicated by lower values)										
245 (4 studies) 39-56 weeks	serious[5]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2,5] due to risk of bias, imprecision	122	123	N/A	N/A	The mean communication in the intervention groups was **0.04 standard deviations lower** (0.29 lower to 0.22 higher)

[1] High risk of performance and response bias as intervention administrator and participants were non-blind, and unclear/unknown risk of detection bias as teacher-rated and blinding of teacher not reported
[2] N<400
[3] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias was unclear/unknown as outcome measure based on interview with non-blind parent rather than direct behavioural observation
[4] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[5] High risk of performance and response bias as intervention administrators and participants were non-blind, and unclear/unknown risk of detection bias as blinding of outcome assessment is unclear

Social skills group versus treatment-as-usual for adaptive behaviour as an indirect outcome

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 Treatment-as-usual	**With Social skills group**		**Risk with Treatment-as-usual**	**Risk difference with Social skills group** (95% CI)
Self-control (measured with: Social Skills Rating System (SSRS): Self-control; Better indicated by lower values)											
68 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	33	35	N/A	N/A	The mean self-control in the intervention groups was **0.63 standard deviations higher** (0.14 to 1.11 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as parent-rated and parents were non-blind and involved in the intervention. There was also a high risk of attrition bias due to a greater drop-out rate in the experimental (N=14; 35%) than in the control (N=5; 14%) group
[2] N<400

LEGO® therapy versus SULP for adaptive behaviour as an indirect outcome

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 LEGO therapy versus Social Use of Language Programme (SULP) for adaptive behaviour as an indirect outcome		Risk with Control	Risk difference with LEGO therapy versus Social Use of Language Programme (SULP) for adaptive behaviour as an indirect outcome (95% CI)
Socialization (measured with: Vineland Adaptive Behaviour Scale (VABS): Socialization; Better indicated by lower values)											
31 (1 study) 18 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	15	16	N/A	N/A	The mean socialization in the intervention groups was **0.32 standard deviations higher** (0.39 lower to 1.03 higher)
Communication (measured with: Vineland Adaptive Behaviour Scale (VABS): Communication; Better indicated by lower values)											
31 (1 study) 18 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	15	16	N/A	N/A	The mean communication in the intervention groups was **0.48 standard deviations higher** (0.23 lower to 1.2 higher)

[1] High risk of performance and response bias as intervention administrator and participants were non-blind, and risk of detection bias is unclear/unknown as although the interviewer was a blinded research assistant, the outcome measure was based on non-blind parent report
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.14 PHARMACOLOGICAL INTERVENTIONS AIMED AT ADAPTIVE BEHAVIOUR

1.14.1 Antipsychotics for adaptive behaviour as an indirect outcome

Aripiprazole versus placebo for adaptive behaviour as an indirect outcome

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 Antipsychotics versus placebo for adaptive behaviour as an indirect outcome		Risk with Control	Risk difference with Antipsychotics versus placebo for adaptive behaviour as an indirect outcome (95% CI)
Adaptive behaviour (aripiprazole) (measured with: PedsQL: Total (change score); Better indicated by lower values)											
243 (2 studies) 8 weeks	serious[1]	very serious[2]	no serious indirectness	serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, inconsistency, imprecision	76	167	N/A	N/A	The mean adaptive behaviour (aripiprazole) in the intervention groups was **0.51 standard deviations higher** (0.21 to 0.8 higher)
Emotional functioning (aripiorazole) (measured with: PedsQL: Emotional functioning (change score); Better indicated by lower values)											
243 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	76	167	N/A	N/A	The mean emotional functioning (aripiorazole) in the intervention groups was **0.41 standard deviations higher** (0.12 to 0.7 higher)
Social functioning (aripiorazole) (measured with: PedsQL: Social functioning (change score); Better indicated by lower values)											
243 (2 studies) 8 weeks	serious[1]	very serious[2]	no serious indirectness	very serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,4] due to risk of bias, inconsistency, imprecision	76	167	N/A	N/A	The mean social functioning (aripiorazole) in the intervention groups was **0.27 standard deviations higher** (0.02 lower to 0.56 higher)
Cognitive functioning (aripiprazole) (measured with: PedsQL: Cognitive functioning (change score); Better indicated by lower values)											
242 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias,	75	167	N/A	N/A	The mean cognitive functioning (aripiprazole) in the intervention groups

						imprecision			was **0.4 standard deviations higher** (0.11 to 0.69 higher)

[1] Risk of detection bias is unclear as blinding of parents not reported
[2] I-squared value indicates substantial to considerable heterogeneity
[3] N<400
[4] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Low dose aripiprazole versus placebo for adaptive behaviour as an indirect outcome

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 Low dose antipsychotics versus placebo for adaptive behaviour as an indirect outcome**		**Risk with Control**	**Risk difference with Low dose antipsychotics versus placebo for adaptive behaviour as an indirect outcome** (95% CI)
Adaptive behaviour (low dose aripiprazole 5mg/day) (measured with: PedsQL: Total (change score); Better indicated by lower values)											
80 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	37	43	N/A	N/A	The mean adaptive behaviour (low dose aripiprazole 5mg/day) in the intervention groups was **0.21 standard deviations higher** (0.23 lower to 0.65 higher)
Emotional functioning (low dose aripiprazole 5mg/day) (measured with: PedsQL: Emotional functioning (change score); Better indicated by lower values)											
80 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	37	43	N/A	N/A	The mean emotional functioning (low dose aripiprazole 5mg/day) in the intervention groups was **0.19 standard deviations higher**

										(0.25 lower to 0.63 higher)
Social functioning (low dose aripiprazole 5mg/day) (measured with: PedsQL: Social functioning (change score); Better indicated by lower values)										
80 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	37 43	N/A	N/A	The mean social functioning (low dose aripiprazole 5mg/day) in the intervention groups was **0 standard deviations higher** (0.43 lower to 0.44 higher)
Cognitive functioning (low dose aripiprazole 5mg/day) (measured with: PedsQL: Cognitive functioning (change score); Better indicated by lower values)										
80 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	37 43	N/A	N/A	The mean cognitive functioning (low dose aripiprazole 5mg/day) in the intervention groups was **0.32 standard deviations higher** (0.12 lower to 0.76 higher)

[1] Risk of detection bias is unclear as blinding of parents not reported
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] N<400

1.15 BIOMEDICAL INTERVENTIONS AIMED AT ADAPTIVE BEHAVIOUR

1.15.1 Complementary therapies for adaptive behaviour as an indirect outcome

Acupuncture/electro-acupuncture versus sham acupuncture/electro-acupuncture for adaptive behaviour as an indirect outcome

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 Acupuncture/Electro-acupuncture versus sham**		**Risk with** / **Risk difference with Acupuncture/Electro-**

								acupuncture/electro-acupuncture for adaptive behaviour as an indirect outcome		**Control**	**acupuncture versus sham acupuncture/electro-acupuncture for adaptive behaviour as an indirect outcome** (95% CI)
Adaptive behaviour (measured with: Functional Independence Measure for Children (WeeFIM): Total (change score); Better indicated by lower values)											
105 (2 studies) 4-9 weeks	no serious risk of bias	very serious[1]	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to inconsistency, imprecision, publication bias	50	55	N/A	N/A	The mean adaptive behaviour in the intervention groups was **0.59 standard deviations higher** (0.19 to 0.98 higher)
Self-care (measured with: Functional Independence Measure for Children (WeeFIM): Self-care (change score); Better indicated by lower values)											
105 (2 studies) 4-9 weeks	no serious risk of bias	very serious[1]	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to inconsistency, imprecision, publication bias	50	55	N/A	N/A	The mean self-care in the intervention groups was **0.56 standard deviations higher** (0.17 to 0.96 higher)
Mobility (measured with: Functional Independence Measure for Children (WeeFIM): Mobility (change score); Better indicated by lower values)											
105 (2 studies) 4-9 weeks	no serious risk of bias	serious[4]	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[2,3,4] due to inconsistency, imprecision, publication bias	50	55	N/A	N/A	The mean mobility in the intervention groups was **0.08 standard deviations lower** (0.46 lower to 0.31 higher)
Cognition (measured with: Functional Independence Measure for Children (WeeFIM): Cognition (change score); Better indicated by lower values)											
105 (2 studies) 4-9 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊕⊖⊖ **LOW**[2,3] due to imprecision, publication bias	50	55	N/A	N/A	The mean cognition in the intervention groups was **0.48 standard deviations higher** (0.09 to 0.87 higher)
Comprehension (measured with: Functional Independence Measure for Children (WeeFIM): Compehension (change score); Better indicated by lower values)											

55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[3,5] due to imprecision, publication bias	25	30	N/A	N/A	The mean comprehension in the intervention groups was **0.51 standard deviations higher** (0.03 lower to 1.05 higher)
Expression (measured with: Functional Independence Measure for Children (WeeFIM): Expression (change score); Better indicated by lower values)											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[3,5] due to imprecision, publication bias	25	30	N/A	N/A	The mean expression in the intervention groups was **0.17 standard deviations higher** (0.36 lower to 0.7 higher)
Social interaction (measured with: Functional Independence Measure for Children (WeeFIM): Social interaction (change score); Better indicated by lower values)											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[3,5] due to imprecision, publication bias	25	30	N/A	N/A	The mean social interaction in the intervention groups was **0.23 standard deviations lower** (0.77 lower to 0.3 higher)
Problem solving (measured with: Functional Independence Measure for Children (WeeFIM): Problem solving (change score); Better indicated by lower values)											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[3,5] due to imprecision, publication bias	25	30	N/A	N/A	The mean problem solving in the intervention groups was **0.24 standard deviations lower** (0.77 lower to 0.3 higher)
Memory (measured with: Functional Independence Measure for Children (WeeFIM): Memory (change score); Better indicated by lower values)											
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[3,5] due to imprecision, publication bias	25	30	N/A	N/A	The mean memory in the intervention groups was **0.13 standard deviations higher** (0.4 lower to 0.67 higher)

Self-care (functional skill) (measured with: Pediatric Evaluation of Disability Inventory (PEDI): Self-care; Better indicated by lower values)										
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[3,5] due to imprecision, publication bias	25	30	N/A	N/A	The mean self-care (functional skill) in the intervention groups was **0.22 standard deviations lower** (0.75 lower to 0.31 higher)
Self-care (independence) (measured with: Pediatric Evaluation of Disability Inventory (PEDI): Self-care (caregiver assistant); Better indicated by lower values)										
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[3,5] due to imprecision, publication bias	25	30	N/A	N/A	The mean self-care (independence) in the intervention groups was **0.44 standard deviations lower** (0.97 lower to 0.1 higher)
Mobility (functional skill) (measured with: Pediatric Evaluation of Disability Inventory (PEDI): Mobility; Better indicated by lower values)										
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[3,5] due to imprecision, publication bias	25	30	N/A	N/A	The mean mobility (functional skill) in the intervention groups was **0.11 standard deviations lower** (0.64 lower to 0.42 higher)
Mobility (independence) (measured with: Pediatric Evaluation of Disability Inventory (PEDI): Mobility (caregiver assistant); Better indicated by lower values)										
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[3,5] due to imprecision, publication bias	25	30	N/A	N/A	The mean mobility (independence) in the intervention groups was **0.19 standard deviations lower** (0.72 lower to 0.35 higher)
Social function (functional skill) (measured with: Pediatric Evaluation of Disability Inventory (PEDI): Social function; Better indicated by lower values)										
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[3,5] due to imprecision,	25	30	N/A	N/A	The mean social function (functional skill) in the intervention groups was **0.04 standard deviations**

						publication bias				**higher** (0.49 lower to 0.57 higher)
Social function (independence) (measured with: Pediatric Evaluation of Disability Inventory (PEDI): Social function (caregiver assistant); Better indicated by lower values)										
55 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[3,5] due to imprecision, publication bias	25	30	N/A	N/A	The mean social function (independence) in the intervention groups was **0.14 standard deviations lower** (0.67 lower to 0.39 higher)

[1] I-squared value indicates substantial to considerable heterogeneity
[2] N<400
[3] High risk of selective reporting bias as trial protocol for WONG2010B states that follow-up measurements will be taken but these are not reported
[4] I-squared value indicates moderate heterogeneity
[5] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Acupuncture/electro-acupuncture and conventional educational programme versus conventional educational programme only for adaptive behaviour as an indirect outcome

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 Acupuncture/electro-acupuncture and conventional educational programme versus conventional educational programme only for adaptive behaviour as an indirect outcome**		**Risk with Control**	**Risk difference with Acupuncture/electro-acupuncture and conventional educational programme versus conventional educational programme only for adaptive behaviour as an indirect outcome** (95% CI)
Adaptive behaviour (measured with: Functional Independence Measure for Children (WeeFIM): Total (change score); Better indicated by lower values)											
64 (2 studies) 8 weeks	serious[1]	very serious[2]	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, inconsistency, imprecision	31	33	N/A	N/A	The mean adaptive behaviour in the intervention groups was **0.41 standard deviations higher** (0.11 lower to 0.93 higher)

Self-care (measured with: Functional Independence Measure for Children (WeeFIM): Self-care (change score); Better indicated by lower values)											
64 (2 studies) 8 weeks	serious[1]	very serious[2]	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, inconsistency, imprecision	31	33	N/A	N/A	The mean self-care in the intervention groups was **0.16 standard deviations higher** (0.35 lower to 0.67 higher)
Mobility (measured with: Functional Independence Measure for Children (WeeFIM): Mobility (change score); Better indicated by lower values)											
64 (2 studies) 8 weeks	serious[1]	very serious[2]	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, inconsistency, imprecision	31	33	N/A	N/A	The mean mobility in the intervention groups was **0.52 standard deviations higher** (0 to 1.05 higher)
Cognition (measured with: Functional Independence Measure for Children (WeeFIM): Cognition (change score); Better indicated by lower values)											
64 (2 studies) 8 weeks	serious[1]	very serious[2]	no serious indirectness	serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,4] due to risk of bias, inconsistency, imprecision	31	33	N/A	N/A	The mean cognition in the intervention groups was **0.62 standard deviations higher** (0.1 to 1.14 higher)
Comprehension (measured with: Functional Independence Measure for Children (WeeFIM): Compehension (change score); Better indicated by lower values)											
36 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	18	18	N/A	N/A	The mean comprehension in the intervention groups was **0.47 standard deviations lower** (1.13 lower to 0.19 higher)
Expression (measured with: Functional Independence Measure for Children (WeeFIM): Expression (change score); Better indicated by lower values)											
36 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	18	18	N/A	N/A	The mean expression in the intervention groups was **0.4 standard deviations higher** (0.26 lower to 1.06 higher)

Social interaction (measured with: Functional Independence Measure for Children (WeeFIM): Social interaction (change score); Better indicated by lower values)										
36 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	18	18	N/A	N/A	The mean social interaction in the intervention groups was **0.4 standard deviations higher** (0.26 lower to 1.06 higher)
Problem solving (measured with: Functional Independence Measure for Children (WeeFIM): Problem solving (change score); Better indicated by lower values)										
36 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	18	18	N/A	N/A	The mean problem solving in the intervention groups was **0.33 standard deviations higher** (0.32 lower to 0.99 higher)
Memory (measured with: Functional Independence Measure for Children (WeeFIM): Memory (change score); Better indicated by lower values)										
36 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	18	18	N/A	N/A	The mean memory in the intervention groups was **0.15 standard deviations lower** (0.81 lower to 0.5 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and the conventional education programme differed for each participant which may introduce bias. The risk of detection bias was also unclear/unknown as all outcome measures were rated by blinded assessors, but some outcome measures involved input from parents who were not blind to treatment allocation or confounding variables and systematic review from which data was extracted does not report which outcome measures relied on non-blind parental report
[2] I-squared value indicates considerable heterogeneity
[3] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[4] N<400

1.15.2 Hormones for adaptive behaviour as an indirect outcome

Secretin versus placebo for adaptive behaviour as an indirect outcome

Quality assessment							Summary of Findings		
Participants	**Risk of**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication**	**Overall**	**Study event rates (%)**	**Relative**	**Anticipated absolute effects**

(studies) Follow up	bias				bias	quality of evidence	With Control	With Secretin versus placebo for adaptive behaviour as an indirect outcome	effect (95% CI)	Risk with Control	Risk difference with Secretin versus placebo for adaptive behaviour as an indirect outcome (95% CI)
Adaptive behaviour (measured with: Vineland Adaptive Behaviour Scale (VABS): Adaptive Composite; Better indicated by lower values)											
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	28	28	N/A	N/A	The mean adaptive behaviour in the intervention groups was **0.08 standard deviations lower** (0.61 lower to 0.44 higher)
Daily living skills (measured with: Vineland Adaptive Behaviour Scale (VABS): Daily Living Skills; Better indicated by lower values)											
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	28	28	N/A	N/A	The mean daily living skills in the intervention groups was **0.11 standard deviations higher** (0.42 lower to 0.63 higher)
Socialization (measured with: Vineland Adaptive Behaviour Scale (VABS): Socialization; Better indicated by lower values)											
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	28	28	N/A	N/A	The mean socialization in the intervention groups was **0.26 standard deviations lower** (0.78 lower to 0.27 higher)
Communication (measured with: Vineland Adaptive Behaviour Scale (VABS): Communication; Better indicated by lower values)											
112 (2 studies) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	56	56	N/A	N/A	The mean communication in the intervention groups was **0.28 standard deviations lower** (0.65 lower to 0.1 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.15.3 Medical procedures for adaptive behaviour as an indirect outcome

Long-term chelation (7-rounds of DMSA therapy) versus short-term chelation (1-round of DMSA therapy and 6-rounds of placebo) for adaptive behaviour as an indirect outcome

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 Short-term chelation (1-round of DMSA therapy and 6-rounds of placebo)	With Long-term chelation (7-rounds of Dimercaptosuccinic Acid [DMSA] therapy)		Risk with Short-term chelation (1-round of DMSA therapy and 6-rounds of placebo)	Risk difference with Long-term chelation (7-rounds of Dimercaptosuccinic Acid [DMSA] therapy) (95% CI)
Adaptive behaviour (measured with: Pervasive Development Disorder Behavior Inventory (PDDBI): Adaptive Behaviours Composite; Better indicated by lower values)											
40 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	25	N/A	N/A	The mean adaptive behaviour in the intervention groups was **0.2 standard deviations lower** (0.84 lower to 0.44 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

HBOT versus attention-placebo for adaptive behaviour as an indirect outcome

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 Attention-placebo control	With Hyperbaric oxygen treatment (HBOT)		Risk with Attention-placebo control	Risk difference with Hyperbaric oxygen treatment (HBOT) (95% CI)
Adaptive behaviour (measured with: Vineland Adaptive Behaviour Scale (VABS): Adaptive Composite (change score); Better indicated by lower values)											

34 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	16	18	N/A	N/A	The mean adaptive behaviour in the intervention groups was **0.18 standard deviations lower** (0.85 lower to 0.5 higher)
Daily living skills (measured with: Vineland Adaptive Behaviour Scale (VABS): Daily Living Skills (change score); Better indicated by lower values)											
34 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	16	18	N/A	N/A	The mean daily living skills in the intervention groups was **0.11 standard deviations higher** (0.56 lower to 0.78 higher)
Socialization (measured with: Vineland Adaptive Behaviour Scale (VABS): Socialization (change score); Better indicated by lower values)											
34 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	16	18	N/A	N/A	The mean socialization in the intervention groups was **0.38 standard deviations lower** (1.06 lower to 0.3 higher)
Communication (measured with: Vineland Adaptive Behaviour Scale (VABS): Communication (change score); Better indicated by lower values)											
34 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	16	18	N/A	N/A	The mean communication in the intervention groups was **0.23 standard deviations higher** (0.45 lower to 0.9 higher)
Clinician-rated positive treatment response (assessed with: Number of participants 'much improved/very improved' on Clinical Global Impression (CGI)-improvement for overall functioning)											

56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	2/26 (7.7%)	9/30 (30%)	**RR 3.9** (0.92 to 16.45)	**Study population**	
										77 per 1000	**223 more per 1000** (from 6 fewer to 1000 more)
										Moderate	
										77 per 1000	**223 more per 1000** (from 6 fewer to 1000 more)
Parent-rated positive treatment response (assessed with: Number of participants 'much improved/very improved' on Parental Global Impression (PGI)-improvement for overall functioning)											
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	4/26 (15.4%)	9/30 (30%)	**RR 1.95** (0.68 to 5.6)	**Study population**	
										154 per 1000	**146 more per 1000** (from 49 fewer to 708 more)
										Moderate	
										154 per 1000	**146 more per 1000** (from 49 fewer to 708 more)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)

1.15.4 Nutritional interventions for adaptive behaviour as an indirect outcome

Omega-3 fatty acids versus placebo for adaptive behaviour as an indirect outcome

Quality assessment							Summary of Findings		
Participants (studies)	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall quality of evidence**	**Study event rates (%)**	**Relative effect**	**Anticipated absolute effects**

Follow up							With Placebo	With Omega-3 fatty acids	(95% CI)	Risk with Placebo	Risk difference with Omega-3 fatty acids (95% CI)
Adaptive skill (measured with: Behavior Assessment System for Children (BASC): Adaptive skill; Better indicated by lower values)											
24 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	12	12	N/A	N/A	The mean adaptive skill in the intervention groups was **0.2 standard deviations lower** (1 lower to 0.6 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Omega-3 fatty acids versus healthy diet control for adaptive behaviour as an indirect outcome

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 Healthy diet control	With Omega-3 fatty acids		Risk with Healthy diet control	Risk difference with Omega-3 fatty acids (95% CI)
Frequency of attending to task/activity (measured with: Behavioural observation; Better indicated by lower values)											
23 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	13	10	N/A	N/A	The mean frequency of attending to task/activity in the intervention groups was **0.65 standard deviations higher** (0.2 lower to 1.5 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Gluten-free and casein-free diet versus treatment-as-usual for adaptive behaviour as an indirect outcome

Quality assessment							Summary of Findings		
Participants	Risk of	Inconsistency	Indirectness	Imprecision	Publication	Overall quality	Study event rates (%)	Relative	Anticipated absolute effects

(studies) Follow up	bias				bias	of evidence	With Treatment-as-usual	With Gluten-free and casein-free diet	effect (95% CI)	Risk with Treatment-as-usual	Risk difference with Gluten-free and casein-free diet (95% CI)
Daily Living Skills (measured with: Vineland Adaptive Behaviour Scale (VABS): Daily Living Skills (change score); Better indicated by lower values)											
55 (1 study) 35 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	29	26	N/A	N/A	The mean daily living skills in the intervention groups was **0.32 standard deviations higher** (0.21 lower to 0.85 higher)
Socialization (measured with: Vineland Adaptive Behaviour Scale (VABS): Socialization (change score); Better indicated by lower values)											
55 (1 study) 35 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	29	26	N/A	N/A	The mean socialization in the intervention groups was **0.05 standard deviations higher** (0.48 lower to 0.58 higher)
Communication (measured with: Vineland Adaptive Behaviour Scale (VABS): Communication (change score); Better indicated by lower values)											
55 (1 study) 35 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	29	26	N/A	N/A	The mean communication in the intervention groups was **0.12 standard deviations lower** (0.65 lower to 0.41 higher)

[1] High risk of performance and response bias as intervention administrators (parents) and participants were non-blind and high risk of detection bias as parent-reported and non-blind to treatment allocation and other potentially confounding factors. There was also a high risk of attrition bias as over twice as many dropouts in the experimental group relative to the controls (32% in experimental group and 15% in the control group)
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.16 PSYCHOSOCIAL INTERVENTIONS AIMED AT SPEECH AND LANGUAGE

1.16.1 AAC interventions for speech and language as a direct outcome

PECS training for teachers versus treatment-as-usual for speech and language as a direct outcome

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 No treatment	**With Picture Exchange Communication System (PECS) training for teachers**		**Risk with No treatment**	**Risk difference with Picture Exchange Communication System (PECS) training for teachers** (95% CI)
Spontaneous child communicative initiations (assessed with: Behavioural observation (odds of being in a higher initiation category))											
0 (1 study) 33 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to risk of bias, imprecision	N/A	N/A	**OR 2.73** (1.22 to 6.09)	**Study population**	
										N/A	N/A
										Moderate	
										0 per 1000	N/A
Spontaneous child communicative initiations (assessed with: Behavioural observation (odds of being in a higher initiation category))											
0 (1 study) 78 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	N/A	N/A	**OR 1.08** (0.3 to 3.89)	**Study population**	
										N/A	N/A
										Moderate	
										0 per	N/A

										1000	
PECS use (assessed with: Behavioural observation (odds of being in a higher category for rate of PECS use))											
0 (1 study) 33 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	N/A	N/A	**OR 3.90** (1.75 to 8.69)	**Study population**	
										N/A	N/A
										Moderate	
										0 per 1000	N/A
PECS use (assessed with: Behavioural observation (odds of being in a higher category for rate of PECS use))											
0 (1 study) 78 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	N/A	N/A	**OR 1.56** (0.46 to 5.3)	**Study population**	
										N/A	N/A
										Moderate	
										0 per 1000	N/A
Speech/vocalisation use (assessed with: Behavioural observation (odds of being in a higher category for rate of speech/vocalisation use))											
0 (1 study) 33 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	N/A	N/A	**OR 1.10** (0.46 to 2.63)	**Study population**	
										N/A	N/A
										Moderate	
										0 per 1000	N/A
Receptive language (assessed with: British Picture Vocabulary test (BPVS): Receptive language (odds of being in a higher category on BPVS))											
0	serious[1]	no serious	no serious	very	undetected	⊕⊖⊖⊖	N/A	N/A	**OR 1.54**	**Study population**	

(1 study) 33 weeks		inconsistency	indirectness	serious[3]		**VERY LOW**[1,3] due to risk of bias, imprecision		(0.52 to 4.55)	N/A	N/A
									Moderate	
									0 per 1000	N/A
Expressive language (assessed with: Expressive One Word Picture Vocabulary Test (EOWPVT) Expressive language (odds of being in a higher category on EOWPVT))										
0 (1 study) 33 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to risk of bias, imprecision	N/A	N/A	**OR 1.01** (0.89 to 1.15)	**Study population**
									N/A	N/A
									Moderate	
									0 per 1000	N/A

[1] High risk of performance, response and detection bias as intervention administrators, participants and outcome assessors were non-blind
[2] Events<300
[3] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm

PECS versus RPMT for speech and language as a direct outcome

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 Responsive Education and Prelinguistic Milieu Teaching (RPMT)	**With Picture Exchange Communication System (PECS)**		**Risk with Responsive Education and Prelinguistic Milieu Teaching (RPMT)**	**Risk difference with Picture Exchange Communication System (PECS)** (95% CI)
Frequency of nonimitative spoken acts (measured with: Behavioural observation; Better indicated by lower values)											

36 (1 study) 26 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	17	19	N/A	N/A	The mean frequency of nonimitative spoken acts in the intervention groups was **0.61 standard deviations higher** (0.06 lower to 1.28 higher)
Frequency of nonimitative spoken acts (measured with: Behavioural observation; Better indicated by lower values)											
36 (1 study) 52 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	17	19	N/A	N/A	The mean frequency of nonimitative spoken acts in the intervention groups was **0.03 standard deviations higher** (0.62 lower to 0.68 higher)
Number of different nonimitative words (measured with: Behavioural observation; Better indicated by lower values)											
36 (1 study) 26 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	17	19	N/A	N/A	The mean number of different nonimitative words in the intervention groups was **0.49 standard deviations higher** (0.18 lower to 1.15 higher)
Number of different nonimitative words (measured with: Behavioural observation; Better indicated by lower values)											

36 (1 study) 52 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	17	19	N/A	N/A	The mean number of different nonimitative words in the intervention groups was **0.08 standard deviations higher** (0.57 lower to 0.74 higher)
Number of picture exchanges (measured with: EScs-Abridged (Early Social Communication Scales-Abridged): Number of picture exchanges; Better indicated by lower values)											
36 (1 study) 26 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	17	19	N/A	N/A	The mean number of picture exchanges in the intervention groups was **0.8 standard deviations higher** (0.12 to 1.48 higher)

[1] High risk of performance bias as intervention administrators were non-blind and comparison groups did not receive the same care apart from the intervention studied (parents in the RPMT group chose to receive more hours of training [mean: 10.6 hours] than parents in the PECS group [mean 7.9 hours]. In addition, the number of hours of 'other intervention' increased between the treatment and follow-up periods, and this increase was greater for the PECS group [4 hours] than for the RPMT group [-0.3 hours]). There was also a high risk of response bias as participants were non-blind and detection bias as identity and blinding of outcome assessors is not reported
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] N<400
[4] High risk of selective reporting bias as only post-intervention (and not 6-month post-intervention follow-up) reported for the only outcome where significant treatment effects observed (number of picture exchanges as assessed by the EScs-Abridged)

1.16.2 Arts-based interventions for speech and language as a direct outcome

Music therapy versus treatment-as-usual for speech and language as a direct outcome

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 Waitlist or treatment-as-usual control	With Music therapy		Risk with Waitlist or treatment-as-usual control	Risk difference with Music therapy (95% CI)
Verbal communication (measured with: Childhood Autism Rating Scale (CARS): Verbal communication; Better indicated by lower values)											
24 (1 study) 30 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	12	12	N/A	N/A	The mean verbal communication in the intervention groups was **0.09 standard deviations lower** (0.89 lower to 0.71 higher)
Non-verbal communication (measured with: Childhood Autism Rating Scale (CARS): Non-verbal communication; Better indicated by lower values)											
24 (1 study) 30 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	12	12	N/A	N/A	The mean non-verbal communication in the intervention groups was **0.35 standard deviations higher** (0.45 lower to 1.16 higher)
Expressive language (music therapy) (measured with: Verbal Production Evaluation Scale (VPES; study-specific): Expressive language; Better indicated by lower values)											
32 (1 study) 4 days	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	14	18	N/A	N/A	The mean expressive language (music therapy) in the intervention groups was **1.22 standard deviations higher** (0.45 to 1.99 higher)
Expressive language (speech therapy) (measured with: Verbal Production Evaluation Scale (VPES; study-specific): Expressive language; Better indicated by lower											

values)										
32 (1 study) 4 days	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	14	18	N/A	N/A	The mean expressive language (speech therapy) in the intervention groups was **1.09 standard deviations higher** (0.33 to 1.84 higher)
[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5) [2] N<400										

1.16.3 Behavioural interventions for speech and language as an indirect outcome

EIBI or EBI (ESDM or P-ESDM) versus treatment-as-usual for speech and language as an indirect outcome

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 Treatment-as-usual	**With ESDM or P-ESDM**		**Risk with Treatment-as-usual**	**Risk difference with ESDM or P-ESDM** (95% CI)
Receptive language (ESDM) (measured with: Mullen Scales of Early Learning (MSEL): Receptive Language; Better indicated by lower values)											
45 (1 study) 104 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	21	24	N/A	N/A	The mean receptive language (esdm) in the intervention groups was **0.6 standard deviations higher** (0 to 1.2 higher)
Expressive language (ESDM) (measured with: Mullen Scales of Early Learning (MSEL): Expressive Language; Better indicated by lower values)											
45 (1 study)	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖	21	24	N/A	N/A	The mean expressive language (esdm) in the

104 weeks						**LOW**[1] due to imprecision				intervention groups was **0.55 standard deviations higher** (0.05 lower to 1.15 higher)
Phrases understood (measured with: MacArthur Communication Developmental Inventories (CDI): Phrases understood; Better indicated by lower values)										
98 (1 study) 12 weeks	serious[2]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	49	49	N/A	N/A	The mean phrases understood in the intervention groups was **0.23 standard deviations lower** (0.63 lower to 0.16 higher)
Vocabulary comprehension (measured with: MacArthur Communication Developmental Inventories (CDI): Vocabulary comprehension; Better indicated by lower values)										
98 (1 study) 12 weeks	serious[2]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	49	49	N/A	N/A	The mean vocabulary comprehension in the intervention groups was **0.19 standard deviations lower** (0.58 lower to 0.21 higher)
Vocabulary production (measured with: MacArthur Communication Developmental Inventories (CDI): Vocabulary production; Better indicated by lower values)										
98 (1 study) 12 weeks	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[2,3] due to risk of bias, imprecision	49	49	N/A	N/A	The mean vocabulary production in the intervention groups was **0.05 standard deviations higher** (0.35 lower to 0.45 higher)
Total gestures produced (measured with: MacArthur Communication Developmental Inventories (CDI): Total gestures produced; Better indicated by lower values)										
98 (1 study) 12 weeks	serious[2]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	49	49	N/A	N/A	The mean total gestures produced in the intervention groups was **0.13 standard deviations lower**

										(0.53 lower to 0.26 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] High risk of performance and response bias as intervention administrators and participants were non-blind and high risk of detection bias as come measure was parent-rated and parents were non-blind and involved in the intervention
[3] N<400

EIBI versus parent training for speech and language as an indirect outcome

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 EIBI versus parent training for speech and language as an indirect outcome**		**Risk with Control**	**Risk difference with EIBI versus parent training for speech and language as an indirect outcome** (95% CI)
Receptive language (measured with: Reynell Developmental Language Scale: Comprehension; Better indicated by lower values)											
28 (1 study) 260 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	13	15	N/A	N/A	The mean receptive language in the intervention groups was **0.48 standard deviations higher** (0.28 lower to 1.23 higher)
Expressive language (measured with: Reynell Developmental Language Scale: Expressive Language; Better indicated by lower values)											
28 (1 study) 260 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	13	15	N/A	N/A	The mean expressive language in the intervention groups was **0.36 standard deviations higher**

										(0.39 lower to 1.11 higher)
Receptive + Expressive language (measured with: Reynell Developmental Language Scale: Total; Better indicated by lower values)										
28 (1 study) 260 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	13	15	N/A	N/A	The mean receptive + expressive language in the intervention groups was **0.63 standard deviations higher** (0.13 lower to 1.39 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Home-based EBI versus centre-based EBI for speech and language as an indirect outcome

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 Home-based versus Centre-based EBI for speech and language as an indirect outcome**		**Risk with Control**	**Risk difference with Home-based versus Centre-based EBI for speech and language as an indirect outcome** (95% CI)
Receptive language (measured with: Reynell Developmental Language Scale: Comprehension; Better indicated by lower values)											
53 (1 study) 40 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	26	27	N/A	N/A	The mean receptive language in the intervention groups was **0.42 standard deviations lower** (0.96 lower to 0.13 higher)
Expressive language (measured with: Reynell Developmental Language Scale: Expressive Language; Better indicated by lower values)											

53 (1 study) 40 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	26	27	N/A	N/A	The mean expressive language in the intervention groups was **0.26 standard deviations lower** (0.8 lower to 0.28 higher)
Everyday language functioning (measured with: Pragmatics Profile: Total Q range; Better indicated by lower values)											
56 (1 study) 40 weeks	serious[2]	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	29	27	N/A	N/A	The mean everyday language functioning in the intervention groups was **0.52 standard deviations lower** (1.06 lower to 0.01 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias in unclear/unknown as although the outcome assessors were blinded, this outcome measure was based on interview with parent and parents were non-blind and were part of the intervention

1.16.4 Educational interventions for speech and language as a direct or indirect outcome

Combined TeachTown and IBI versus IBI-only for speech and language as a direct outcome

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 IBI-only	**With Combined computer-assisted educational intervention and intensive behavioural intervention (IBI) day class program**		**Risk with IBI-only**	**Risk difference with Combined computer-assisted educational intervention and intensive behavioural intervention (IBI) day class program** (95% CI)
Receptive language (measured with: Peabody Picture Vocabulary Test, 3rd Ed. (PPVT-III): Total; Better indicated by lower values)										

46 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	24	22	N/A	N/A	The mean receptive language in the intervention groups was **0.33 standard deviations higher** (0.25 lower to 0.92 higher)
Receptive language (preschool subgroup analysis) (measured with: Peabody Picture Vocabulary Test, 3rd Ed. (PPVT-III): Total; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	12	11	N/A	N/A	The mean receptive language (preschool subgroup analysis) in the intervention groups was **0.4 standard deviations higher** (0.43 lower to 1.22 higher)
Receptive language (K-1 subgroup analysis) (measured with: Peabody Picture Vocabulary Test, 3rd Ed. (PPVT-III): Total; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	12	11	N/A	N/A	The mean receptive language (k-1 subgroup analysis) in the intervention groups was **0.27 standard deviations higher** (0.55 lower to 1.09 higher)
Receptive language (measured with: Brigance Inventory of Child Development: Receptive language; Better indicated by lower values)											
46 (1 study) 13 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	24	22	N/A	N/A	The mean receptive language in the intervention groups was **0.09 standard deviations higher** (0.49 lower to 0.67 higher)
Receptive language (preschool subgroup analysis) (measured with: Brigance Inventory of Child Development: Receptive language; Better indicated by lower values)											

23 (1 study) 13 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	12	11	N/A	N/A	The mean receptive language (preschool subgroup analysis) in the intervention groups was **0.02 standard deviations lower** (0.84 lower to 0.8 higher)
Receptive language (K-1 subgroup analysis) (measured with: Brigance Inventory of Child Development: Receptive language; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	12	11	N/A	N/A	The mean receptive language (k-1 subgroup analysis) in the intervention groups was **0.2 standard deviations higher** (0.62 lower to 1.02 higher)
Expressive language (measured with: Expressive Vocabulary Test (EVT): Total; Better indicated by lower values)											
46 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	24	22	N/A	N/A	The mean expressive language in the intervention groups was **0.27 standard deviations higher** (0.31 lower to 0.85 higher)
Expressive language (preschool subgroup analysis) (measured with: Expressive Vocabulary Test (EVT): Total; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	12	11	N/A	N/A	The mean expressive language (preschool subgroup analysis) in the intervention groups was **0.33 standard deviations higher** (0.5 lower to 1.15 higher)
Expressive language (K-1 subgroup analysis) (measured with: Expressive Vocabulary Test (EVT): Total; Better indicated by lower values)											
23 (1 study)	serious[1]	no serious	no serious	very	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2]	12	11	N/A	N/A	The mean expressive language (k-1 subgroup analysis) in the

13 weeks		inconsistency	indirectness	serious[2]		due to risk of bias, imprecision					intervention groups was **0.22 standard deviations higher** (0.6 lower to 1.04 higher)
Expressive language (measured with: Brigance Inventory of Child Development: Expressive language; Better indicated by lower values)											
46 (1 study) 13 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	24	22	N/A	N/A	The mean expressive language in the intervention groups was **0.01 standard deviations higher** (0.57 lower to 0.59 higher)
Expressive language (preschool subgroup analysis) (measured with: Brigance Inventory of Child Development: Expressive language; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	12	11	N/A	N/A	The mean expressive language (preschool subgroup analysis) in the intervention groups was **0.07 standard deviations higher** (0.75 lower to 0.89 higher)
Expressive language (K-1 subgroup analysis) (measured with: Brigance Inventory of Child Development: Expressive language; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	12	11	N/A	N/A	The mean expressive language (k-1 subgroup analysis) in the intervention groups was **0.05 standard deviations lower** (0.87 lower to 0.77 higher)

[1] High risk of performance and response bias as intervention administrators and participants non-blind. Risk of detection bias is unclear/unknown as the identity and blinding of outcome assessors not reported.
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] High risk of performance and response bias as intervention administrators and participants non-blind. Risk of detection bias is unclear/unknown as the identity and blinding of outcome assessors not reported. In addition, for the Brigance Inventory of Child Development scale there are no independent reliability and/or validity data reported

LEAP training versus manual-only control for speech and language as an indirect outcome

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 Intervention-manual-only control	**With Inclusive educational intervention (LEAP) training**		**Risk with Intervention-manual-only control**	**Risk difference with Inclusive educational intervention (LEAP) training** (95% CI)
Language (measured with: Preschool Language Scale-4 (PLS-4): Total; Better indicated by lower values)											
294 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	117	177	N/A	N/A	The mean language in the intervention groups was **0.94 standard deviations higher** (0.7 to 1.19 higher)
Receptive language (measured with: Mullen Scales of Early Learning (MSEL): Receptive Language Age (months); Better indicated by lower values)											
294 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	117	177	N/A	N/A	The mean receptive language in the intervention groups was **1.1 standard deviations higher** (0.85 to 1.35 higher)
Expressive language (measured with: Mullen Scales of Early Learning (MSEL): Expressive Language Age (months); Better indicated by lower values)											
294 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias,	117	177	N/A	N/A	The mean expressive language in the intervention groups was **0.49 standard**

						imprecision				**deviations higher** (0.25 to 0.73 higher)

[1] High risk of performance and response bias as intervention administrators and participants non-blind. In addition, risk of detection bias is unclear/unknown as identity and blinding of outcome assessors not reported
[2] N<400

1.16.5 Parent training for speech and language as a direct or indirect outcome

Parent training versus treatment-as-usual for speech and language as a direct or indirect outcome

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 Parent training versus treatment-as-usual for speech and language**		**Risk with Control**	**Risk difference with Parent training versus treatment-as-usual for speech and language** (95% CI)
Receptive language (measured with: Mullen Scales of Early Learning (MSEL): Receptive Language or MacArthur Communication Developmental Inventories (CDI): Vocabulary Comprehension or Reynell Developmental Language Scale: Comprehension; Better indicated by lower values)											
147 (3 studies) 12-52 weeks	serious[1]	very serious[2]	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, inconsistency, imprecision	57	90	N/A	N/A	The mean receptive language in the intervention groups was **0.2 standard deviations lower** (0.54 lower to 0.14 higher)
Receptive language (direct outcome) (measured with: Mullen Scales of Early Learning (MSEL): Receptive Language; Better indicated by lower values)											
20 (1 study) 12 weeks	serious[4]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to risk of bias,	10	10	N/A	N/A	The mean receptive language (direct outcome) in the intervention groups was **0.09 standard deviations**

						imprecision				**higher** (0.78 lower to 0.97 higher)
Receptive language (indirect outcome) (measured with: MacArthur Communication Developmental Inventories (CDI): Vocabulary Comprehension; Better indicated by lower values)										
24 (1 study) 52 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,5] due to risk of bias, imprecision	12	12	N/A	N/A	The mean receptive language (indirect outcome) in the intervention groups was **0.71 standard deviations higher** (0.12 lower to 1.54 higher)
Receptive language (indirect outcome; PEC+PEBM combined) (measured with: Reynell Developmental Language Scale: Comprehension; Better indicated by lower values)										
103 (1 study) 46 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[6]	undetected	⊕⊕⊖⊖ **LOW**[1,6] due to risk of bias, imprecision	35	68	N/A	N/A	The mean receptive language (indirect outcome; pec+pebm combined) in the intervention groups was **0.5 standard deviations lower** (0.91 to 0.08 lower)
Expressive language (measured with: Mullen Scales of Early Learning (MSEL): Expressive Language or MacArthur Communication Developmental Inventories (CDI): Vocabulary Production or Reynell Developmental Language Scale: Expressive Language; Better indicated by lower values)										
147 (3 studies) 12-52 weeks	serious[1]	serious[7]	no serious indirectness	serious[6]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,6,7] due to risk of bias, inconsistency, imprecision	57	90	N/A	N/A	The mean expressive language in the intervention groups was **0.14 standard deviations lower** (0.48 lower to 0.2 higher)
Expressive language (direct outcome) (measured with: Mullen Scales of Early Learning (MSEL): Expressive Language; Better indicated by lower values)										

20 (1 study) 12 weeks	serious[4]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4] due to risk of bias, imprecision	10	10	N/A	N/A	The mean expressive language (direct outcome) in the intervention groups was **0.15 standard deviations lower** (1.03 lower to 0.73 higher)
Expressive language (indirect outcome) (measured with: MacArthur Communication Developmental Inventories (CDI): Vocabulary Production; Better indicated by lower values)											
24 (1 study) 52 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,5] due to risk of bias, imprecision	12	12	N/A	N/A	The mean expressive language (indirect outcome) in the intervention groups was **0.56 standard deviations higher** (0.26 lower to 1.38 higher)
Expressive language (indirect outcome; PEC+PEBM combined) (measured with: Reynell Developmental Language Scale: Expressive Language; Better indicated by lower values)											
103 (1 study) 46 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	35	68	N/A	N/A	The mean expressive language (indirect outcome; pec+pebm combined) in the intervention groups was **0.31 standard deviations lower** (0.72 lower to 0.1 higher)
Overall language rating of non-verbal (<5 words) (indirect outcome) (assessed with: Dichotomous: Overall language rating (based on ADI-R) of non-verbal (<5 words))											
24 (1 study) 52 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[8]	undetected	⊕⊖⊖⊖ **VERY LOW**[5,8] due to risk of bias, imprecision	9/12 (75%)	4/12 (33.3%)	**RR 0.44** (0.19 to 1.05)	**Study population**	
										750 per 1000	**420 fewer per 1000** (from 608 fewer to 37 more)
										Moderate	

									750 per 1000	**420 fewer per 1000** (from 608 fewer to 37 more)
Overall language rating of single word speech (indirect outcome) (assessed with: Dichotomous: Overall language rating (based on ADI-R) of single words)										
24 (1 study) 52 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[8]	undetected	⊕⊖⊖⊖ **VERY LOW**[5,8] due to risk of bias, imprecision	3/12 (25%)	5/12 (41.7%)	**RR 1.67** (0.51 to 5.46)	**Study population**
									250 per 1000	**167 more per 1000** (from 123 fewer to 1000 more)
									Moderate	
									250 per 1000	**167 more per 1000** (from 123 fewer to 1000 more)
Overall language rating of phrase speech (indirect outcome) (assessed with: Dichotomous: Overall language rating (based on ADI-R) of phrase speech)										
24 (1 study) 52 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[8]	undetected	⊕⊖⊖⊖ **VERY LOW**[5,8] due to risk of bias, imprecision	0/12 (0%)	3/12 (25%)	**RR 7** (0.4 to 122.44)	**Study population**
									0 per 1000	N/A
									Moderate	
									0 per 1000	N/A
Total gestures produced (indirect outcome) (measured with: MacArthur Communication Developmental Inventories (CDI): Total gestures produced; Better indicated by lower values)										
24 (1 study) 52 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,5] due to risk of bias,	12	12	N/A	N/A	The mean total gestures produced (indirect outcome) in the intervention groups was

						imprecision				**0.58 standard deviations higher** (0.24 lower to 1.4 higher)

[1] High risk of selection bias as baseline differences in TONGE2006/2012 between groups on this outcome measure
[2] I-squared value indicates considerable heterogeneity
[3] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[4] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear/unknown as the identity and blinding of outcome assessor/s are not reported
[5] High risk of performance and response bias as intervention administrators and participants were non-blind and high risk of detection bias as outcome measure was parent-rated and parents were non-blind and involved in the intervention
[6] N<400
[7] I-squared value indicates moderate heterogeneity
[8] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1,25)

Parent and day-care staff training versus standard day-care for speech and language as an indirect outcome

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 Parent and day-care staff training versus standard day-care for speech and language as an indirect outcome**		**Risk with Control**	**Risk difference with Parent and day-care staff training versus standard day-care for speech and language as an indirect outcome** (95% CI)
Language (measured with: Early Intervention Developmental Profile (EIDP)/Preschool Developmental Profile (PSDP): Language; Better indicated by lower values)											
35 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	19	16	N/A	N/A	The mean language in the intervention groups was **0.66 standard deviations higher** (0.03 lower to 1.34 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.16.6 Social-communication interventions for speech and language as an indirect outcome

Caregiver-mediated social communication intervention versus treatment-as-usual for speech and language as an indirect outcome

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 Caregiver-mediated social-communication interventions versus treatment-as-usual for speech and language as an indirect outcome		Risk with Control	Risk difference with Caregiver-mediated social-communication interventions versus treatment-as-usual for speech and language as an indirect outcome (95% CI)
Receptive language (clinician-rated) (measured with: Mullen Scales of Early Learning (MSEL): Receptive Language Age (months) or Preschool Language Scale-3 (PLS-3): Auditory Comprehension; Better indicated by lower values)											
225 (3 studies) 39-56 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	112	113	N/A	N/A	The mean receptive language (clinician-rated) in the intervention groups was **0.04 standard deviations higher** (0.23 lower to 0.30 higher)
Receptive language (parent-rated) (measured with: MacArthur Communication Developmental Inventories (CDI): Vocabulary Comprehension; Better indicated by lower values)											
180 (2 studies) 52-56 weeks	serious[2]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	89	91	N/A	N/A	The mean receptive language (parent-rated) in the intervention groups was **0.16 standard deviations higher** (0.13 lower to 0.45 higher)
Expressive language (clinician-rated) (measured with: Mullen Scales of Early Learning (MSEL): Expressive Language Age (months) or Preschool Language Scale-3 (PLS-3): Expressive Communication; Better indicated by lower values)											
225 (3 studies) 39-56 weeks	no serious risk of	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to	112	113	N/A	N/A	The mean expressive language (clinician-rated) in the intervention groups was

	bias					imprecision				**0.03 standard deviations higher** (0.23 lower to 0.29 higher)
Expressive language (parent-rated) (measured with: MacArthur Communication Developmental Inventories (CDI): Vocabulary Production; Better indicated by lower values)										
180 (2 studies) 52-56 weeks	serious[2]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	89	91	N/A	N/A	The mean expressive language (parent-rated) in the intervention groups was **0.05 standard deviations higher** (0.24 lower to 0.34 higher)

[1] N<400
[2] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as this outcome measure was parent-rated and parents were non-blind

Social skills group versus treatment-as-usual for speech and language as an indirect outcome

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 Treatment-as-usual	**With Social skills group**		**Risk with Treatment-as-usual**	**Risk difference with Social skills group** (95% CI)
Idiomatic language (measured with: Comprehensive Assessment of Spoken Language (CASL): Idiomatic Language; Better indicated by lower values)											
34 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	16	18	N/A	N/A	The mean idiomatic language in the intervention groups was **0.05 standard deviations higher** (0.62 lower to 0.73

										higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as researcher-rated and researchers were non-blind and no reliability or validity data for the use of this scale in this age group (only for >11 years)
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Joint attention training and EBI/EIBI versus EBI/EIBI only for speech and language as an indirect outcome

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 Combined joint attention training and EBI/EIBI versus EBI/EIBI only for speech and language as an indirect outcome**		**Risk with Control**	**Risk difference with Combined joint attention training and EBI/EIBI versus EBI/EIBI only for speech and language as an indirect outcome** (95% CI)
Receptive language (measured with: Reynell Developmental Language Scale: Comprehension or Mullen Scales of Early Learning (MSEL): Receptive language; Better indicated by lower values)											
85 (2 studies) 6-26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	41	44	N/A	N/A	The mean receptive language in the intervention groups was **0.27 standard deviations higher** (0.16 lower to 0.69 higher)
Receptive language (measured with: Reynell Developmental Language Scale: Comprehension or Mullen Scales of Early Learning (MSEL): Receptive language; Better indicated by lower values)											
85 (2 studies) 26-52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	41	44	N/A	N/A	The mean receptive language in the intervention groups was **0.23 standard deviations higher** (0.2 lower to 0.65 higher)

Receptive language (measured with: Reynell Developmental Language Scale: Comprehension; Better indicated by lower values)										
36 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	16 20	N/A	N/A	The mean receptive language in the intervention groups was **0.36 standard deviations higher** (0.31 lower to 1.02 higher)
Expressive language (measured with: Reynell Developmental Language Scale: Expressive Language or Mullen Scales of Early Learning (MSEL): Expressive Language; Better indicated by lower values)										
85 (2 studies) 6-26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	41 44	N/A	N/A	The mean expressive language in the intervention groups was **0.19 standard deviations higher** (0.23 lower to 0.62 higher)
Expressive language (measured with: Reynell Developmental Language Scale: Expressive Language or Mullen Scales of Early Learning (MSEL): Expressive Language; Better indicated by lower values)										
85 (2 studies) 26-52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	41 44	N/A	N/A	The mean expressive language in the intervention groups was **0.29 standard deviations higher** (0.14 lower to 0.72 higher)
Expressive language (measured with: Reynell Developmental Language Scale: Expressive Language; Better indicated by lower values)										
36 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	16 20	N/A	N/A	The mean expressive language in the intervention groups was **0.57 standard deviations higher** (0.1 lower to 1.25 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.17 BIOMEDICAL INTERVENTIONS AIMED AT SPEECH AND LANGUAGE

1.17.1 Complementary therapies for speech and language as a direct or indirect outcome

Acupuncture/acupressure and language therapy versus language therapy only for speech and language as a direct outcome

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 Acupuncture/Acupressure and language therapy versus language therapy only for the coexisting problem of speech and language as a direct outcome		Risk with Control	Risk difference with Acupuncture/Acupressure and language therapy versus language therapy only for the coexisting problem of speech and language as a direct outcome (95% CI)
Receptive semantics (measured with: Arabic Language Test: Receptive Semantics; Better indicated by lower values)											
20 (1 study) 39 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	10	10	N/A	N/A	The mean receptive semantics in the intervention groups was **0.66 standard deviations higher** (0.24 lower to 1.57 higher)
Expressive semantics (measured with: Arabic Language Test: Expressive Semantics; Better indicated by lower values)											
20 (1 study) 39 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	10	10	N/A	N/A	The mean expressive semantics in the intervention groups was **0.08 standard deviations lower** (0.96 lower to 0.79 higher)
Attention level (measured with: Arabic Language Test: Attention Level; Better indicated by lower values)											
20	serious[1]	no serious	no serious	very	undetected	⊕⊖⊖⊖	10	10	N/A	N/A	The mean attention level in the

(1 study) 39 weeks		inconsistency	indirectness	serious[2]		**VERY LOW**[1,2] due to risk of bias, imprecision				intervention groups was **0.36 standard deviations higher** (0.53 lower to 1.24 higher)
Positive treatment response for vocalization (assessed with: Dichotomous: Frequency of improvement in basic developmental assessment)										
30 (1 study) 39 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	2/14 (14.3%)	1/16 (6.3%)	**RR 0.44** (0.04 to 4.32)	**Study population** **143 per 1000** | **80 fewer per 1000** (from 137 fewer to 474 more) **Moderate** **143 per 1000** | **80 fewer per 1000** (from 137 fewer to 475 more)
Positive treatment response for babbling (assessed with: Dichotomous: Frequency of improvement in basic developmental assessment)										
30 (1 study) 39 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	4/14 (28.6%)	2/16 (12.5%)	**RR 0.44** (0.09 to 2.04)	**Study population** **286 per 1000** | **160 fewer per 1000** (from 260 fewer to 297 more) **Moderate** **286 per 1000** | **160 fewer per 1000** (from 260 fewer to 297 more)
Positive treatment response for speech (assessed with: Dichotomous: Frequency of improvement in basic developmental assessment)										
30 (1 study) 39 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias,	2/14 (14.3%)	8/16 (50%)	**RR 3.5** (0.89 to 13.82)	**Study population** **143 per 1000** | **357 more per 1000** (from 16 fewer to 1000 more)

						imprecision				**Moderate**	
										143 per 1000	**358 more per 1000** (from 16 fewer to 1000 more)
Positive treatment response for speech comprehension (assessed with: Dichotomous: Frequency of improvement in China Rehabilitation Research Council (CRRC) sign-significance relations scale)											
30 (1 study) 39 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	5/14 (35.7%)	5/16 (31.3%)	**RR 0.88** (0.32 to 2.4)	**Study population**	
										357 per 1000	**43 fewer per 1000** (from 243 fewer to 500 more)
										Moderate	
										357 per 1000	**43 fewer per 1000** (from 243 fewer to 500 more)
Positive treatment response for speech expression (assessed with: Dichotomous: Frequency of improvement in China Rehabilitation Research Council (CRRC) sign-significance relations scale)											
30 (1 study) 39 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	3/14 (21.4%)	4/16 (25%)	**RR 1.17** (0.31 to 4.34)	**Study population**	
										214 per 1000	**36 more per 1000** (from 148 fewer to 716 more)
										Moderate	
										214 per 1000	**36 more per 1000** (from 148 fewer to 715 more)
Positive treatment response for speech imitation (assessed with: Dichotomous: Frequency of improvement in China Rehabilitation Research Council (CRRC) sign-significance relations scale)											
30	serious[1]	no serious	no serious	very	undetected	⊕⊖⊖⊖	2/14	1/16	**RR 0.44**	**Study population**	

(1 study) 39 weeks		inconsistency	indirectness	serious[3]		**VERY LOW**[1,3] due to risk of bias, imprecision	(14.3%)	(6.3%)	(0.04 to 4.32)	**143 per 1000**	**80 fewer per 1000** (from 137 fewer to 474 more)
										Moderate	
										143 per 1000	**80 fewer per 1000** (from 137 fewer to 475 more)
Positive treatment response for vocabulary comprehension (assessed with: Dichotomous: Frequency of improvement in China Rehabilitation Research Council (CRRC) sign-significance relations scale)											
30 (1 study) 39 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	0/14 (0%)	5/16 (31.3%)	**RR 9.71** (0.58 to 161.31)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Positive treatment response for vocabulary expression (assessed with: Dichotomous: Frequency of improvement in China Rehabilitation Research Council (CRRC) sign-significance relations scale)											
30 (1 study) 39 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	0/14 (0%)	5/16 (31.3%)	**RR 9.71** (0.58 to 161.31)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Positive treatment response for phrase comprehension (assessed with: Dichotomous: Frequency of improvement in China Rehabilitation Research Council (CRRC) sign-significance relations scale)											
30	serious[1]	no serious	no serious	very	undetected	⊕⊖⊖⊖	0/14	1/16	**RR 2.65**	**Study population**	

(1 study) 39 weeks		inconsistency	indirectness	serious[3]		**VERY LOW**[1,3] due to risk of bias, imprecision	(0%)	(6.3%)	(0.12 to 60.21)	**0 per 1000**	N/A
										Moderate	
										0 per 1000	N/A
Positive treatment response for phrase expression (assessed with: Dichotomous: Frequency of improvement in China Rehabilitation Research Council (CRRC) sign-significance relations scale)											
30 (1 study) 39 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	0/14 (0%)	1/16 (6.3%)	**RR 2.65** (0.12 to 60.21)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Positive treatment response for communication attitude (assessed with: Dichotomous: Frequency of improvement in China Rehabilitation Research Council (CRRC) sign-significance relations scale)											
30 (1 study) 39 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[1,4] due to risk of bias, imprecision	8/14 (57.1%)	15/16 (93.8%)	**RR 1.64** (1.02 to 2.63)	**Study population**	
										571 per 1000	**366 more per 1000** (from 11 more to 931 more)
										Moderate	
										571 per 1000	**365 more per 1000** (from 11 more to 931 more)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear/unknown as identity and blinding of outcome assessors not reported and no independent reliability or validity data for this outcome measure
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)
[4] Events<300

Acupuncture/electro-acupuncture versus sham acupuncture/electro-acupuncture for speech and language as an indirect outcome

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 Acupuncture/Electro-acupuncture versus sham acupuncture/electro-acupuncture for speech and language as an indirect outcome**		**Risk with Control**	**Risk difference with Acupuncture/Electro-acupuncture versus sham acupuncture/electro-acupuncture for speech and language as an indirect outcome** (95% CI)
Receptive language (measured with: Reynell Developmental Language Scale (change score): Comprehension score; Better indicated by lower values)											
50 (1 study) 9 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	25	25	N/A	N/A	The mean receptive language in the intervention groups was **0.18 standard deviations lower** (0.73 lower to 0.38 higher)
Receptive language (measured with: Reynell Developmental Language Scale (change score): Comprehension age (years); Better indicated by lower values)											
105 (2 studies) 4-9 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊕⊖⊖ **LOW**[2,3] due to imprecision, publication bias	50	55	N/A	N/A	The mean receptive language in the intervention groups was **0.39 standard deviations higher** (0 to 0.78 higher)
Expressive language (measured with: Reynell Developmental Language Scale (change score): Expression score; Better indicated by lower values)											
50 (1 study) 9 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	25	25	N/A	N/A	The mean expressive language in the intervention groups was **0.42 standard deviations higher** (0.14 lower to 0.98 higher)

Expressive language (measured with: Reynell Developmental Language Scale (change score): Expression age (years); Better indicated by lower values)										
105 (2 studies) 4-9 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊕⊖⊖ **LOW**[2,3] due to imprecision, publication bias	50	55	N/A	N/A	The mean expressive language in the intervention groups was **0.11 standard deviations higher** (0.28 lower to 0.49 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] N<400
[3] High risk of selective reporting bias in WONG2010B as trial protocol includes a follow-up but no follow-up data reported

1.17.2 Hormones for speech and language as an indirect outcome

Secretin versus placebo for speech and language as an indirect outcome

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 Secretin versus placebo for speech and language as an indirect outcome**		**Risk with Control**	**Risk difference with Secretin versus placebo for speech and language as an indirect outcome** (95% CI)
Receptive language (measured with: Preschool Language Scale-3 (PLS-3): Auditory Comprehension (change score) or Mullen Scales of Early Learning (MSEL): Receptive Language or MSEL/PPVT-III: Receptive Language (months; change score); Better indicated by lower values)											
187 (3 studies) 3-6 weeks	no serious risk of bias	serious[1]	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to inconsistency, imprecision	96	91	N/A	N/A	The mean receptive language in the intervention groups was **0.02 standard deviations lower** (0.31 lower to 0.27 higher)
Expressive language (measured with: Preschool Language Scale-3 (PLS-3): Expressive Communication (change score) or Behavioural observation: Mean length of utterance (MLU) or Expressive One Word Picture Vocabulary Test-Revised (EOWPVT-R) Expressive language (change score); Better indicated by lower values)											
212 (3 studies)	no serious	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2]	100	112	N/A	N/A	The mean expressive language in the

3-6 weeks	risk of bias					due to imprecision					intervention groups was **0.16 standard deviations lower** (0.43 lower to 0.11 higher)
Receptive and expressive language (measured with: Preschool Language Scale-3 (PLS-3): Total (change score); Better indicated by lower values)											
85 (1 study) 3 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊕⊖⊖ **LOW**[3] due to imprecision	44	41	N/A	N/A	The mean receptive and expressive language in the intervention groups was **0.28 standard deviations higher** (0.15 lower to 0.71 higher)
Vocabulary (measured with: Behavioural observation: Type token ratio or MacArthur Communication Developmental Inventories (CDI): Vocabulary (change score); Better indicated by lower values)											
115 (2 studies) 4-6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	53	62	N/A	N/A	The mean vocabulary in the intervention groups was **0.06 standard deviations lower** (0.43 lower to 0.31 higher)
Positive treatment response (assessed with: Dichotomous: Positive treatment response (improvement >=4 points on PLS-3))											
95 (1 study) 3 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4] due to imprecision	10/48 (20.8%)	16/47 (34%)	**RR 1.63** (0.83 to 3.23)	**Study population**	
										208 per 1000	**131 more per 1000** (from 35 fewer to 465 more)
										Moderate	
										208 per 1000	**131 more per 1000** (from 35 fewer to 464 more)

[1] I-squared value indicates moderate heterogeneity
[2] N<400
[3] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[4] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)

1.17.3 Medical procedures for speech and language as an indirect outcome

Long-term chelation (7-rounds of DMSA therapy) versus short-term chelation (1-round of DMSA therapy and 6-rounds of placebo) for speech and language as an indirect outcome

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 Short-term chelation (1-round of DMSA therapy and 6-rounds of placebo)	**With Long-term chelation (7-rounds of Dimercaptosuccinic Acid [DMSA] therapy)**		**Risk with Short-term chelation (1-round of DMSA therapy and 6-rounds of placebo)**	**Risk difference with Long-term chelation (7-rounds of Dimercaptosuccinic Acid [DMSA] therapy)** (95% CI)
Semantic Pragmatic Problems (measured with: Pervasive Development Disorder Behavior Inventory (PDDBI): Semantic Pragmatic Problems; Better indicated by lower values)											
40 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⨁⨁⨀⨀ **LOW**[1] due to imprecision	15	25	N/A	N/A	The mean semantic pragmatic problems in the intervention groups was **0.44 standard deviations higher** (0.2 lower to 1.09 higher)
Expressive Language (measured with: Pervasive Development Disorder Behavior Inventory (PDDBI): Expressive Language; Better indicated by lower values)											
40 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⨁⨁⨀⨀ **LOW**[1] due to imprecision	15	25	N/A	N/A	The mean expressive language in the intervention groups was **0.26 standard deviations lower** (0.91 lower to 0.38 higher)

Learning, Memory, and Receptive Language (measured with: Pervasive Development Disorder Behavior Inventory (PDDBI): Learning, Memory, and Receptive Language; Better indicated by lower values)										
40 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	25	N/A	N/A	The mean learning, memory, and receptive language in the intervention groups was **0.12 standard deviations lower** (0.76 lower to 0.52 higher)
[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)										

HBOT versus attention-placebo for speech and language as an indirect outcome

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 Attention-placebo control	**With Hyperbaric oxygen treatment (HBOT)**		**Risk with Attention-placebo control**	**Risk difference with Hyperbaric oxygen treatment (HBOT)** (95% CI)
Receptive language (measured with: Peabody Picture Vocabulary Test, 3rd Ed. (PPVT-III): Total (change score); Better indicated by lower values)											
27 (1 study) 15 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	13	14	N/A	N/A	The mean receptive language in the intervention groups was **0.45 standard deviations lower** (1.22 lower to 0.31 higher)
[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)											

1.17.4 Nutritional interventions for speech and language as an indirect outcome

Omega-3 fatty acids versus placebo for speech and language as an indirect outcome

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 Placebo	**With Omega-3 fatty acids**		**Risk with Placebo**	**Risk difference with Omega-3 fatty acids** (95% CI)
Receptive language (measured with: Peabody Picture Vocabulary Test, 3rd Ed. (PPVT-III): Total; Better indicated by lower values)											
25 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	12	13	N/A	N/A	The mean receptive language in the intervention groups was **0.52 standard deviations lower** (1.32 lower to 0.28 higher)
Expressive language (measured with: Expressive Vocabulary Test (EVT): Total; Better indicated by lower values)											
25 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	12	13	N/A	N/A	The mean expressive language in the intervention groups was **0.69 standard deviations lower** (1.51 lower to 0.12 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Omega-3 fatty acids versus healthy diet control for speech and language as an indirect outcome

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 Healthy diet control	**With Omega-3 fatty acids**		**Risk with Healthy diet control**	**Risk difference with Omega-3 fatty acids** (95% CI)

Receptive language (measured with: Mullen Scales of Early Learning (MSEL): Receptive Language; Better indicated by lower values)										
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A	The mean receptive language in the intervention groups was **0.21 standard deviations higher** (0.61 lower to 1.04 higher)
Expressive language (measured with: Mullen Scales of Early Learning (MSEL): Expressive Language; Better indicated by lower values)										
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A	The mean expressive language in the intervention groups was **0.36 standard deviations higher** (0.47 lower to 1.19 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as the outcome assessor for this outcome measure was not blinded.
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Multivitamin/ mineral supplement versus placebo for speech and language as an indirect outcome

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 Placebo	**With Multivitamin and mineral supplement**		**Risk with Placebo**	**Risk difference with Multivitamin and mineral supplement** (95% CI)
Receptive language improvement (measured with: Parent Global Impressions-Revised (PGI-R): Receptive language improvement; Better indicated by lower values)											
104 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	51	53	N/A	N/A	The mean receptive language improvement in the intervention groups was **0.43 standard deviations higher** (0.04 to 0.82 higher)

Expressive language improvement (measured with: Parent Global Impressions-Revised (PGI-R): Expressive language improvement; Better indicated by lower values)										
104 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	51	53	N/A	N/A	The mean expressive language improvement in the intervention groups was **0.37 standard deviations higher** (0.02 lower to 0.76 higher)

[1] N<400
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

L-carnosine supplement versus placebo for speech and language as an indirect outcome

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 Placebo	**With L-carnosine supplement**		**Risk with Placebo**	**Risk difference with L-carnosine supplement** (95% CI)
Receptive language (measured with: Receptive One-Word Picture Vocabulary Test (ROWPVT) Receptive language (raw score); Better indicated by lower values)											
31 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	14	N/A	N/A	The mean receptive language in the intervention groups was **0.25 standard deviations higher** (0.46 lower to 0.96 higher)
Receptive language (measured with: Receptive One-Word Picture Vocabulary Test (ROWPVT) Receptive language (age adjusted score); Better indicated by lower values)											
31 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	14	N/A	N/A	The mean receptive language in the intervention groups was **0.2 standard deviations higher** (0.5 lower to 0.91 higher)
Expressive language (measured with: Expressive One Word Picture Vocabulary Test (EOWPVT) Expressive language (raw score); Better indicated by lower values)											

31 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	14	N/A	N/A	The mean expressive language in the intervention groups was **0.2 standard deviations higher** (0.51 lower to 0.91 higher)
Expressive language (measured with: Expressive One Word Picture Vocabulary Test (EOWPVT) Expressive language (age adjusted score); Better indicated by lower values)											
31 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	17	14	N/A	N/A	The mean expressive language in the intervention groups was **0.21 standard deviations higher** (0.5 lower to 0.92 higher)
[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)											

1.17.5 Sensory interventions for speech and language as an indirect outcome

Auditory integration training versus attention-placebo (structured listening) for speech and language as an indirect outcome

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 Attention-placebo (structured listening) control	**With Auditory integration training**		**Risk with Attention-placebo (structured listening) control**	**Risk difference with Auditory integration training** (95% CI)
Receptive language (measured with: Peabody Picture Vocabulary Test (PPVT): Total; Better indicated by lower values)											
80 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean receptive language in the intervention groups was **0.24 standard deviations lower**

											(0.68 lower to 0.2 higher)
Receptive language (measured with: Peabody Picture Vocabulary Test (PPVT): Total; Better indicated by lower values)											
80 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean receptive language in the intervention groups was **0.32 standard deviations lower** (0.76 lower to 0.12 higher)
Receptive language (measured with: Peabody Picture Vocabulary Test (PPVT): Total; Better indicated by lower values)											
80 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	40	40	N/A	N/A	The mean receptive language in the intervention groups was **0.5 standard deviations lower** (0.94 to 0.05 lower)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] N<400

Neurofeedback versus treatment-as-usual for speech and language as an indirect outcome

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 Treatment-as-usual	**With Neurofeedback**		**Risk with Treatment-as-usual**	**Risk difference with Neurofeedback** (95% CI)
Parent-rated speech production (measured with: Children's Communication Checklist (CCC-2): Speech production; Better indicated by lower values)											
20 (1 study)	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias strongly	⊕⊖⊖⊖ **VERY LOW**[1,2,3]	10	10	N/A	N/A	The mean parent-rated speech production in

20 weeks					suspected [3]	due to risk of bias, imprecision, publication bias					the intervention groups was **0.38 standard deviations lower** (1.26 lower to 0.51 higher)
Teacher-rated speech production (measured with: Children's Communication Checklist (CCC-2): Speech production; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated speech production in the intervention groups was **0.75 standard deviations higher** (0.16 lower to 1.67 higher)
Parent-rated syntax (measured with: Children's Communication Checklist (CCC-2): Syntax; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean parent-rated syntax in the intervention groups was **0.54 standard deviations lower** (1.44 lower to 0.35 higher)
Teacher-rated syntax (measured with: Children's Communication Checklist (CCC-2): Syntax; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated syntax in the intervention groups was **0.2 standard deviations higher** (0.68 lower to 1.08 higher)
Parent-rated semantics (measured with: Children's Communication Checklist (CCC-2): Semantics; Better indicated by lower values)											
20 (1 study)	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias strongly	⊕⊖⊖⊖ **VERY LOW**[1,2,3]	10	10	N/A	N/A	The mean parent-rated semantics in the

20 weeks					suspected [3]	due to risk of bias, imprecision, publication bias					intervention groups was **0.89 standard deviations lower** (1.82 lower to 0.04 higher)
Teacher-rated semantics (measured with: Children's Communication Checklist (CCC-2): Semantics; Better indicated by lower values)											
20 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	10	10	N/A	N/A	The mean teacher-rated semantics in the intervention groups was **1.12 standard deviations higher** (0.17 to 2.08 higher)
Parent-rated coherence (measured with: Children's Communication Checklist (CCC-2): Coherence; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean parent-rated coherence in the intervention groups was **0.68 standard deviations lower** (1.59 lower to 0.23 higher)
Teacher-rated coherence (measured with: Children's Communication Checklist (CCC-2): Coherence; Better indicated by lower values)											
20 (1 study) 20 weeks	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	10	10	N/A	N/A	The mean teacher-rated coherence in the intervention groups was **0.89 standard deviations higher** (0.04 lower to 1.82 higher)

[1] High risk of performance, response and detection bias as intervention administrators, participants and outcome assessors were non-blind. The risk of other bias due to potential conflict of interest is also high as neurofeedback equipment provided by manufacturer for trial.
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] High risk of selective reporting bias as data cannot be extracted for 6-month follow-up
[4] N<400

1.18 PSYCHOSOCIAL INTERVENTIONS AIMED AT IQ AND ACADEMIC SKILLS

1.18.1 Behavioural interventions for IQ and/or academic skills as a direct or indirect outcome

EIBI or EBI (ESDM or P-ESDM) versus treatment-as-usual for IQ as a direct or indirect outcome

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 Treatment-as-usual	**With ESDM or P-ESDM**		**Risk with Treatment-as-usual**	**Risk difference with ESDM or P-ESDM** (95% CI)
IQ (ESDM or P-ESDM) (measured with: Mullen Scales of Early Learning (MSEL): Early learning composite score or developmental quotient; Better indicated by lower values)											
143 (2 studies) 12-104 weeks	serious[1]	serious[2]	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, inconsistency, imprecision	70	73	N/A	N/A	The mean iq (esdm or p-esdm) in the intervention groups was **0.25 standard deviations higher** (0.08 lower to 0.58 higher)
Verbal developmental quotient (P-ESDM) (measured with: Mullen Scales of Early Learning (MSEL): Verbal developmental quotient; Better indicated by lower values)											
98 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[1,4] due to risk of bias, imprecision	49	49	N/A	N/A	The mean verbal developmental quotient (p-esdm) in the intervention groups was **0.1 standard deviations higher** (0.3 lower to 0.5 higher)
Non-verbal developmental quotient (P-ESDM) (measured with: Mullen Scales of Early Learning (MSEL): Non-verbal developmental quotient; Better indicated by lower values)											
98 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[1,4] due to risk of bias,	49	49	N/A	N/A	The mean non-verbal developmental quotient (p-esdm) in the intervention

						imprecision				groups was **0.08 standard deviations higher** (0.31 lower to 0.48 higher)

[1] High risk of performance and response bias as intervention administrators and participants were nonblind, and risk of detection bias is unclear/unknown as identity and blinding of outcome assessors not reported
[2] I-squared value indicates moderate heterogeneity
[3] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[4] N<400

EIBI versus parent training for IQ and academic skills as an indirect outcome

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 EIBI versus parent training for IQ as an indirect outcome**		**Risk with Control**	**Risk difference with EIBI versus parent training for IQ as an indirect outcome** (95% CI)
IQ (measured with: Bayley Scales of Infant Development: Mental Development Index; Better indicated by lower values)											
28 (1 study) 260 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	13	15	N/A	N/A	The mean iq in the intervention groups was **0.74 standard deviations higher** (0.04 lower to 1.51 higher)
Academic skills (measured with: Wechsler Individualized Achievement Test (WIAT): Total; Better indicated by lower values)											
28 (1 study) 260 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	13	15	N/A	N/A	The mean academic skills in the intervention groups was **0.84 standard deviations higher** (0.06 to 1.62 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

[2] N<400

1.18.2 Educational interventions for IQ as an indirect outcome

LEAP training versus manual-only control for IQ as an indirect outcome

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 Intervention-manual-only control	With Inclusive educational intervention (LEAP) training		Risk with Intervention-manual-only control	Risk difference with Inclusive educational intervention (LEAP) training (95% CI)
IQ (measured with: Mullen Scales of Early Learning (MSEL): Early-learning composite score; Better indicated by lower values)											
294 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	117	177	N/A	N/A	The mean iq in the intervention groups was **0.87 standard deviations higher** (0.63 to 1.12 higher)

[1] High risk of performance and response bias as intervention administrators and participants non-blind. In addition, risk of detection bias is unclear/unknown as identity and blinding of outcome assessors not reported
[2] N<400

1.18.3 Parent training for IQ as an indirect outcome

Parent training versus treatment-as-usual for IQ as an indirect outcome

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 Parent training versus treatment-as-usual for IQ		Risk with Control	Risk difference with Parent training versus treatment-as-usual for IQ (95% CI)

IQ (measured with: Griffiths Scale of Mental Development: D and E scales (NVIQ NVMA/age) or Psychoeducational Profile-Revised (PEP-R): Developmental Quotient (DQ) or Mullen Scales of Early Learning (MSEL): Developmental quotient; Better indicated by lower values)										
147 (3 studies) 12-52 weeks	no serious risk of bias	serious[1]	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to inconsistency, imprecision	57	90	N/A	N/A	The mean iq in the intervention groups was **0.04 standard deviations higher** (0.3 lower to 0.38 higher)
[1] I-squared value indicates moderate heterogeneity [2] N<400										

Combined parent training and early intervention centre programme versus early intervention centre programme only for IQ as an indirect outcome

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 Early intervention centre programme only	**With Combined parent training and early intervention centre programme**		**Risk with Early intervention centre programme only**	**Risk difference with Combined parent training and early intervention centre programme** (95% CI)
IQ (mixed ASD & DD sample) (measured with: Bayley Scales of Infant Development-Second Edition or Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R); Better indicated by lower values)											
59 (1 study) 40 weeks	no serious risk of bias	no serious inconsistency	serious[1]	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to indirectness, imprecision	29	30	N/A	N/A	The mean iq (mixed asd & dd sample) in the intervention groups was **0.35 standard deviations higher** (0.17 lower to 0.86 higher)
IQ (ASD-only sample) (measured with: Bayley Scales of Infant Development-Second Edition or Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R); Better indicated by lower values)											
39	no	no serious	no serious	very	undetected	⊕⊕⊖⊖	21	18	N/A	N/A	The mean iq (asd-

(1 study) 40 weeks	serious risk of bias	inconsistency	indirectness	serious[2]		**LOW**[2] due to imprecision				only sample) in the intervention groups was **0.43 standard deviations higher** (0.21 lower to 1.07 higher)
IQ (mixed ASD & DD sample) (measured with: Bayley Scales of Infant Development-Second Edition or Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R); Better indicated by lower values)										
54 (1 study) 108 weeks	no serious risk of bias	no serious inconsistency	serious[1]	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to indirectness, imprecision	26	28	N/A	N/A	The mean iq (mixed asd & dd sample) in the intervention groups was **0.37 standard deviations higher** (0.17 lower to 0.91 higher)

[1] Population was indirect (as the sample included participants with developmental delay or language delay without autism)
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.18.4 Social-communication interventions for IQ as an indirect outcome

Caregiver-mediated social communication intervention versus treatment-as-usual for IQ as an indirect outcome

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 Reciprocal social-communication interventions versus treatment-as-usual for IQ as an indirect outcome**		**Risk with Control**	**Risk difference with Reciprocal social-communication interventions versus treatment-as-usual for IQ as an indirect outcome** (95% CI)
IQ (measured with: Mullen Scales of Early Learning (MSEL): Early-learning composite score; Better indicated by lower values)											

49 (1 study) 39 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	25	24	N/A	N/A	The mean iq in the intervention groups was **0.06 standard deviations lower** (0.62 lower to 0.5 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear/unknown as identity and blinding of outcome assessors is not reported
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Joint attention training and EIBI versus EIBI only for IQ as an indirect outcome

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 Combined joint attention training and EIBI versus EIBI only for IQ as an indirect outcome**		**Risk with Control**	**Risk difference with Combined joint attention training and EIBI versus EIBI only for IQ as an indirect outcome** (95% CI)
IQ (measured with: Mullen Scales of Early Learning (MSEL): Developmental quotient; Better indicated by lower values)											
36 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	16	20	N/A	N/A	The mean iq in the intervention groups was **0.54 standard deviations higher** (0.13 lower to 1.21 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.19 PHARMACOLOGICAL INTERVENTIONS AIMED AT ACADEMIC SKILLS

1.19.1 Antipsychotics for academic skills as an indirect outcome

Risperidone versus placebo for academic skills as an indirect outcome

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 Antipsychotics versus placebo for academic skills		Risk with Control	Risk difference with Antipsychotics versus placebo for academic skills (95% CI)
Maths problem-solving (measured with: Classroom Analogue Task: Total number of maths problems correctly calculated; Better indicated by lower values)											
38 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	18	20	N/A	N/A	The mean maths problem-solving in the intervention groups was **0.45 standard deviations lower** (1.1 lower to 0.19 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.20 BIOMEDICAL INTERVENTIONS AIMED AT IQ

1.20.1 Complementary therapies for IQ as a direct outcome

Acupuncture/electro-acupuncture versus sham acupuncture/electro-acupuncture for IQ as a direct outcome

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 Acupuncture/Electro-acupuncture versus sham acupuncture/electro-acupuncture for IQ as a direct outcome		Risk with Control	Risk difference with Acupuncture/Electro-acupuncture versus sham acupuncture/electro-acupuncture for IQ as a direct outcome (95% CI)
General quotient/FIQ (measured with: Griffiths Mental Development Scale (change score): General Quotient or Leiter International Performance Scale-Revised: Visualization and Reasoning: Battery (IQ/Composite Score) (change score); Better indicated by lower values)											
105	no	no serious	no serious	very	reporting bias	⊕⊖⊖⊖	50	55	N/A	N/A	The mean general

(2 studies) 4-9 weeks	serious risk of bias	inconsistency	indirectness	serious[1]	strongly suspected [2]	**VERY LOW**[1,2] due to imprecision, publication bias				quotient/fiq in the intervention groups was **0.23 standard deviations higher** (0.15 lower to 0.62 higher)
Mental age in months (measured with: Griffiths Mental Development Scale (change score): Mental age (months); Better indicated by lower values)										
50 (1 study) 9 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	25	25	N/A	N/A	The mean mental age in months in the intervention groups was **0.43 standard deviations higher** (0.13 lower to 0.99 higher)
Locomotor (measured with: Griffiths Mental Developmental Scale: Locomotor (change score); Better indicated by lower values)										
50 (1 study) 9 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	25	25	N/A	N/A	The mean locomotor in the intervention groups was **0.2 standard deviations lower** (0.76 lower to 0.35 higher)
Personal-Social (measured with: Griffiths Mental Developmental Scale: Personal-Social (change score); Better indicated by lower values)										
50 (1 study) 9 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	25	25	N/A	N/A	The mean personal-social in the intervention groups was **0.53 standard deviations higher** (0.03 lower to 1.1 higher)
Hearing and speech (measured with: Griffiths Mental Developmental Scale: Hearing & Speech (change score); Better indicated by lower values)										
50 (1 study) 9 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	25	25	N/A	N/A	The mean hearing and speech in the intervention groups was **0.15 standard deviations higher** (0.4 lower to 0.71 higher)
Eye and hand coordination (measured with: Griffiths Mental Developmental Scale: Eye & Hand Coordination (change score); Better indicated by lower values)										

50 (1 study) 9 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	25	25	N/A	N/A	The mean eye and hand coordination in the intervention groups was **0.12 standard deviations higher** (0.44 lower to 0.67 higher)
Performance (measured with: Griffiths Mental Developmental Scale: Performance (change score); Better indicated by lower values)										
50 (1 study) 9 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	25	25	N/A	N/A	The mean performance in the intervention groups was **0.41 standard deviations higher** (0.15 lower to 0.97 higher)
Practical reasoning (measured with: Griffiths Mental Developmental Scale: Practical Reasoning (change score); Better indicated by lower values)										
50 (1 study) 9 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	25	25	N/A	N/A	The mean practical reasoning in the intervention groups was **0.32 standard deviations higher** (0.23 lower to 0.88 higher)
Attention and memory (measured with: Leiter International Performance Scale-Revised: Attention and Memory: Battery (Composite Score); Better indicated by lower values)										
55 (1 study) 4 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	25	30	N/A	N/A	The mean attention and memory in the intervention groups was **0.04 standard deviations lower** (0.57 lower to 0.49 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] High risk of selective reporting bias as trial protocol for WONG2010B states that follow-up measurements will be taken but these are not reported

1.20.2 Hormones for IQ as an indirect outcome

Secretin versus placebo for IQ as an indirect outcome

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 Secretin versus placebo for IQ as an indirect outcome		Risk with Control	Risk difference with Secretin versus placebo for IQ as an indirect outcome (95% CI)
IQ (measured with: Merrill-Palmer Scale; Better indicated by lower values)											
42 (1 study) 6 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	23	19	N/A	N/A	The mean iq in the intervention groups was **0.31 standard deviations lower** (0.92 lower to 0.3 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.20.3 Nutritional intervention for IQ as an indirect outcome

Multivitamin/ mineral supplement versus placebo for IQ as an indirect outcome

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 Placebo	With Multivitamin and mineral supplement		Risk with Placebo	Risk difference with Multivitamin and mineral supplement (95% CI)
Cognition improvement (measured with: Parent Global Impressions-Revised (PGI-R): Cognition improvement; Better indicated by lower values)											
104 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	51	53	N/A	N/A	The mean cognition improvement in the intervention groups was **0.32 standard deviations higher** (0.06 lower to 0.71 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.20.4 Sensory intervention for IQ as an indirect outcome

Auditory integration training versus attention-placebo (structured listening) for IQ as an indirect outcome

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 Attention-placebo (structured listening) control	With Auditory integration training		Risk with Attention-placebo (structured listening) control	Risk difference with Auditory integration training (95% CI)
PIQ (measured with: Leiter International Performance Scale (LIPS): Total; Better indicated by lower values)											
80 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean piq in the intervention groups was **0.16 standard deviations lower** (0.6 lower to 0.28 higher)
PIQ (measured with: Leiter International Performance Scale (LIPS): Total; Better indicated by lower values)											
80 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean piq in the intervention groups was **0.17 standard deviations lower** (0.61 lower to 0.26 higher)
PIQ (measured with: Leiter International Performance Scale (LIPS): Total; Better indicated by lower values)											
80 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean piq in the intervention groups was **0.22 standard deviations lower** (0.66 lower to 0.22 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.21 PYCHOSOCIAL INTERVENTIONS AIMED AT SENSORY SENSITIVITIES

1.21.1 Animal-based interventions for sensory sensitivities as an indirect outcome

Horseback riding versus waitlist control for sensory sensitivities as an indirect outcome

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 Waitlist control	**With Horseback riding**		**Risk with Waitlist control**	**Risk difference with Horseback riding** (95% CI)
Sensory problems (measured with: Sensory Profile: Total; Better indicated by lower values)											
34 (1 study) 12 weeks	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	15	19	N/A	N/A	The mean sensory problems in the intervention groups was **0.45 standard deviations higher** (0.23 lower to 1.14 higher)
Sensory seeking (measured with: Sensory Profile: Sensory seeking; Better indicated by lower values)											
34 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	15	19	N/A	N/A	The mean sensory seeking in the intervention groups was **0.89 standard deviations higher** (0.17 to 1.6 higher)
Sensory sensitivity (measured with: Sensory Profile: Sensory sensitivity; Better indicated by lower values)											
34	serious[1]	no serious	no serious	very	reporting bias	⊕⊖⊖⊖	15	19	N/A	N/A	The mean sensory

(1 study) 12 weeks		inconsistency	indirectness	serious[2]	strongly suspected [3]	**VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias				sensitivity in the intervention groups was **0.39 standard deviations higher** (0.29 lower to 1.08 higher)

[1] High risk of performance and response bias as intervention administrators and participants non-blind. There is also a high risk of detection bias as outcome measures are parent-rated and parents non-blind
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] High risk of selective reporting bias as data not reported for selected subscales: low endurance/tone, oral sensory sensitivity, and poor registration subscales of the Sensory Profile scale
[4] N<400

1.21.2 Educational interventions for sensory sensitivities as an indirect outcome

Combined TeachTown and IBI versus IBI-only for sensory sensitivities as an indirect outcome

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 IBI-only	**With Combined computer-assisted educational intervention and intensive behavioural intervention (IBI) day class program**		**Risk with IBI-only**	**Risk difference with Combined computer-assisted educational intervention and intensive behavioural intervention (IBI) day class program** (95% CI)
Auditory processing (measured with: Brigance Inventory of Child Development: Auditory processing; Better indicated by lower values)											
46 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	24	22	N/A	N/A	The mean auditory processing in the intervention groups was **0.21 standard deviations higher** (0.37 lower to 0.79 higher)
Auditory processing (preschool subgroup analysis) (measured with: Brigance Inventory of Child Development: Auditory processing; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of	12	11	N/A	N/A	The mean auditory processing (preschool subgroup analysis) in the intervention groups was

						bias, imprecision				**0.13 standard deviations higher** (0.69 lower to 0.95 higher)
Auditory processing (K-1 subgroup analysis) (measured with: Brigance Inventory of Child Development: Auditory processing; Better indicated by lower values)										
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	12 11	N/A	N/A	The mean auditory processing (k-1 subgroup analysis) in the intervention groups was **0.29 standard deviations higher** (0.54 lower to 1.11 higher)

[1] High risk of performance and response bias as intervention administrators and participants non-blind. Risk of detection bias is unclear/unknown as the identity and blinding of outcome assessors not reported. In addition, for the Brigance Inventory of Child Development scale there are no independent reliability and/or validity data reported
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.22 BIOMEDICAL INTERVENTIONS AIMED AT SENSORY SENSITIVITIES

1.22.1 Complementary therapies for sensory sensitivities as a direct outcome

Qigong massage training versus waitlist for sensory sensitivities as a direct outcome

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 Qigong massage versus waitlist for the coexisting problem or disorder of sensory sensitivities as a direct outcome**		**Risk with Control**	**Risk difference with Qigong massage versus waitlist for the coexisting problem or disorder of sensory sensitivities as a direct outcome** (95% CI)
Sensory impairment (measured with: Pervasive Development Disorder Behavior Inventory (PDDBI): Sensory; Better indicated by lower values)										
79 (2 studies) 17-22 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias,	39	40	N/A	N/A	The mean sensory impairment in the intervention groups was **0.8 standard deviations**

						imprecision				**lower** (1.27 to 0.34 lower)
Sensory impairment (measured with: Sense and Self-Regulation Checklist (SSC): Sense checklist; Better indicated by lower values)										
87 (2 studies) 17-22 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to risk of bias, imprecision	39 48	N/A	N/A	The mean sensory impairment in the intervention groups was **1.11 standard deviations lower** (1.56 to 0.65 lower)

[1] High risk of selection bias in SILVA2009 as although groups were assigned using a random number generator, there were caveats to the randomisation (five sets of siblings were co-assigned due to parental involvement in the treatment and different geographical areas were assigned separately to meet the 'therapist to participant requirements'). Groups were also not comparable at baseline for measures of parent rated social communication and autism composite and teacher rated sensory problems. There was also a high risk of performance and response bias as intervention administrators and participants were non-blind, and an unclear or high risk of detection bias due to unclear blinding or non-blind outcome assessment
[2] N<400

1.22.2 Sensory interventions for sensory sensitivities as a direct outcome

Auditory integration training versus attention-placebo (structured listening) for sensory sensitivities as a direct outcome

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 Attention-placebo (structured listening) control	**With Auditory integration training**		**Risk with Attention-placebo (structured listening) control**	**Risk difference with Auditory integration training** (95% CI)
Sound sensitivity (measured with: Sound Sensitivity Questionnaire (SSQ): Total; Better indicated by lower values)											
80 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⨁⨁⨀⨀ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean sound sensitivity in the intervention groups was **0.27 standard deviations lower**

											(0.71 lower to 0.17 higher)
Sound sensitivity (measured with: Sound Sensitivity Questionnaire (SSQ): Total; Better indicated by lower values)											
80 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean sound sensitivity in the intervention groups was **0.13 standard deviations lower** (0.57 lower to 0.31 higher)
Sound sensitivity (measured with: Sound Sensitivity Questionnaire (SSQ): Total; Better indicated by lower values)											
80 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean sound sensitivity in the intervention groups was **0.12 standard deviations higher** (0.32 lower to 0.56 higher)
Sound sensitivity (measured with: Sound Sensitivity Questionnaire (SSQ): Total; Better indicated by lower values)											
80 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean sound sensitivity in the intervention groups was **0.2 standard deviations higher** (0.24 lower to 0.64 higher)
Sound distress (measured with: Sound Sensitivity Questionnaire (SSQ): Sound distress; Better indicated by lower values)											
80 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	40	40	N/A	N/A	The mean sound distress in the intervention groups was

											0.02 standard deviations lower (0.46 lower to 0.41 higher)
Sound distress (measured with: Sound Sensitivity Questionnaire (SSQ): Sound distress; Better indicated by lower values)											
80 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	40	40	N/A	N/A	The mean sound distress in the intervention groups was **0 standard deviations higher** (0.44 lower to 0.44 higher)
Sound distress (measured with: Sound Sensitivity Questionnaire (SSQ): Sound distress; Better indicated by lower values)											
80 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean sound distress in the intervention groups was **0.43 standard deviations higher** (0.01 lower to 0.87 higher)
Sound distress (measured with: Sound Sensitivity Questionnaire (SSQ): Sound distress; Better indicated by lower values)											
80 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean sound distress in the intervention groups was **0.2 standard deviations higher** (0.24 lower to 0.63 higher)
Sensory self-stimulation (measured with: Sensory Problems Checklist (SP): Total; Better indicated by lower values)											
80 (1 study)	no serious	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1]	40	40	N/A	N/A	The mean sensory self-stimulation in

4 weeks	risk of bias					due to imprecision					the intervention groups was **0.07 standard deviations higher** (0.36 lower to 0.51 higher)
Sensory self-stimulation (measured with: Sensory Problems Checklist (SP): Total; Better indicated by lower values)											
80 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean sensory self-stimulation in the intervention groups was **0.1 standard deviations higher** (0.34 lower to 0.54 higher)
Sensory self-stimulation (measured with: Sensory Problems Checklist (SP): Total; Better indicated by lower values)											
80 (1 study) 26 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	40	40	N/A	N/A	The mean sensory self-stimulation in the intervention groups was **0.05 standard deviations higher** (0.39 lower to 0.49 higher)
Sensory self-stimulation (measured with: Sensory Problems Checklist (SP): Total; Better indicated by lower values)											
80 (1 study) 52 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	40	40	N/A	N/A	The mean sensory self-stimulation in the intervention groups was **0.22 standard deviations higher** (0.22 lower to 0.66 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] N<400

Sensory integration therapy versus treatment-as-usual for sensory sensitivities as a direct outcome

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 Treatment-as-usual	With Sensory integration therapy		Risk with Treatment-as-usual	Risk difference with Sensory integration therapy (95% CI)
Sensory problems (measured with: Sensory Evaluation Form for Children with Autism: Total; Better indicated by lower values)											
30 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	15	15	N/A	N/A	The mean sensory problems in the intervention groups was **2 standard deviations lower** (2.9 to 1.11 lower)

[1] High risk of performance and response bias as intervention administrators and participants non-blind, and risk of detection bias is unclear/unknown as the identity and blinding of outcome assessor is not reported
[2] N<400

1.23 PSYCHOSOCIAL INTERVENTIONS AIMED AT MOTOR SKILLS

1.23.1 Animal-based interventions for motor skills as an indirect outcome

Horseback riding versus waitlist control for motor skills as an indirect outcome

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 Waitlist control	With Horseback riding		Risk with Waitlist control	Risk difference with Horseback riding (95% CI)
Fine motor/perception (measured with: Sensory Profile: Fine motor/perception; Better indicated by lower values)											

34 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	15	19	N/A	N/A	The mean fine motor/perception in the intervention groups was **0.22 standard deviations higher** (0.45 lower to 0.9 higher)

[1] High risk of performance and response bias as intervention administrators and participants non-blind. There is also a high risk of detection bias as outcome measures are parent-rated and parents non-blind
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.23.2 Behavioural interventions for motor skills as an indirect outcome

EIBI (ESDM) versus treatment-as-usual for motor skills as an indirect outcome

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 Behaviour-focused intervention versus treatment-as-usual for fine and gross motor skills as an indirect outcome**		**Risk with Control**	**Risk difference with Behaviour-focused intervention versus treatment-as-usual for fine and gross motor skills as an indirect outcome** (95% CI)
Fine motor skills (measured with: Mullen Scales of Early Learning (MSEL): Fine Motor; Better indicated by lower values)											
45 (1 study) 104 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	21	24	N/A	N/A	The mean fine motor skills in the intervention groups was **0.45 standard deviations higher** (0.15 lower to 1.04 higher)
Motor skills (measured with: Vineland Adaptive Behaviour Scale (VABS): Motor Skills; Better indicated by lower values)											
45 (1 study) 104 weeks	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊖⊖ **LOW**[2,3] due to risk of bias, imprecision	21	24	N/A	N/A	The mean motor skills in the intervention groups was **0.78 standard deviations higher** (0.17 to 1.39 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] High risk of performance and response bias as intervention administrators and participants were non-blind and risk of detection bias is unclear/unknown as although outcome assessors were blinded the outcome measure was based on interview with (non-blind) parent rather than direct observation
[3] N<400

1.23.3 Educational interventions for motor skills as an indirect outcome

LEAP training versus manual-only control for motor skills as an indirect outcome

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 Intervention-manual-only control	**With Inclusive educational intervention (LEAP) training**		**Risk with Intervention-manual-only control**	**Risk difference with Inclusive educational intervention (LEAP) training** (95% CI)
Fine motor skills (measured with: Mullen Scales of Early Learning (MSEL): Fine Motor Age (months); Better indicated by lower values)											
294 (1 study) 104 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⨁⨁⨀⨀ **LOW**[1,2] due to risk of bias, imprecision	117	177	N/A	N/A	The mean fine motor skills in the intervention groups was **0.69 standard deviations higher** (0.45 to 0.93 higher)

[1] High risk of performance and response bias as intervention administrators and participants non-blind. In addition, risk of detection bias is unclear/unknown as identity and blinding of outcome assessors not reported
[2] N<400

1.23.4 Parent training for motor skills as an indirect outcome

Parent training versus treatment-as-usual for motor skills as an indirect outcome

Quality assessment							Summary of Findings		
Participants	**Risk of**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication**	**Overall**	**Study event rates (%)**	**Relative**	**Anticipated absolute effects**

(studies) Follow up	bias				bias	quality of evidence	With Control	With Parent training versus treatment-as-usual for motor skills as an indirect outcome	effect (95% CI)	Risk with Control	Risk difference with Parent training versus treatment-as-usual for motor skills as an indirect outcome (95% CI)
Motor skills (PEC+PEBM combined) (measured with: Vineland Adaptive Behaviour Scale (VABS): Motor Skills; Better indicated by lower values)											
103 (1 study) 46 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	35	68	N/A	N/A	The mean motor skills (pec+pebm combined) in the intervention groups was **0.11 standard deviations higher** (0.3 lower to 0.52 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias is unclear/unknown as although the study included a blinded clinician outcome assessor this outcome measure was based on parental interview and simultaneous child observation and parents non-blind
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Parent and day-care staff training versus standard day-care for motor skills as an indirect outcome

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 Parent and day-care staff training versus standard day-care for fine and gross motor skills as an indirect outcome		Risk with Control	Risk difference with Parent and day-care staff training versus standard day-care for fine and gross motor skills as an indirect outcome (95% CI)
Fine motor skills (measured with: Early Intervention Developmental Profile (EIDP)/Preschool Developmental Profile (PSDP): Perceptual/Fine Motor; Better indicated by lower values)											
35 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	19	16	N/A	N/A	The mean fine motor skills in the intervention groups was **0.01 standard deviations higher** (0.66 lower to 0.67 higher)
Gross motor skills (measured with: Early Intervention Developmental Profile (EIDP)/Preschool Developmental Profile (PSDP): Gross Motor; Better indicated by lower values)											
35	no	no serious	no serious	very	undetected	⊕⊕⊖⊖	19	16	N/A	N/A	The mean gross motor skills

(1 study) 12 weeks	serious risk of bias	inconsistency	indirectness	serious[1]		**LOW**[1] due to imprecision				in the intervention groups was **0.18 standard deviations lower** (0.85 lower to 0.48 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.23.5 Social-communication interventions for motor skills as an indirect outcome

Caregiver-mediated social-communication intervention versus treatment-as-usual for motor skills as an indirect outcome

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 Reciprocal social-communication interventions versus treatment-as-usual for fine and gross motor skills as an indirect outcome**		**Risk with Control**	**Risk difference with Reciprocal social-communication interventions versus treatment-as-usual for fine and gross motor skills as an indirect outcome** (95% CI)
Fine motor skills (measured with: Mullen Scales of Early Learning (MSEL): Fine Motor Age (months); Better indicated by lower values)											
50 (1 study) 39 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	25	25	N/A	N/A	The mean fine motor skills in the intervention groups was **0.02 standard deviations higher** (0.53 lower to 0.58 higher)
Motor skills (measured with: Vineland Adaptive Behaviour Scale (VABS): Motor Skills; Better indicated by lower values)											
39 (1 study) 39 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	20	19	N/A	N/A	The mean motor skills in the intervention groups was **0.19 standard deviations higher** (0.44 lower to 0.82 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias unclear/unknown as identity and blinding of outcome assessors not reported
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] High risk of performance and response bias as intervention administrators and participants non-blind, and risk of detection bias unclear/unknown as outcome measure based on parent interview rather than direct behaviour observation and parents non-blind and involved in the intervention

1.24 BIOMEDICAL INTERVENTIONS AIMED AT MOTOR SKILLS

1.24.1 Hormones for motor skills as an indirect outcome

Secretin versus placebo for motor skills as an indirect outcome

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 Secretin versus placebo for fine and gross motor skills as an indirect outcome		Risk with Control	Risk difference with Secretin versus placebo for fine and gross motor skills as an indirect outcome (95% CI)
Fine motor skills (measured with: Mullen/DTVP-2: Fine motor (months); Better indicated by lower values)											
56 (1 study) 4 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	28	28	N/A	N/A	The mean fine motor skills in the intervention groups was **0.04 standard deviations lower** (0.57 lower to 0.48 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.24.2 Nutritional interventions for motor skills as an indirect outcome

Omega-3 fatty acids versus healthy diet control for motor skills as an indirect outcome

Quality assessment							Summary of Findings		
Participants	Risk of	Inconsistency	Indirectness	Imprecision	Publication	Overall quality	Study event rates (%)	Relative	Anticipated absolute effects

(studies) Follow up	bias				bias	of evidence	With Healthy diet control	With Omega-3 fatty acids	effect (95% CI)	Risk with Healthy diet control	Risk difference with Omega-3 fatty acids (95% CI)
Fine motor (measured with: Mullen Scales of Early Learning (MSEL): Fine Motor; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A	The mean fine motor in the intervention groups was **0.03 standard deviations lower** (0.86 lower to 0.79 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as the outcome assessor for this outcome measure was not blinded.
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Gluten-free and casein-free diet versus treatment-as-usual for motor skills as an indirect outcome

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 Treatment-as-usual	With Gluten-free and casein-free diet		Risk with Treatment-as-usual	Risk difference with Gluten-free and casein-free diet (95% CI)
Motor impairment (measured with: Movement Assessment Battery for Children: Test of Motor Impairment (TOMI); Better indicated by lower values)											
20 (1 study) 52 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	10	10	N/A	N/A	The mean motor impairment in the intervention groups was **0.12 standard deviations lower** (1 lower to 0.76 higher)

[1] High risk of performance and response bias as intervention administrators (parents) and participants were non-blind and unclear/unknown risk of detection bias as identity and blinding of outcome assessors not reported
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.25 PSYCHOSOCIAL INTERVENTIONS AIMED AT COEXISTING MENTAL HEALTH PROBLEMS

1.25.1 Cognitive-behavioural interventions for anxiety as a direct outcome

CBT versus treatment-as-usual for anxiety as a direct outcome

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 Waitlist or treatment-as-usual	**With CBT**		**Risk with Waitlist or treatment-as-usual**	**Risk difference with CBT** (95% CI)
No longer meet anxiety disorder diagnosis (assessed with: Number of participants who no longer met DSM-IV criteria for anxiety disorder)											
87 (2 studies) 16-24 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	2/42 (4.8%)	29/45 (64.4%)	**RR 11.82** (3.14 to 44.5)	**Study population**	
										48 per 1000	**515 more per 1000** (from 102 more to 1000 more)
										Moderate	
										44 per 1000	**476 more per 1000** (from 94 more to 1000 more)
Improvement in anxiety symptoms (assessed with: Clinical global impressions scale: Improvement ratings)											
83	no serious	no serious	no serious	serious[1]	undetected	⊕⊕⊕⊖	4/46	23/37	**RR 7.2**	**Study population**	

(2 studies) 16-24 weeks	risk of bias	inconsistency	indirectness			**MODERATE**[1] due to imprecision	(8.7%)	(62.2%)	(2.74 to 18.91)	**87 per 1000**	**539 more per 1000** (from 151 more to 1000 more)
										Moderate	
										87 per 1000	**539 more per 1000** (from 151 more to 1000 more)
Self-reported anxiety (measured with: Spence Childrens Anxiety Scale (SCAS) or Multidimensional Anxiety Scale for Children (MASC): Child version; Better indicated by lower values)											
83 (2 studies) 16-24 weeks	serious[2]	very serious[3]	no serious indirectness	serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3,4] due to risk of bias, inconsistency, imprecision	41	42	N/A	N/A	The mean self-reported anxiety in the intervention groups was **1.06 standard deviations lower** (1.58 to 0.55 lower)
Parent-reported anxiety (measured with: Spence Childrens Anxiety Scale: Parent Version (SCAS-P) or Multidimensional Anxiety Scale for Children (MASC): Parent version; Better indicated by lower values)											
149 (3 studies) 6-24 weeks	serious[5]	very serious[3]	no serious indirectness	serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[3,4,5] due to risk of bias, inconsistency, imprecision	61	88	N/A	N/A	The mean parent-reported anxiety in the intervention groups was **0.99 standard deviations lower** (1.39 to 0.6 lower)
Clinician-rated anxiety (measured with: Anxiety Disorders Interview Schedule for Children - Clinical Severity Rating Scale (ADIS-CSR) or Anxiety Disorders Interview Schedule for Children - Parent Version (ADIS-P): Principle anxiety diagnosis; Better indicated by lower values)											
79 (2 studies) 16-24 weeks	no serious risk of bias	very serious[6]	no serious indirectness	serious[4]	undetected	⊕⊖⊖⊖ **VERY LOW**[4,6] due to inconsistency,	45	34	N/A	N/A	The mean clinician-rated anxiety in the intervention groups was **1.19 standard deviations lower**

						imprecision					(1.7 to 0.68 lower)
Chronic anxiety (measured with: Revised Children's Manifest Anxiety Scale (RCMAS); Better indicated by lower values)											
47 (1 study) 24 weeks	serious[2]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[2,4] due to risk of bias, imprecision	19	28	N/A	N/A	The mean chronic anxiety in the intervention groups was **3.29 standard deviations lower** (4.19 to 2.38 lower)
Social anxiety (measured with: Spence Child Anxiety Scale-Parent (SCAS-P): Social phobia or Anxiety Disorders Interview Schedule for Children - Parent Version (ADIS-P): Social ; Better indicated by lower values)											
109 (2 studies) 6-24 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[7]	undetected	⊕⊖⊖⊖ **VERY LOW**[5,7] due to risk of bias, imprecision	43	66	N/A	N/A	The mean social anxiety in the intervention groups was **0.2 standard deviations lower** (0.59 lower to 0.2 higher)
Separation anxiety (measured with: Spence Child Anxiety Scale-Parent (SCAS-P): Separation or Anxiety Disorders Interview Schedule for Children - Parent Version (ADIS-P): Separation; Better indicated by lower values)											
109 (2 studies) 6-24 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[7]	undetected	⊕⊖⊖⊖ **VERY LOW**[5,7] due to risk of bias, imprecision	43	66	N/A	N/A	The mean separation anxiety in the intervention groups was **0.39 standard deviations lower** (0.78 lower to 0.01 higher)
Generalised Anxiety Disorder (measured with: Spence Child Anxiety Scale-Parent (SCAS-P): GAD or Anxiety Disorders Interview Schedule for Children - Parent Version (ADIS-P): GAD; Better indicated by lower values)											

87 (2 studies) 6-24 weeks	serious[5]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4,5] due to risk of bias, imprecision	43	44	N/A	N/A	The mean generalised anxiety disorder in the intervention groups was **0.66 standard deviations lower** (1.1 to 0.22 lower)
Anxiety relating to a specific phobia (measured with: Anxiety Disorders Interview Schedule for Children - Parent Version (ADIS-P): Specific phobia; Better indicated by lower values)											
43 (1 study) 24 weeks	serious[8]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4,8] due to risk of bias, imprecision	23	20	N/A	N/A	The mean anxiety relating to a specific phobia in the intervention groups was **0.99 standard deviations lower** (1.63 to 0.36 lower)
Panic (measured with: Spence Child Anxiety Scale-Parent (SCAS-P): Panic; Better indicated by lower values)											
66 (1 study) 6 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[7]	undetected	⊕⊖⊖⊖ **VERY LOW**[5,7] due to risk of bias, imprecision	20	46	N/A	N/A	The mean panic in the intervention groups was **0.15 standard deviations higher** (0.37 lower to 0.68 higher)
Panic (measured with: Spence Child Anxiety Scale-Parent (SCAS-P): Panic; Better indicated by lower values)											
66 (1 study) 12 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[7]	undetected	⊕⊖⊖⊖ **VERY LOW**[5,7] due to risk of bias, imprecision	20	46	N/A	N/A	The mean panic in the intervention groups was **0.13 standard deviations lower** (0.65 lower to 0.4 higher)
Fear of personal injury (measured with: Spence Child Anxiety Scale-Parent (SCAS-P): Personal Injury; Better indicated by lower values)											

66 (1 study) 6 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[7]	undetected	⊕⊖⊖⊖ **VERY LOW**[5,7] due to risk of bias, imprecision	20	46	N/A	N/A	The mean fear of personal injury in the intervention groups was **0.2 standard deviations higher** (0.32 lower to 0.73 higher)
Fear of personal injury (measured with: Spence Child Anxiety Scale-Parent (SCAS-P): Personal Injury; Better indicated by lower values)										
66 (1 study) 12 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[7]	undetected	⊕⊖⊖⊖ **VERY LOW**[5,7] due to risk of bias, imprecision	20	46	N/A	N/A	The mean fear of personal injury in the intervention groups was **0.31 standard deviations lower** (0.84 lower to 0.22 higher)
OCD (measured with: Spence Child Anxiety Scale-Parent (SCAS-P): OCD; Better indicated by lower values)										
66 (1 study) 6 weeks	serious[5]	no serious inconsistency	no serious indirectness	very serious[7]	undetected	⊕⊖⊖⊖ **VERY LOW**[5,7] due to risk of bias, imprecision	20	46	N/A	N/A	The mean ocd in the intervention groups was **0.33 standard deviations lower** (0.86 lower to 0.19 higher)
OCD (measured with: Spence Child Anxiety Scale-Parent (SCAS-P): OCD; Better indicated by lower values)										
66 (1 study) 12 weeks	serious[5]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4,5] due to risk of bias, imprecision	20	46	N/A	N/A	The mean ocd in the intervention groups was **1 standard deviations lower** (1.55 to 0.45 lower)
Emotional symptoms (measured with: Strengths and Difficulties Questionnaire: Parent Version: Emotional; Better indicated by lower values)										

47 (1 study) 24 weeks	serious[5]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4,5] due to risk of bias, imprecision	19	28	N/A	N/A	The mean emotional symptoms in the intervention groups was **4.29 standard deviations lower** (5.37 to 3.21 lower)
Emotional symptoms (measured with: Strengths and Difficulties Questionnaire: Teacher Version: Emotional; Better indicated by lower values)											
47 (1 study) 24 weeks	serious[9]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4,9] due to risk of bias, imprecision	19	28	N/A	N/A	The mean emotional symptoms in the intervention groups was **2.75 standard deviations lower** (3.57 to 1.93 lower)
Self-directed negative thoughts (measured with: Children's Automatic Thoughts Scale (CATS): Internalising; Better indicated by lower values)											
47 (1 study) 24 weeks	serious[2]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[2,4] due to risk of bias, imprecision	19	28	N/A	N/A	The mean self-directed negative thoughts in the intervention groups was **4.61 standard deviations lower** (5.75 to 3.48 lower)
Outward-directed negative thoughts (measured with: Children's Automatic Thoughts Scale (CATS): Hostile Intent; Better indicated by lower values)											
47 (1 study) 24 weeks	serious[2]	no serious inconsistency	no serious indirectness	very serious[7]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,7] due to risk of bias, imprecision	19	28	N/A	N/A	The mean outward-directed negative thoughts in the intervention groups was **0.33 standard deviations lower** (0.91 lower to 0.26 higher)

[1] Total events less than 300
[2] High risk of performance, response and detection bias. Self-report and children were not blind to treatment allocation or confounding factors.
[3] I squared value is considerable at 96% (p=0.00001)

[4] Total N less than 400
[5] High risk of performance, response and detection bias. Parent-rated and parents were not blind to treatment allocation or confounding factors.
[6] I squared value is considerable at 91% (p=0.00001)
[7] Total N is less than 400. 95% confidence interval crosses both line of no effect and measure of appreciable benefit/harm (SMD -0.5/0.5)
[8] High risk of performance and response bias as intervention administrators and participants were non-blind, and risk of detection bias was unclear/unknown as although outcome assessors were blind to treatment allocation the outcome measure was based on interview with parents who were involved in the intervention and not blind to treatment allocation
[9] High risk of performance, response and detection bias. Teacher-rated and unclear if teachers were blind to treatment allocation. Teachers are not blind to confounding factors.

1.26 PHARMACOLOGICAL INTERVENTIONS AIMED AT COEXISTING MENTAL HEALTH PROBLEMS

1.26.1 SNRIs for ADHD as a direct outcome

Atomoxetine versus placebo for ADHD as a direct outcome

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 Selective noradrenaline reuptake inhibitors versus placebo for hyperactivity/ADHD symptoms as a direct outcome**		**Risk with Control**	**Risk difference with Selective noradrenaline reuptake inhibitors versus placebo for hyperactivity/ADHD symptoms as a direct outcome** (95% CI)
Hyperactivity (parent-rated) (measured with: Aberrant Behaviour Checklist (ABC): Hyperactivity & Noncompliance; Better indicated by lower values)											
88 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	45	43	N/A	N/A	The mean hyperactivity (parent-rated) in the intervention groups was **0.19 standard deviations lower** (0.61 lower to 0.22 higher)

Hyperactivity (teacher-rated) (measured with: Conners' Teacher Rating Scale - Revised: Short Form (CTRS-R:S): Hyperactivity; Better indicated by lower values)											
72 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	36	36	N/A	N/A	The mean hyperactivity (teacher-rated) in the intervention groups was **0.12 standard deviations lower** (0.59 lower to 0.34 higher)
ADHD symptoms (parent-rated) (measured with: DSM-IV ADHD Rating Scale (ADHD-RS): Total; Better indicated by lower values)											
90 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊕⊖ **MODERATE**[2] due to imprecision	47	43	N/A	N/A	The mean adhd symptoms (parent-rated) in the intervention groups was **0.48 standard deviations lower** (0.9 to 0.06 lower)
ADHD symptoms (teacher-rated) (measured with: Conners' Teacher Rating Scale - Revised: Short Form (CTRS-R:S): ADHD; Better indicated by lower values)											
72 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	36	36	N/A	N/A	The mean adhd symptoms (teacher-rated) in the intervention groups was **0.15 standard deviations lower** (0.61 lower to 0.31 higher)
Inattention (teacher-rated) (measured with: Conners' Teacher Rating Scale - Revised: Short Form (CTRS-R:S): Cognitive/Attention; Better indicated by lower values)											
70 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	36	34	N/A	N/A	The mean inattention (teacher-rated) in the intervention groups was **0.37 standard deviations higher** (0.11 lower to 0.84 higher)

Oppositional (teacher-rated) (measured with: Conners' Teacher Rating Scale - Revised: Short Form (CTRS-R:S): Oppositional; Better indicated by lower values)											
72 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	36	36	N/A	N/A	The mean oppositional (teacher-rated) in the intervention groups was **0.1 standard deviations higher** (0.36 lower to 0.56 higher)
Improvement in ADHD symptoms (Clinician-rated) (measured with: Clinical Global Impression Scale-ADHD-Improvement (CGI-ADHD-I); Better indicated by lower values)											
89 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	46	43	N/A	N/A	The mean improvement in adhd symptoms (clinician-rated) in the intervention groups was **0.39 standard deviations lower** (0.81 lower to 0.03 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[2] N<400

1.27 BIOMEDICAL INTERVENTIONS AIMED AT COEXISTING MENTAL HEALTH PROBLEMS

1.27.1 Nutritional interventions for ADHD as an indirect outcome

Omega-3 fatty acids versus healthy diet control for ADHD as an indirect outcome

Quality assessment							Summary of Findings		
Participants	**Risk of**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication**	**Overall quality**	**Study event rates (%)**	**Relative**	**Anticipated absolute effects**

(studies) Follow up	bias				bias	of evidence	With Healthy diet control	With Omega-3 fatty acids	effect (95% CI)	Risk with Healthy diet control	Risk difference with Omega-3 fatty acids (95% CI)
ADHD symptoms (measured with: Child Behavior Checklist 1.5 - 5 (CBCL/1.5-5): ADHD; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A	The mean adhd symptoms in the intervention groups was **0.3 standard deviations lower** (1.13 lower to 0.53 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as the outcome assessor for this outcome measure was not blinded.
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Gluten-free and casein-free diet versus treatment-as-usual for ADHD as an indirect outcome

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 Treatment-as-usual	With Gluten-free and casein-free diet		Risk with Treatment-as-usual	Risk difference with Gluten-free and casein-free diet (95% CI)
Inattention (measured with: Attention-Deficit Hyperactivity Disorder-IV rating scale based on DSM-IV criteria (ADHD-IV): Inattention (change score); Better indicated by lower values)											
55 (1 study) 35 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias,	29	26	N/A	N/A	The mean inattention in the intervention groups was **0.59 standard deviations lower**

						imprecision				(1.13 to 0.05 lower)
Hyperactivity (measured with: Attention-Deficit Hyperactivity Disorder-IV rating scale based on DSM-IV criteria (ADHD-IV): Hyperactivity (change score); Better indicated by lower values)										
55 (1 study) 35 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	29	26	N/A	N/A	The mean hyperactivity in the intervention groups was **0.5 standard deviations lower** (1.04 lower to 0.04 higher)

[1] High risk of performance and response bias as intervention administrators (parents) and participants were non-blind and high risk of detection bias as parent-reported and non-blind to treatment allocation and other potentially confounding factors. There was also a high risk of attrition bias as over twice as many dropouts in the experimental group relative to the controls (32% in experimental group and 15% in the control group)
[2] N<400
[3] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.27.2 Nutritional interventions for anxiety as an indirect outcome

Omega-3 fatty acids versus placebo for anxiety as an indirect outcome

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 Placebo	**With Omega-3 fatty acids**		**Risk with Placebo**	**Risk difference with Omega-3 fatty acids** (95% CI)
Internalizing (measured with: Behavior Assessment System for Children (BASC): Internalizing; Better indicated by lower values)											

24 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	12	12	N/A	N/A	The mean internalizing in the intervention groups was **0.48 standard deviations lower** (1.3 lower to 0.33 higher)
[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)											

Omega-3 fatty acids versus healthy diet control for anxiety as an indirect outcome

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 Healthy diet control	**With Omega-3 fatty acids**		**Risk with Healthy diet control**	**Risk difference with Omega-3 fatty acids** (95% CI)
Internalizing (measured with: Child Behavior Checklist 1.5 - 5 (CBCL/1.5-5): Internalizing; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A	The mean internalizing in the intervention groups was **0.17 standard deviations lower** (0.99 lower to 0.66 higher)
Anxious/Depressed (measured with: Child Behavior Checklist 1.5 - 5 (CBCL/1.5-5): Anxious/Depressed; Better indicated by lower values)											
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A	The mean anxious/depressed in the intervention groups was **0.23 standard deviations lower** (1.05 lower to 0.6 higher)

Affective (measured with: Child Behavior Checklist 1.5 - 5 (CBCL/1.5-5): Affective; Better indicated by lower values)										
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A	The mean affective in the intervention groups was **0.07 standard deviations higher** (0.76 lower to 0.89 higher)
Anxiety (measured with: Child Behavior Checklist 1.5 - 5 (CBCL/1.5-5): Anxiety; Better indicated by lower values)										
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A	The mean anxiety in the intervention groups was **0.16 standard deviations lower** (0.99 lower to 0.66 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as the outcome assessor for this outcome measure was not blinded
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.27.3 Medical procedures for anxiety as an indirect outcome

Long-term chelation (7-rounds of DMSA therapy) versus short-term chelation (1-round of DMSA therapy and 6-rounds of placebo) for anxiety as an indirect outcome

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 Short-term chelation (1-round of DMSA therapy and 6-rounds	**With Long-term chelation (7-rounds of Dimercaptosuccinic Acid [DMSA] therapy)**		**Risk with Short-term chelation (1-round of DMSA therapy and 6-rounds of**	**Risk difference with Long-term chelation (7-rounds of Dimercaptosuccinic Acid [DMSA] therapy)** (95% CI)

							of placebo)			placebo)	
Specific Fears (measured with: Pervasive Development Disorder Behavior Inventory (PDDBI): Specific Fears; Better indicated by lower values)											
40 (1 study) 17 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	15	25	N/A	N/A	The mean specific fears in the intervention groups was **0.11 standard deviations lower** (0.75 lower to 0.53 higher)
[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)											

1.28 PSYCHOSOCIAL AND PHARMACOLOGICAL INTERVENTIONS AIMED AT COEXISTING MEDICAL OR FUNCTIONAL PROBLEMS

1.28.1 Cognitive-behavioural interventions for sleep problems as a direct outcome

CBT versus placebo for sleep problems as a direct outcome

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 CBT versus placebo**		**Risk with Control**	**Risk difference with CBT versus placebo** (95% CI)
Sleep onset latency (measured with: Actigraph; Better indicated by lower values)											
65 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32	33	N/A	N/A	The mean sleep onset latency in the intervention groups was **0.68 standard deviations lower**

										(1.18 to 0.18 lower)
Wake after sleep onset (measured with: Actigraph; Better indicated by lower values)										
65 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	32 33	N/A	N/A	The mean wake after sleep onset in the intervention groups was **0.24 standard deviations lower** (0.73 lower to 0.24 higher)
Nap time (measured with: Actigraph; Better indicated by lower values)										
65 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32 33	N/A	N/A	The mean nap time in the intervention groups was **0.81 standard deviations lower** (1.32 to 0.3 lower)
Bedtime (measured with: Actigraph; Better indicated by lower values)										
65 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32 33	N/A	N/A	The mean bedtime in the intervention groups was **0.89 standard deviations lower** (1.4 to 0.38 lower)
Total sleep time (measured with: Actigraph; Better indicated by lower values)										
65 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32 33	N/A	N/A	The mean total sleep time in the intervention groups was **0.62 standard deviations higher** (0.12 to 1.12 higher)
Sleep efficiency (measured with: Actigraph; Better indicated by lower values)										
65 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32 33	N/A	N/A	The mean sleep efficiency in the intervention groups was **1.98 standard deviations higher**

										(1.38 to 2.58 higher)
Sleep problems (measured with: Children's Sleep Habits Questionnaire (CSHQ): Total score; Better indicated by lower values)										
65 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	32	33	N/A	N/A	The mean sleep problems in the intervention groups was **1.01 standard deviations lower** (1.53 to 0.5 lower)
Bed resistance (measured with: Children's Sleep Habits Questionnaire (CSHQ): Bed resistance; Better indicated by lower values)										
65 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	32	33	N/A	N/A	The mean bed resistance in the intervention groups was **1.18 standard deviations lower** (1.71 to 0.65 lower)
Sleep onset delay (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep onset delay; Better indicated by lower values)										
65 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	32	33	N/A	N/A	The mean sleep onset delay in the intervention groups was **0.94 standard deviations lower** (1.45 to 0.42 lower)
Sleep anxiety (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep anxiety; Better indicated by lower values)										
65 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	32	33	N/A	N/A	The mean sleep anxiety in the intervention groups was **0.43 standard deviations lower** (0.92 lower to 0.06 higher)
Night-wakings (measured with: Children's Sleep Habits Questionnaire (CSHQ): Night-wakings; Better indicated by lower values)										
65 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	32	33	N/A	N/A	The mean night-wakings in the intervention groups was **0.84 standard deviations lower**

										(1.34 to 0.33 lower)
Sleep duration (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep duration; Better indicated by lower values)										
65 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	32 33	N/A	N/A	The mean sleep duration in the intervention groups was **0.23 standard deviations higher** (0.26 lower to 0.71 higher)
Parasomnias (measured with: Children's Sleep Habits Questionnaire (CSHQ): Parasomnias; Better indicated by lower values)										
65 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	32 33	N/A	N/A	The mean parasomnias in the intervention groups was **0.34 standard deviations higher** (0.15 lower to 0.83 higher)
Sleep disordered breathing (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep disordered breathing; Better indicated by lower values)										
65 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	32 33	N/A	N/A	The mean sleep disordered breathing in the intervention groups was **0 standard deviations higher** (0.49 lower to 0.49 higher)
Daytime sleepiness (measured with: Children's Sleep Habits Questionnaire (CSHQ): Daytime sleepiness; Better indicated by lower values)										
65 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	32 33	N/A	N/A	The mean daytime sleepiness in the intervention groups was **0.5 standard deviations lower** (1 to 0.01 lower)
Positive treatment response - Sleep onset latency (assessed with: Number of participants who showed sleep onset latency <30 min or reduction of sleep onset latency =>50% based on actigraph data)										
65	no serious	no serious	no serious	very	undetected	⊕⊕⊖⊖	0/32 3/33	**RR 6.79**	**Study population**	

(1 study) 12 weeks	risk of bias	inconsistency	indirectness	serious[4]		**LOW**[4] due to imprecision	(0%)	(9.1%)	(0.36 to 126.5)	**0 per 1000**	N/A
										Moderate	
										0 per 1000	N/A
Positive treatment response - Sleep efficiency (assessed with: Number of participants who showed =>85% for sleep efficiency based on actigraph data)											
65 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	undetected	⊕⊕⊖⊖ **LOW**[4] due to imprecision	0/32 (0%)	3/33 (9.1%)	**RR 6.79** (0.36 to 126.5)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A

[1] N<400
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] High risk of performance and response bias as intervention administrators and participants non-blind, and high risk of detection bias as parent-completed and parents non-blind and involved in the intervention
[4] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)

1.28.2 Melatonin for sleep problems as a direct outcome

Melatonin versus placebo for sleep problems as a direct outcome

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 Melatonin versus placebo for the coexisting problem of sleep**		**Risk with Control**	**Risk difference with Melatonin versus placebo for the coexisting problem of sleep** (95% CI)
Sleep onset latency (measured with: Actigraph; Better indicated by lower values)											

66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32	34	N/A	N/A	The mean sleep onset latency in the intervention groups was **1.23 standard deviations lower** (1.75 to 0.7 lower)
Wake after sleep onset (measured with: Actigraph; Better indicated by lower values)											
66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32	34	N/A	N/A	The mean wake after sleep onset in the intervention groups was **0.82 standard deviations lower** (1.32 to 0.31 lower)
Nap time (measured with: Actigraph; Better indicated by lower values)											
66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32	34	N/A	N/A	The mean nap time in the intervention groups was **0.57 standard deviations lower** (1.06 to 0.08 lower)
Bed time (measured with: Actigraph; Better indicated by lower values)											
66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32	34	N/A	N/A	The mean bed time in the intervention groups was **1.08 standard deviations lower** (1.6 to 0.56 lower)
Total sleep time (measured with: Actigraph; Better indicated by lower values)											
66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32	34	N/A	N/A	The mean total sleep time in the intervention groups was **1.45 standard deviations higher** (0.9 to 1.99 higher)

Sleep efficiency (measured with: Actigraph; Better indicated by lower values)										
66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32 34	N/A	N/A	The mean sleep efficiency in the intervention groups was **2.47 standard deviations higher** (1.82 to 3.12 higher)
Sleep problems (measured with: Children's Sleep Habits Questionnaire (CSHQ): Total score; Better indicated by lower values)										
66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32 34	N/A	N/A	The mean sleep problems in the intervention groups was **1.81 standard deviations lower** (2.39 to 1.23 lower)
Bed resistance (measured with: Children's Sleep Habits Questionnaire (CSHQ): Bed resistance; Better indicated by lower values)										
66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32 34	N/A	N/A	The mean bed resistance in the intervention groups was **1.72 standard deviations lower** (2.29 to 1.15 lower)
Sleep onset delay (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep onset delay; Better indicated by lower values)										
66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32 34	N/A	N/A	The mean sleep onset delay in the intervention groups was **1.58 standard deviations lower** (2.14 to 1.03 lower)
Sleep anxiety (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep anxiety; Better indicated by lower values)										
66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	32 34	N/A	N/A	The mean sleep anxiety in the intervention groups was **0.37 standard deviations lower**

										(0.86 lower to 0.12 higher)
Night-wakings (measured with: Children's Sleep Habits Questionnaire (CSHQ): Night-wakings; Better indicated by lower values)										
66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32	34	N/A	N/A	The mean night-wakings in the intervention groups was **2.88 standard deviations lower** (3.58 to 2.18 lower)
Sleep duration (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep duration; Better indicated by lower values)										
66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32	34	N/A	N/A	The mean sleep duration in the intervention groups was **1.39 standard deviations lower** (1.93 to 0.85 lower)
Parasomnias (measured with: Children's Sleep Habits Questionnaire (CSHQ): Parasomnias; Better indicated by lower values)										
66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	32	34	N/A	N/A	The mean parasomnias in the intervention groups was **0.11 standard deviations higher** (0.37 lower to 0.6 higher)
Sleep disordered breathing (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep disordered breathing; Better indicated by lower values)										
66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	32	34	N/A	N/A	The mean sleep disordered breathing in the intervention groups was **0.11 standard deviations lower** (0.59 lower to 0.38 higher)
Daytime sleepiness (measured with: Children's Sleep Habits Questionnaire (CSHQ): Daytime sleepiness; Better indicated by lower values)										
66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32	34	N/A	N/A	The mean daytime sleepiness in the intervention groups was **0.72 standard deviations lower**

											(1.21 to 0.22 lower)
Sleep onset latency (measured with: Sleep diary (study-specific); Better indicated by lower values)											
49 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	24	25	N/A	N/A	The mean sleep onset latency in the intervention groups was **0.76 standard deviations lower** (1.35 to 0.18 lower)
Total sleep time (measured with: Sleep diary (study-specific); Better indicated by lower values)											
47 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	24	23	N/A	N/A	The mean total sleep time in the intervention groups was **0.15 standard deviations higher** (0.43 lower to 0.72 higher)
Positive treatment response - Sleep onset latency (assessed with: Number of participants who showed sleep onset latency <30 min or reduction of sleep onset latency =>50% based on actigraph data)											
66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊕⊖ **MODERATE**[3] due to imprecision	0/32 (0%)	13/34 (38.2%)	**RR 25.46** (1.58 to 411.3)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Positive treatment response - Sleep efficiency (assessed with: Number of participants who showed =>85% for sleep efficiency based on actigraph data)											
66 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	undetected	⊕⊕⊕⊖ **MODERATE**[3] due to imprecision	0/32 (0%)	16/34 (47.1%)	**RR 31.11** (1.94 to 498.04)	**Study population**	
										0 per 1000	N/A
										Moderate	

									0 per 1000	N/A

[1] N<400
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] Events<300

Melatonin versus CBT for sleep problems as a direct outcome

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 Melatonin versus CBT for coexisting problem of sleep**		**Risk with Control**	**Risk difference with Melatonin versus CBT for coexisting problem of sleep** (95% CI)
Sleep onset latency (measured with: Actigraph; Better indicated by lower values)											
67 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	33	34	N/A	N/A	The mean sleep onset latency in the intervention groups was **0.54 standard deviations lower** (1.03 to 0.05 lower)
Wake after sleep onset (measured with: Actigraph; Better indicated by lower values)											
67 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	33	34	N/A	N/A	The mean wake after sleep onset in the intervention groups was **0.73 standard deviations lower** (1.22 to 0.23 lower)
Nap time (measured with: Actigraph; Better indicated by lower values)											
67 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to	33	34	N/A	N/A	The mean nap time in the intervention groups was **0.16 standard deviations**

						imprecision				**higher** (0.32 lower to 0.64 higher)
Bed time (measured with: Actigraph; Better indicated by lower values)										
67 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	33	34	N/A	N/A	The mean bed time in the intervention groups was **0.23 standard deviations lower** (0.71 lower to 0.25 higher)
Total sleep time (measured with: Actigraph; Better indicated by lower values)										
67 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	33	34	N/A	N/A	The mean total sleep time in the intervention groups was **0.76 standard deviations higher** (0.26 to 1.26 higher)
Sleep efficiency (measured with: Actigraph; Better indicated by lower values)										
67 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	33	34	N/A	N/A	The mean sleep efficiency in the intervention groups was **0.89 standard deviations higher** (0.39 to 1.4 higher)
Sleep problems (measured with: Children's Sleep Habits Questionnaire (CSHQ): Total score; Better indicated by lower values)										
67 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	33	34	N/A	N/A	The mean sleep problems in the intervention groups was **0.94 standard deviations lower** (1.45 to 0.44 lower)
Bed resistance (measured with: Children's Sleep Habits Questionnaire (CSHQ): Bed resistance; Better indicated by lower values)										
67 (1 study)	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3]	33	34	N/A	N/A	The mean bed resistance in the intervention groups

12 weeks						due to risk of bias, imprecision					was **0.5 standard deviations lower** (0.99 to 0.01 lower)
Sleep onset delay (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep onset delay; Better indicated by lower values)											
67 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	33	34	N/A	N/A	The mean sleep onset delay in the intervention groups was **0.65 standard deviations lower** (1.14 to 0.15 lower)
Sleep anxiety (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep anxiety; Better indicated by lower values)											
67 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	33	34	N/A	N/A	The mean sleep anxiety in the intervention groups was **0.02 standard deviations higher** (0.46 lower to 0.5 higher)
Night-wakings (measured with: Children's Sleep Habits Questionnaire (CSHQ): Night-wakings; Better indicated by lower values)											
67 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	33	34	N/A	N/A	The mean night-wakings in the intervention groups was **1.86 standard deviations lower** (2.44 to 1.28 lower)
Sleep duration (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep duration; Better indicated by lower values)											
67 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	33	34	N/A	N/A	The mean sleep duration in the intervention groups was **1.74 standard deviations lower** (2.31 to 1.18 lower)

Parasomnias (measured with: Children's Sleep Habits Questionnaire (CSHQ): Parasomnias; Better indicated by lower values)										
67 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	33	34	N/A	N/A	The mean parasomnias in the intervention groups was **0.23 standard deviations lower** (0.71 lower to 0.25 higher)
Sleep disordered breathing (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep disordered breathing; Better indicated by lower values)										
67 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	33	34	N/A	N/A	The mean sleep disordered breathing in the intervention groups was **0.11 standard deviations lower** (0.59 lower to 0.37 higher)
Daytime sleepiness (measured with: Children's Sleep Habits Questionnaire (CSHQ): Daytime sleepiness; Better indicated by lower values)										
67 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	33	34	N/A	N/A	The mean daytime sleepiness in the intervention groups was **0.26 standard deviations lower** (0.74 lower to 0.22 higher)
Positive treatment response - Sleep onset latency (assessed with: Number of participants who showed sleep onset latency <30 min or reduction of sleep onset latency =>50% based on actigraph data)										
67 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	3/33 (9.1%)	13/34 (38.2%)	**RR 4.21** (1.32 to 13.42)	**Study population**	
										91 per 1000	**292 more per 1000** (from 29 more to 1000 more)
										Moderate	
										91 per 1000	**292 more per 1000** (from 29 more to 1000 more)

Positive treatment response - Sleep efficiency (assessed with: Number of participants who showed =>85% for sleep efficiency based on actigraph data)										
67 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	3/33 (9.1%) 16/34 (47.1%)	**RR 5.18** (1.66 to 16.13)	**Study population**	
									91 per 1000	**380 more per 1000** (from 60 more to 1000 more)
									Moderate	
									91 per 1000	**380 more per 1000** (from 60 more to 1000 more)

[1] N<400
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] High risk of performance and response bias as intervention administrators and participants non-blind, and high risk of detection bias as parent-completed and parents non-blind and involved in the intervention
[4] Events<300

1.28.3 Combined cognitive-behavioural intervention and melatonin for sleep problems as a direct outcome

COMB versus placebo for sleep problems as a direct outcome

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 Combined melatonin and CBT versus placebo**		**Risk with Control**	**Risk difference with Combined melatonin and CBT versus placebo** (95% CI)
Sleep onset latency (measured with: Actigraph; Better indicated by lower values)											
67 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32	35	N/A	N/A	The mean sleep onset latency in the intervention groups was **1.86 standard deviations lower**

											(2.44 to 1.29 lower)
Wake after sleep onset (measured with: Actigraph; Better indicated by lower values)											
67 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32	35	N/A	N/A	The mean wake after sleep onset in the intervention groups was **1.29 standard deviations lower** (1.82 to 0.76 lower)
Nap time (measured with: Actigraph; Better indicated by lower values)											
67 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32	35	N/A	N/A	The mean nap time in the intervention groups was **0.95 standard deviations lower** (1.45 to 0.44 lower)
Bedtime (measured with: Actigraph; Better indicated by lower values)											
67 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32	35	N/A	N/A	The mean bedtime in the intervention groups was **1.32 standard deviations lower** (1.85 to 0.79 lower)
Total sleep time (measured with: Actigraph; Better indicated by lower values)											
67 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32	35	N/A	N/A	The mean total sleep time in the intervention groups was **2.33 standard deviations higher** (1.7 to 2.96 higher)
Sleep efficiency (measured with: Actigraph; Better indicated by lower values)											
67 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	32	35	N/A	N/A	The mean sleep efficiency in the intervention groups was **2.8 standard deviations**

										higher (2.12 to 3.49 higher)
Sleep problems (measured with: Children's Sleep Habits Questionnaire (CSHQ): Total score; Better indicated by lower values)										
67 (1 study) 12 weeks	serious[2]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	32	35	N/A	N/A	The mean sleep problems in the intervention groups was **4.44 standard deviations lower** (5.35 to 3.53 lower)
Bed resistance (measured with: Children's Sleep Habits Questionnaire (CSHQ): Bed resistance; Better indicated by lower values)										
67 (1 study) 12 weeks	serious[2]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	32	35	N/A	N/A	The mean bed resistance in the intervention groups was **3.34 standard deviations lower** (4.09 to 2.58 lower)
Sleep onset delay (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep onset delay; Better indicated by lower values)										
67 (1 study) 12 weeks	serious[2]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	32	35	N/A	N/A	The mean sleep onset delay in the intervention groups was **2.21 standard deviations lower** (2.82 to 1.59 lower)
Sleep anxiety (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep anxiety; Better indicated by lower values)										
67 (1 study) 12 weeks	serious[2]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	32	35	N/A	N/A	The mean sleep anxiety in the intervention groups was **1.74 standard deviations lower** (2.3 to 1.17 lower)
Night-wakings (measured with: Children's Sleep Habits Questionnaire (CSHQ): Night-wakings; Better indicated by lower values)										
67	serious[2]	no serious	no serious	serious[1]	undetected	⊕⊕⊖⊖	32	35	N/A	N/A	The mean night-wakings

(1 study) 12 weeks		inconsistency	indirectness			**LOW**[1,2] due to risk of bias, imprecision					in the intervention groups was **3.96 standard deviations lower** (4.8 to 3.12 lower)
Sleep duration (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep duration; Better indicated by lower values)											
67 (1 study) 12 weeks	serious[2]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	32	35	N/A	N/A	The mean sleep duration in the intervention groups was **1.73 standard deviations lower** (2.29 to 1.16 lower)
Parasomnias (measured with: Children's Sleep Habits Questionnaire (CSHQ): Parasomnias; Better indicated by lower values)											
67 (1 study) 12 weeks	serious[2]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	32	35	N/A	N/A	The mean parasomnias in the intervention groups was **0.16 standard deviations lower** (0.64 lower to 0.32 higher)
Sleep disordered breathing (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep disordered breathing; Better indicated by lower values)											
67 (1 study) 12 weeks	serious[2]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	32	35	N/A	N/A	The mean sleep disordered breathing in the intervention groups was **0.03 standard deviations higher** (0.45 lower to 0.51 higher)
Daytime sleepiness (measured with: Children's Sleep Habits Questionnaire (CSHQ): Daytime sleepiness; Better indicated by lower values)											
67 (1 study) 12 weeks	serious[2]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	32	35	N/A	N/A	The mean daytime sleepiness in the intervention groups was **1.15 standard deviations lower** (1.67 to 0.63 lower)

Positive treatment response - Sleep onset latency (assessed with: Number of participants who showed sleep onset latency <30 min or reduction of sleep onset latency =>50% based on actigraph data)										
67 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	0/32 (0%)	30/35 (85.7%)	**RR 55.92** (3.56 to 878.39)	**Study population** **0 per 1000** N/A **Moderate** **0 per 1000** N/A
Positive treatment response - Sleep efficiency (assessed with: Number of participants who showed =>85% for sleep efficiency based on actigraph data)										
67 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	0/32 (0%)	22/35 (62.9%)	**RR 41.25** (2.6 to 653.27)	**Study population** **0 per 1000** N/A **Moderate** **0 per 1000** N/A

[1] N<400
[2] High risk of performance and response bias as intervention administrators and participants non-blind, and high risk of detection bias as parent-completed and parents non-blind and involved in the intervention
[3] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[4] Events<300

COMB versus CBT-only for sleep problems as a direct outcome

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 Combined melatonin and CBT versus CBT-only for coexisting problem of**		**Risk with Control**	**Risk difference with Combined melatonin and CBT versus CBT-only for coexisting problem of sleep** (95% CI)

							sleep				
Sleep onset latency (measured with: Actigraph; Better indicated by lower values)											
68 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	33	35	N/A	N/A	The mean sleep onset latency in the intervention groups was **1.15 standard deviations lower** (1.67 to 0.64 lower)
Wake after sleep onset (measured with: Actigraph; Better indicated by lower values)											
68 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	33	35	N/A	N/A	The mean wake after sleep onset in the intervention groups was **1.4 standard deviations lower** (1.94 to 0.87 lower)
Nap time (measured with: Actigraph; Better indicated by lower values)											
68 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	33	35	N/A	N/A	The mean nap time in the intervention groups was **0.13 standard deviations lower** (0.61 lower to 0.35 higher)
Bed time (measured with: Actigraph; Better indicated by lower values)											
68 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	33	35	N/A	N/A	The mean bed time in the intervention groups was **0.47 standard deviations lower** (0.95 lower to 0.01 higher)
Total sleep time (measured with: Actigraph; Better indicated by lower values)											
68 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	33	35	N/A	N/A	The mean total sleep time in the intervention groups was **1.46 standard deviations higher**

										(0.93 to 2 higher)
Sleep efficiency (measured with: Actigraph; Better indicated by lower values)										
68 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	33 35	N/A	N/A	The mean sleep efficiency in the intervention groups was **1.33 standard deviations higher** (0.81 to 1.86 higher)
Sleep problems (measured with: Children's Sleep Habits Questionnaire (CSHQ): Total score; Better indicated by lower values)										
68 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	33 35	N/A	N/A	The mean sleep problems in the intervention groups was **3.1 standard deviations lower** (3.81 to 2.38 lower)
Bed resistance (measured with: Children's Sleep Habits Questionnaire (CSHQ): Bed resistance; Better indicated by lower values)										
68 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	33 35	N/A	N/A	The mean bed resistance in the intervention groups was **1.7 standard deviations lower** (2.26 to 1.14 lower)
Sleep onset delay (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep onset delay; Better indicated by lower values)										
68 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	33 35	N/A	N/A	The mean sleep onset delay in the intervention groups was **1.23 standard deviations lower** (1.75 to 0.71 lower)
Sleep anxiety (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep anxiety; Better indicated by lower values)										
68 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of	33 35	N/A	N/A	The mean sleep anxiety in the intervention groups was **1.55 standard deviations**

						bias, imprecision					**lower** (2.1 to 1.01 lower)
Night-wakings (measured with: Children's Sleep Habits Questionnaire (CSHQ): Night-wakings; Better indicated by lower values)											
68 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⨁⨁⨀⨀ **LOW**[1,3] due to risk of bias, imprecision	33	35	N/A	N/A	The mean night-wakings in the intervention groups was **2.66 standard deviations lower** (3.32 to 2 lower)
Sleep duration (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep duration; Better indicated by lower values)											
68 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⨁⨁⨀⨀ **LOW**[1,3] due to risk of bias, imprecision	33	35	N/A	N/A	The mean sleep duration in the intervention groups was **2.09 standard deviations lower** (2.68 to 1.49 lower)
Parasomnias (measured with: Children's Sleep Habits Questionnaire (CSHQ): Parasomnias; Better indicated by lower values)											
68 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⨁⨁⨀⨀ **LOW**[1,3] due to risk of bias, imprecision	33	35	N/A	N/A	The mean parasomnias in the intervention groups was **0.48 standard deviations lower** (0.96 lower to 0 higher)
Sleep disordered breathing (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep disordered breathing; Better indicated by lower values)											
68 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⨁⨁⨀⨀ **LOW**[1,3] due to risk of bias, imprecision	33	35	N/A	N/A	The mean sleep disordered breathing in the intervention groups was **0.03 standard deviations higher** (0.45 lower to 0.5 higher)
Daytime sleepiness (measured with: Children's Sleep Habits Questionnaire (CSHQ): Daytime sleepiness; Better indicated by lower values)											
68 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⨁⨁⨀⨀ **LOW**[1,3] due to risk of bias, imprecision	33	35	N/A	N/A	The mean daytime sleepiness in the intervention groups was **0.61 standard deviations**

										lower (1.09 to 0.12 lower)
Positive treatment response - Sleep onset latency (assessed with: Number of participants who showed sleep onset latency <30 min or reduction of sleep onset latency =>50% based on actigraph data)										
68 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	3/33 (9.1%)	30/35 (85.7%)	**RR 9.43** (3.18 to 27.97)	**Study population**
										91 per 1000 **766 more per 1000** (from 198 more to 1000 more)
										Moderate
										91 per 1000 **767 more per 1000** (from 198 more to 1000 more)
Positive treatment response - Sleep efficiency (=>85% for sleep efficiency) (assessed with: Number of participants who showed =>85% for sleep efficiency based on actigraph data)										
68 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	3/33 (9.1%)	22/35 (62.9%)	**RR 6.91** (2.28 to 20.95)	**Study population**
										91 per 1000 **537 more per 1000** (from 116 more to 1000 more)
										Moderate
										91 per 1000 **538 more per 1000** (from 116 more to 1000 more)

[1] N<400
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] High risk of performance and response bias as intervention administrators and participants non-blind, and high risk of detection bias as parent-completed and parents non-blind and involved in the intervention
[4] Events<300

COMB versus melatonin-only for sleep problems as a direct outcome

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 Combined melatonin and CBT versus melatonin-only for coexisting problem of sleep		Risk with Control	Risk difference with Combined melatonin and CBT versus melatonin-only for coexisting problem of sleep (95% CI)
Sleep onset latency (measured with: Actigraph; Better indicated by lower values)											
69 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	34	35	N/A	N/A	The mean sleep onset latency in the intervention groups was **0.59 standard deviations lower** (1.07 to 0.11 lower)
Wake after sleep onset (measured with: Actigraph; Better indicated by lower values)											
69 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	34	35	N/A	N/A	The mean wake after sleep onset in the intervention groups was **0.68 standard deviations lower** (1.17 to 0.19 lower)
Nap time (measured with: Actigraph; Better indicated by lower values)											
69 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	34	35	N/A	N/A	The mean nap time in the intervention groups was **0.27 standard deviations lower** (0.75 lower to 0.2 higher)
Bed time (measured with: Actigraph; Better indicated by lower values)											
69 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	34	35	N/A	N/A	The mean bed time in the intervention groups was **0.22 standard deviations lower** (0.69 lower to 0.25 higher)

Total sleep time (measured with: Actigraph; Better indicated by lower values)										
69 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊕⊖ **MODERATE**[1] due to imprecision	34	35	N/A	N/A	The mean total sleep time in the intervention groups was **0.61 standard deviations higher** (0.13 to 1.1 higher)
Sleep efficiency (measured with: Actigraph; Better indicated by lower values)										
69 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊕⊖⊖ **LOW**[2] due to imprecision	34	35	N/A	N/A	The mean sleep efficiency in the intervention groups was **0.42 standard deviations higher** (0.06 lower to 0.9 higher)
Sleep problems (measured with: Children's Sleep Habits Questionnaire (CSHQ): Total score; Better indicated by lower values)										
69 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	34	35	N/A	N/A	The mean sleep problems in the intervention groups was **1.42 standard deviations lower** (1.95 to 0.89 lower)
Bed resistance (measured with: Children's Sleep Habits Questionnaire (CSHQ): Bed resistance; Better indicated by lower values)										
69 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	34	35	N/A	N/A	The mean bed resistance in the intervention groups was **1.1 standard deviations lower** (1.61 to 0.59 lower)
Sleep onset delay (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep onset delay; Better indicated by lower values)										
69 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	34	35	N/A	N/A	The mean sleep onset delay in the intervention groups was **0.57 standard deviations lower**

										(1.06 to 0.09 lower)
Sleep anxiety (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep anxiety; Better indicated by lower values)										
69 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	34	35	N/A	N/A	The mean sleep anxiety in the intervention groups was **1.33 standard deviations lower** (1.85 to 0.8 lower)
Night-wakings (measured with: Children's Sleep Habits Questionnaire (CSHQ): Night-wakings; Better indicated by lower values)										
69 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1,3] due to risk of bias, imprecision	34	35	N/A	N/A	The mean night-wakings in the intervention groups was **0.6 standard deviations lower** (1.08 to 0.12 lower)
Sleep duration (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep duration; Better indicated by lower values)										
69 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	34	35	N/A	N/A	The mean sleep duration in the intervention groups was **0.44 standard deviations lower** (0.92 lower to 0.03 higher)
Parasomnias (measured with: Children's Sleep Habits Questionnaire (CSHQ): Parasomnias; Better indicated by lower values)										
69 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	34	35	N/A	N/A	The mean parasomnias in the intervention groups was **0.27 standard deviations lower** (0.74 lower to 0.21 higher)
Sleep disordered breathing (measured with: Children's Sleep Habits Questionnaire (CSHQ): Sleep disordered breathing; Better indicated by lower values)										
69 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	34	35	N/A	N/A	The mean sleep disordered breathing in the intervention groups was **0.09 standard deviations higher**

										(0.38 lower to 0.56 higher)
Daytime sleepiness (measured with: Children's Sleep Habits Questionnaire (CSHQ): Daytime sleepiness; Better indicated by lower values)										
69 (1 study) 12 weeks	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[2,3] due to risk of bias, imprecision	34	35	N/A	N/A	The mean daytime sleepiness in the intervention groups was **0.27 standard deviations lower** (0.74 lower to 0.21 higher)
Positive treatment response - Sleep onset latency (assessed with: Number of participants who showed sleep onset latency <30 min or reduction of sleep onset latency =>50% based on actigraph data)										
69 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊕⊖ **MODERATE**[4] due to imprecision	13/34 (38.2%)	30/35 (85.7%)	**RR 2.24** (1.43 to 3.51)	**Study population**	
										382 per 1000	**474 more per 1000** (from 164 more to 960 more)
										Moderate	
										382 per 1000	**474 more per 1000** (from 164 more to 959 more)
Positive treatment response - Sleep efficiency (assessed with: Number of participants who showed =>85% for sleep efficiency based on actigraph data)										
69 (1 study) 12 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊕⊖⊖ **LOW**[5] due to imprecision	16/34 (47.1%)	22/35 (62.9%)	**RR 1.34** (0.86 to 2.07)	**Study population**	
										471 per 1000	**160 more per 1000** (from 66 fewer to 504 more)
										Moderate	
										471 per 1000	**160 more per 1000** (from 66 fewer to 504 more)

[1] N<400
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[3] High risk of performance and response bias as intervention administrators and participants non-blind, and high risk of detection bias as parent-completed and parents non-blind and involved in the intervention

[4] Events<300
[5] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)

1.28.4 SNRIs for sleep problems as an indirect outcome

Atomoxetine versus placebo for sleep problems as an indirect outcome

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 Selective noradrenaline reuptake inhibitors versus placebo for sleep problems as an indirect outcome		Risk with Control	Risk difference with Selective noradrenaline reuptake inhibitors versus placebo for sleep problems as an indirect outcome (95% CI)
Time to fall asleep (measured with: Sleep Measure Scale: Time to fall asleep; Better indicated by lower values)											
89 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	46	43	N/A	N/A	The mean time to fall asleep in the intervention groups was **0.29 standard deviations lower** (0.7 lower to 0.13 higher)
Total hours of sleep (measured with: Sleep Measure Scale: Total hours of sleep; Better indicated by lower values)											
88 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	46	42	N/A	N/A	The mean total hours of sleep in the intervention groups was **0.13 standard deviations lower** (0.55 lower to 0.29 higher)
Difficulty falling asleep (measured with: Sleep Measure Scale: Difficulty falling asleep; Better indicated by lower values)											
89 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	46	43	N/A	N/A	The mean difficulty falling asleep in the intervention groups was **0.17 standard deviations higher**

										(0.24 lower to 0.59 higher)
Quality of sleep (measured with: Sleep Measure Scale: Quality of sleep; Better indicated by lower values)										
89 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	46	43	N/A	N/A	The mean quality of sleep in the intervention groups was **0.23 standard deviations lower** (0.65 lower to 0.18 higher)
Functional outcome during day (measured with: Sleep Measure Scale: Functional outcome during day; Better indicated by lower values)										
89 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	46	43	N/A	N/A	The mean functional outcome during day in the intervention groups was **0.18 standard deviations lower** (0.6 lower to 0.24 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.29 BIOMEDICAL INTERVENTIONS AIMED AT COEXISTING MEDICAL OR FUNCTIONAL PROBLEMS

1.29.1 Nutritional interventions for sleep problems as an indirect outcome

Multivitamin/mineral supplement versus placebo for sleep problems as an indirect outcome

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 Placebo	**With Multivitamin and mineral supplement**		**Risk with Placebo**	**Risk difference with Multivitamin and mineral supplement** (95% CI)

Sleep improvement (measured with: Parent Global Impressions-Revised (PGI-R): Sleep improvement; Better indicated by lower values)										
104 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	51 53	N/A	N/A	The mean sleep improvement in the intervention groups was **0.18 standard deviations higher** (0.2 lower to 0.57 higher)
[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)										

Omega-3 fatty acids versus healthy diet control for sleep problems as an indirect outcome

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 Healthy diet control	**With Omega-3 fatty acids**		**Risk with Healthy diet control** / **Risk difference with Omega-3 fatty acids** (95% CI)
Sleep problems (measured with: Child Behavior Checklist 1.5 - 5 (CBCL/1.5-5): Sleep problems; Better indicated by lower values)										
23 (1 study) 13 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	13	10	N/A	N/A / The mean sleep problems in the intervention groups was **1.11 standard deviations higher** (0.21 to 2 higher)
[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as the outcome assessor for this outcome measure was not blinded [2] N<400										

1.29.2 Hormones for gastrointestinal symptoms as an indirect outcome

Secretin versus placebo for gastrointestinal symptoms as an indirect outcome

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 Secretin versus placebo for gastrointestinal symptoms as an indirect outcome		Risk with Control	Risk difference with Secretin versus placebo for gastrointestinal symptoms as an indirect outcome (95% CI)
Number of gastrointestinal problems (measured with: GI symptoms questionnaire: Total (change score); Better indicated by lower values)											
95 (1 study) 3 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	48	47	N/A	N/A	The mean number of gastrointestinal problems in the intervention groups was **0.18 standard deviations lower** (0.59 lower to 0.22 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.29.3 Nutritional interventions for gastrointestinal symptoms as a direct or indirect outcome

Immunoglobulin versus placebo for gastrointestinal symptoms as a direct outcome

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 Placebo	With Immunoglobulin (dosages combined)		Risk with Placebo	Risk difference with Immunoglobulin (dosages

										combined) (95% CI)
Positive treatment response (assessed with: Dichotomous measure of 'moderately or substantially improved' on at least two of last 4 assessments or 'somewhat improved' for all of last 4 assessments of the Modified Global Improvement Scale [MGIS] for GI symptoms)										
125 (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	14/31 (45.2%)	31/94 (33%)	**RR 0.73** (0.45 to 1.18)	**Study population**
										452 per 1000 / **122 fewer per 1000** (from 248 fewer to 81 more)
										Moderate
										452 per 1000 / **122 fewer per 1000** (from 249 fewer to 81 more)

[1] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)
[2] High risk of selective reporting bias as continuous data could not be extracted for the MGIS scale

Multivitamin/ mineral supplement versus placebo for gastrointestinal symptoms as an indirect outcome

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 Placebo	**With Multivitamin and mineral supplement**		**Risk with Placebo**	**Risk difference with Multivitamin and mineral supplement** (95% CI)
Gastrointestinal symptom improvement (measured with: Parent Global Impressions-Revised (PGI-R): GI improvement; Better indicated by lower values)											
104 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to	51	53	N/A	N/A	The mean gastrointestinal symptom improvement in the intervention groups was **0.3 standard deviations**

						imprecision				**higher** (0.09 lower to 0.68 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.30 PSYCHOSOCIAL INTERVENTIONS AIMED AT IMPROVING THE IMPACT OF AUTISM ON THE FAMILY

1.30.1 Behavioural interventions for improving the impact of autism on the family as an indirect outcome

Home-based EBI versus centre-based EBI for improving the impact of autism on the family as an indirect outcome

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 Home-based versus Centre-based EBI for improving the impact on the family as an indirect outcome**		**Risk with Control**	**Risk difference with Home-based versus Centre-based EBI for improving the impact on the family as an indirect outcome** (95% CI)
Family quality of life (measured with: Beach Family Quality of Life Questionnaire: Total; Better indicated by lower values)											
44 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	23	21	N/A	N/A	The mean family quality of life in the intervention groups was **0.16 standard deviations higher** (0.43 lower to 0.76 higher)
Family quality of life (family interaction) (measured with: Beach Family Quality of Life Questionnaire: Family interaction; Better indicated by lower values)											
44 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias,	23	21	N/A	N/A	The mean family quality of life (family interaction) in the intervention groups was **0.14 standard deviations**

						imprecision				**higher** (0.45 lower to 0.73 higher)
Family quality of life (parenting) (measured with: Beach Family Quality of Life Questionnaire: Parenting; Better indicated by lower values)										
44 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	23	21	N/A	N/A	The mean family quality of life (parenting) in the intervention groups was **0 standard deviations higher** (0.59 lower to 0.59 higher)
Family quality of life (emotional wellbeing) (measured with: Beach Family Quality of Life Questionnaire: Emotional wellbeing; Better indicated by lower values)										
44 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	23	21	N/A	N/A	The mean family quality of life (emotional wellbeing) in the intervention groups was **0.22 standard deviations higher** (0.38 lower to 0.81 higher)
Family quality of life (physical wellbeing) (measured with: Beach Family Quality of Life Questionnaire: Physical wellbeing; Better indicated by lower values)										
44 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	23	21	N/A	N/A	The mean family quality of life (physical wellbeing) in the intervention groups was **0 standard deviations higher** (0.59 lower to 0.59 higher)
Family quality of life (disability support) (measured with: Beach Family Quality of Life Questionnaire: Disability support; Better indicated by lower values)										
44 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	23	21	N/A	N/A	The mean family quality of life (disability support) in the intervention groups was **0.1 standard deviations higher** (0.49 lower to 0.69 higher)
Parental coping skills (measured with: Parent Perception Questionnaire: Total; Better indicated by lower values)										
46 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of	23	23	N/A	N/A	The mean parental coping skills in the intervention groups was **0.15 standard deviations**

						bias, imprecision				**lower** (0.73 lower to 0.43 higher)
Parental coping skills (confidence) (measured with: Parent Perception Questionnaire: Confidence; Better indicated by lower values)										
46 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	23	23	N/A	N/A	The mean parental coping skills (confidence) in the intervention groups was **0 standard deviations higher** (0.58 lower to 0.58 higher)
Parental coping skills (coping) (measured with: Parent Perception Questionnaire: Coping; Better indicated by lower values)										
46 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	23	23	N/A	N/A	The mean parental coping skills (coping) in the intervention groups was **0.33 standard deviations higher** (0.25 lower to 0.91 higher)
Parental coping skills (knowledge) (measured with: Parent Perception Questionnaire: Knowledge; Better indicated by lower values)										
46 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	23	23	N/A	N/A	The mean parental coping skills (knowledge) in the intervention groups was **0.52 standard deviations lower** (1.11 lower to 0.07 higher)
Parental coping skills (understanding) (measured with: Parent Perception Questionnaire: Understanding; Better indicated by lower values)										
46 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	23	23	N/A	N/A	The mean parental coping skills (understanding) in the intervention groups was **0.26 standard deviations lower** (0.84 lower to 0.32 higher)
Parental coping skills (family issues) (measured with: Parent Perception Questionnaire: Family issues; Better indicated by lower values)										
46 (1 study)	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2]	23	23	N/A	N/A	The mean parental coping skills (family issues) in the

40 weeks						due to risk of bias, imprecision					intervention groups was **0.23 standard deviations higher** (0.35 lower to 0.81 higher)
Parental coping skills (planning) (measured with: Parent Perception Questionnaire: Planning; Better indicated by lower values)											
46 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	23	23	N/A	N/A	The mean parental coping skills (planning) in the intervention groups was **0.09 standard deviations lower** (0.67 lower to 0.49 higher)
Parental stress (measured with: Parenting Stress Index-3rd Edition (PSI): Total; Better indicated by lower values)											
40 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	20	20	N/A	N/A	The mean parental stress in the intervention groups was **0.26 standard deviations lower** (0.89 lower to 0.36 higher)
Parental stress (defensive responding) (measured with: Parenting Stress Index (PSI): Defensive responding; Better indicated by lower values)											
40 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	20	20	N/A	N/A	The mean parental stress (defensive responding) in the intervention groups was **0.21 standard deviations lower** (0.83 lower to 0.42 higher)
Parental stress (parental distress) (measured with: Parenting Stress Index (PSI): Parental distress; Better indicated by lower values)											
40 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	20	20	N/A	N/A	The mean parental stress (parental distress) in the intervention groups was **0.22 standard deviations lower** (0.84 lower to 0.4 higher)
Parental stress (parent-child dysfunctional interaction) (measured with: Parenting Stress Index (PSI): Parent-child dysfunctional interaction; Better indicated by											

lower values)										
40 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	20 20	N/A	N/A	The mean parental stress (parent-child dysfunctional interaction) in the intervention groups was **0.15 standard deviations lower** (0.77 lower to 0.47 higher)
Parental stress (difficult child) (measured with: Parenting Stress Index (PSI): Difficult child; Better indicated by lower values)										
40 (1 study) 40 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	20 20	N/A	N/A	The mean parental stress (difficult child) in the intervention groups was **0.35 standard deviations lower** (0.98 lower to 0.27 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and unclear/unknown risk of detection bias as although the outcome assessors were blinded, this outcome measure was based on interview with parent and parents were non-blind and were part of the intervention
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.30.2 Cognitive-behavioural interventions for improving the impact of autism on the family as an indirect outcome

CBT versus waitlist for improving the impact of autism on the family as an indirect outcome

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 Behaviour-focused intervention versus treatment-as-usual for improving the impact on the family as an indirect outcome**		**Risk with Control**	**Risk difference with Behaviour-focused intervention versus treatment-as-usual for improving the impact on the family as an indirect outcome** (95% CI)
Parent intrusiveness/child independence (measured with: Parent-Child Interaction Questionnaire (PCIQ): Parent Intrusiveness ; Better indicated by lower values)											

40 (1 study) 16 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	20	20	N/A	N/A	The mean parent intrusiveness/child independence in the intervention groups was **0.68 standard deviations lower** (1.32 to 0.04 lower)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome assessors were non-blind parents
[2] N<400

1.30.3 Parent training for improving the impact of autism on the family as a direct or indirect outcome

Parent training versus treatment as usual for improving the impact of autism on the family as a direct or indirect outcome

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 Parent training versus treatment-as-usual for improving the impact of autism on the family**		**Risk with Control**	**Risk difference with Parent training versus treatment-as-usual for improving the impact of autism on the family** (95% CI)
Parental stress (direct or indirect outcome) (measured with: Parenting Stress Thermometer or Parental Stress Inventory: Total or Parenting Stress Index-3rd Edition (PSI): Total; Better indicated by lower values)											
143 (3 studies) 12-52 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	55	88	N/A	N/A	The mean parental stress (direct or indirect outcome) in the intervention groups was **0.39 standard deviations lower** (0.73 to 0.04 lower)
Parental stress (direct outcome; combined PEBM+PEC post-intervention) (measured with: Parenting Stress Thermometer; Better indicated by lower values)											
103 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias,	35	68	N/A	N/A	The mean parental stress (direct outcome; combined pebm+pec post-intervention) in the intervention groups was

						imprecision					**0.42 standard deviations lower** (0.84 to 0.01 lower)
Parental stress (indirect outcome) (measured with: Parental Stress Inventory: Total or Parenting Stress Index-3rd Edition (PSI): Total; Better indicated by lower values)											
40 (2 studies) 12-52 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	20	20	N/A	N/A	The mean parental stress (indirect outcome) in the intervention groups was **0.30 standard deviations lower** (0.93 lower to 0.32 higher)
Parental mental health (combined PEBM+PEC groups) (measured with: General Health Questionnaire (GHQ-28): Total; Better indicated by lower values)											
103 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	35	68	N/A	N/A	The mean parental mental health (combined pebm+pec groups) in the intervention groups was **0.26 standard deviations lower** (0.67 lower to 0.15 higher)
Parental mental health (combined PEBM+PEC groups) (measured with: General Health Questionnaire (GHQ-28): Total; Better indicated by lower values)											
103 (1 study) 46 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	35	68	N/A	N/A	The mean parental mental health (combined pebm+pec groups) in the intervention groups was **0.45 standard deviations lower** (0.86 to 0.03 lower)
Parental somatic symptoms (combined PEBM+PEC groups) (measured with: General Health Questionnaire (GHQ-28): Somatic symptoms; Better indicated by lower values)											
103 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	35	68	N/A	N/A	The mean parental somatic symptoms (combined pebm+pec groups) in the intervention groups was **0.19 standard deviations lower**

											(0.6 lower to 0.22 higher)
Parental somatic symptoms (combined PEBM+PEC groups) (measured with: General Health Questionnaire (GHQ-28): Somatic symptoms; Better indicated by lower values)											
103 (1 study) 46 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	35	68	N/A	N/A	The mean parental somatic symptoms (combined pebm+pec groups) in the intervention groups was **0.22 standard deviations lower** (0.63 lower to 0.19 higher)
Parental anxiety and insomnia (combined PEBM+PEC groups) (measured with: General Health Questionnaire (GHQ-28): Anxiety and insomnia; Better indicated by lower values)											
103 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	35	68	N/A	N/A	The mean parental anxiety and insomnia (combined pebm+pec groups) in the intervention groups was **0.16 standard deviations lower** (0.57 lower to 0.25 higher)
Parental anxiety and insomnia (combined PEBM+PEC groups) (measured with: General Health Questionnaire (GHQ-28): Anxiety and insomnia; Better indicated by lower values)											
103 (1 study) 46 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	35	68	N/A	N/A	The mean parental anxiety and insomnia (combined pebm+pec groups) in the intervention groups was **0.54 standard deviations lower** (0.95 to 0.12 lower)
Parental social dysfunction (combined PEBM+PEC groups) (measured with: General Health Questionnaire (GHQ-28): Social dysfunction; Better indicated by lower values)											
103 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias,	35	68	N/A	N/A	The mean parental social dysfunction (combined pebm+pec groups) in the intervention groups was

						imprecision					**0.65 standard deviations lower** (1.07 to 0.23 lower)
Parental social dysfunction (combined PEBM+PEC groups) (measured with: General Health Questionnaire (GHQ-28): Social dysfunction; Better indicated by lower values)											
103 (1 study) 46 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	35	68	N/A	N/A	The mean parental social dysfunction (combined pebm+pec groups) in the intervention groups was **0.37 standard deviations lower** (0.78 lower to 0.04 higher)
Parental severe depression (combined PEBM+PEC groups) (measured with: General Health Questionnaire (GHQ-28): Severe depression; Better indicated by lower values)											
103 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	35	68	N/A	N/A	The mean parental severe depression (combined pebm+pec groups) in the intervention groups was **0.09 standard deviations higher** (0.32 lower to 0.49 higher)
Parental severe depression (combined PEBM+PEC groups) (measured with: General Health Questionnaire (GHQ-28): Severe depression; Better indicated by lower values)											
103 (1 study) 46 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	35	68	N/A	N/A	The mean parental severe depression (combined pebm+pec groups) in the intervention groups was **0.14 standard deviations lower** (0.55 lower to 0.27 higher)
General family function (combined PEBM+PEC groups) (measured with: McMaster Family Assessment Device (FAD); Better indicated by lower values)											
103 (1 study) 20 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of	35	68	N/A	N/A	The mean general family function (combined pebm+pec groups) in the intervention

						bias, imprecision				groups was **0.31 standard deviations lower** (0.72 lower to 0.1 higher)
General family function (combined PEBM+PEC groups) (measured with: McMaster Family Assessment Device (FAD); Better indicated by lower values)										
103 (1 study) 46 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	35	68	N/A	N/A	The mean general family function (combined pebm+pec groups) in the intervention groups was **0.14 standard deviations lower** (0.55 lower to 0.27 higher)

[1] High risk of performance and response bias as intervention administrators and participants non-blind and high risk of detection bias as parent-completed and parents involved in intervention and not blinded
[2] N<400
[3] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Parent and day-care staff training versus standard day-care for improving the impact of autism on the family as an indirect outcome

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 Parent and day-care staff training versus standard day-care for improving the impact of autism on the family as an indirect outcome**		**Risk with Control**	**Risk difference with Parent and day-care staff training versus standard day-care for improving the impact of autism on the family as an indirect outcome** (95% CI)
Maternal stress (measured with: Stress-Arousal Checklist: Mothers' Stress; Better indicated by lower values)											
35 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	19	16	N/A	N/A	The mean maternal stress in the intervention groups was **0.06 standard deviations lower**

										(0.73 lower to 0.61 higher)
Maternal arousal (measured with: Stress-Arousal Checklist: Mothers' Arousal; Better indicated by lower values)										
35 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	19 16	N/A	N/A	The mean maternal arousal in the intervention groups was **0.18 standard deviations higher** (0.48 lower to 0.85 higher)
Paternal stress (measured with: Stress-Arousal Checklist: Fathers' Stress; Better indicated by lower values)										
35 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	19 16	N/A	N/A	The mean paternal stress in the intervention groups was **0.14 standard deviations higher** (0.53 lower to 0.8 higher)
Paternal arousal (measured with: Stress-Arousal Checklist: Fathers' Arousal; Better indicated by lower values)										
35 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	19 16	N/A	N/A	The mean paternal arousal in the intervention groups was **0.51 standard deviations higher** (0.16 lower to 1.19 higher)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as the reliability and validity of this outcome measure is unclear and parent-completed and parents involved in the intervention so non-blind
[2] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.31 PHARMACOLOGICAL INTERVENTIONS AIMED AT IMPROVING THE IMPACT OF AUTISM ON THE FAMILY

1.31.1 SNRIs for improving the impact of autism on the family as an indirect outcome

Atomoxetine versus placebo for improving the impact of autism on the family as an indirect outcome

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 Selective noradrenaline reuptake inhibitors versus placebo for improving the impact of autism on the family as an indirect outcome		Risk with Control	Risk difference with Selective noradrenaline reuptake inhibitors versus placebo for improving the impact of autism on the family as an indirect outcome (95% CI)
Parental mental health (measured with: General Health Questionnaire (GHQ-28): Total; Better indicated by lower values)											
89 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	46	43	N/A	N/A	The mean parental mental health in the intervention groups was **0.24 standard deviations lower** (0.66 lower to 0.18 higher)
Parental stress (measured with: Nijmeegse Ouderlijke Stress Index (NOSI): Total; Better indicated by lower values)											
77 (1 study) 8 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	39	38	N/A	N/A	The mean parental stress in the intervention groups was **0.24 standard deviations lower** (0.69 lower to 0.21 higher)

[1] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

1.32 BIOMEDICAL INTERVENTIONS AIMED AT IMPROVING THE IMPACT OF AUTISM ON THE FAMILY

1.32.1 Complementary therapies for improving the impact of autism on the family as an indirect outcome

Qigong massage training versus waitlist for improving the impact of autism on the family as an indirect outcome

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 Qigong massage versus waitlist for impact on family as an indirect outcome		Risk with Control	Risk difference with Qigong massage versus waitlist for impact on family as an indirect outcome (95% CI)
Parental stress (measured with: Autism Parenting Stress Index (ASPI); Better indicated by lower values)											
41 (1 study) 17 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	18	23	N/A	N/A	The mean parental stress in the intervention groups was **0.78 standard deviations lower** (1.42 to 0.14 lower)

[1] High risk of performance and response bias as intervention administrators and participants were non-blind, and high risk of detection bias as outcome assessors were parents who were delivering the intervention and the outcome measure was created for this study so reliability and validity is unknown
[2] N<400

1.33 ADVERSE EVENTS ASSOCIATED WITH PHARMACOLOGICAL INTERVENTIONS

1.33.1 Adverse events associated with anticonvulsants

Adverse events associated with divalproex versus placebo

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 Adverse events associated with anticonvulsants		Risk with Control	Risk difference with Adverse events associated with anticonvulsants (95% CI)
Any adverse event (assessed with: Number of participants experiencing any side effect during the trial (measured using checklist derived from Physicians' Desk Reference, 1997))											
30 (1 study) 8 weeks	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,	11/14 (78.6%)	15/16 (93.8%)	**RR 1.19** (0.88 to 1.61)	**Study population**	
										786 per	**149 more per 1000**

						imprecision, publication bias			**1000**	(from 94 fewer to 479 more)
									Moderate	
									786 per 1000	**149 more per 1000** (from 94 fewer to 479 more)
More than one adverse event (assessed with: Number of participants experiencing more than one adverse event during the trial (measured using physical examination))										
27 (1 study) 12 weeks	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	2/11 (18.2%)	5/16 (31.3%)	**RR 1.72** (0.4 to 7.32)	**Study population**	
									182 per 1000	**131 more per 1000** (from 109 fewer to 1000 more)
									Moderate	
									182 per 1000	**131 more per 1000** (from 109 fewer to 1000 more)
Discontinuation due to adverse event (assessed with: Number of participants who discontinued due to adverse event)										
57 (2 studies) 8-12 weeks	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	0/25 (0%)	2/32 (6.3%)	**RR 2.37** (0.26 to 21.43)	**Study population**	
									0 per 1000	N/A
									Moderate	
									0 per 1000	N/A
Weight gain (measured with: Number of kilograms or pounds that participants gained during the trial; Better indicated by lower values)										
57 (2 studies) 8-12 weeks	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,	25	32	N/A	N/A	The mean weight gain in the intervention groups was **0.29 standard**

						publication bias				**deviations higher** (0.24 lower to 0.82 higher)

[1] High risk of detection bias as unclear if follow-up duration (=<12 weeks) is sufficient to observe potential longer term adverse events
[2] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)
[3] Trial funded by pharmaceutical company and/or study drugs were provided by pharmaceutical company and/or authors are consultants to pharmaceutical companies

1.33.2 Adverse events associated with antidepressants

Adverse events associated with citalopram versus placebo

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 Adverse events associated with antidepressants		Risk with Control	Risk difference with Adverse events associated with antidepressants (95% CI)
Any adverse event (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	66/76 (86.8%)	71/73 (97.3%)	**RR 1.12** (1.02 to 1.23)	**Study population**	
										868 per 1000	**104 more per 1000** (from 17 more to 200 more)
										Moderate	
										868 per 1000	**104 more per 1000** (from 17 more to 200 more)
Nightmares (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	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	0/76 (0%)	5/73 (6.8%)	**RR 11.45** (0.64 to 203.38)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Increased energy level (assessed with: Safety Monitoring Uniform Report Form)											

149 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	15/76 (19.7%)	28/73 (38.4%)	**RR 1.94** (1.13 to 3.33)	**Study population**	
										197 per 1000	**186 more per 1000** (from 26 more to 460 more)
										Moderate	
										197 per 1000	**185 more per 1000** (from 26 more to 459 more)
Anger or irritability (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	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	13/76 (17.1%)	18/73 (24.7%)	**RR 1.44** (0.76 to 2.73)	**Study population**	
										171 per 1000	**75 more per 1000** (from 41 fewer to 296 more)
										Moderate	
										171 per 1000	**75 more per 1000** (from 41 fewer to 296 more)
Aggression or hostility (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	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	13/76 (17.1%)	17/73 (23.3%)	**RR 1.36** (0.71 to 2.6)	**Study population**	
										171 per 1000	**62 more per 1000** (from 50 fewer to 274 more)
										Moderate	
										171 per 1000	**62 more per 1000** (from 50 fewer to 274 more)
Headache or migraine (assessed with: Safety Monitoring Uniform Report Form)											

149 (1 study) 12 weeks	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	10/76 (13.2%)	15/73 (20.5%)	**RR 1.56** (0.75 to 3.25)	**Study population**	
										132 per 1000	**74 more per 1000** (from 33 fewer to 296 more)
										Moderate	
										132 per 1000	**74 more per 1000** (from 33 fewer to 297 more)
Restlessness or difficulty settling down (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	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	7/76 (9.2%)	13/73 (17.8%)	**RR 1.93** (0.82 to 4.57)	**Study population**	
										92 per 1000	**86 more per 1000** (from 17 fewer to 329 more)
										Moderate	
										92 per 1000	**86 more per 1000** (from 17 fewer to 328 more)
Disinhibited, impulsive, or intrusive behaviour (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	5/76 (6.6%)	14/73 (19.2%)	**RR 2.92** (1.11 to 7.68)	**Study population**	
										66 per 1000	**126 more per 1000** (from 7 more to 439 more)
										Moderate	
										66 per 1000	**127 more per 1000** (from 7 more to 441 more)
Silliness (assessed with: Safety Monitoring Uniform Report Form)											

149 (1 study) 12 weeks	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	10/76 (13.2%)	9/73 (12.3%)	**RR 0.94** (0.4 to 2.17)	**Study population**	
										132 per 1000	**8 fewer per 1000** (from 79 fewer to 154 more)
										Moderate	
										132 per 1000	**8 fewer per 1000** (from 79 fewer to 154 more)
Anxiety (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	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	9/76 (11.8%)	8/73 (11%)	**RR 0.93** (0.38 to 2.27)	**Study population**	
										118 per 1000	**8 fewer per 1000** (from 73 fewer to 150 more)
										Moderate	
										118 per 1000	**8 fewer per 1000** (from 73 fewer to 150 more)
Mood lability (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	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	9/76 (11.8%)	7/73 (9.6%)	**RR 0.81** (0.32 to 2.06)	**Study population**	
										118 per 1000	**22 fewer per 1000** (from 81 fewer to 126 more)
										Moderate	
										118 per 1000	**22 fewer per 1000** (from 80 fewer to 125 more)
Increased speech (assessed with: Safety Monitoring Uniform Report Form)											

149 (1 study) 12 weeks	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	4/76 (5.3%)	8/73 (11%)	**RR 2.08** (0.66 to 6.62)	**Study population**	
										53 per 1000	**57 more per 1000** (from 18 fewer to 296 more)
										Moderate	
										53 per 1000	**57 more per 1000** (from 18 fewer to 298 more)
Decreased attention and concentration (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	2/76 (2.6%)	9/73 (12.3%)	**RR 4.68** (1.05 to 20.96)	**Study population**	
										26 per 1000	**97 more per 1000** (from 1 more to 525 more)
										Moderate	
										26 per 1000	**96 more per 1000** (from 1 more to 519 more)
Hyperactivity (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	2/76 (2.6%)	9/73 (12.3%)	**RR 4.68** (1.05 to 20.96)	**Study population**	
										26 per 1000	**97 more per 1000** (from 1 more to 525 more)
										Moderate	
										26 per 1000	**96 more per 1000** (from 1 more to 519 more)
Stereotypy (assessed with: Safety Monitoring Uniform Report Form)											

149 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	1/76 (1.3%)	8/73 (11%)	**RR 8.33** (1.07 to 64.95)	**Study population**	
										13 per 1000	**96 more per 1000** (from 1 more to 841 more)
										Moderate	
										13 per 1000	**95 more per 1000** (from 1 more to 831 more)
Diarrhoea or loose stools (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	9/76 (11.8%)	19/73 (26%)	**RR 2.2** (1.06 to 4.54)	**Study population**	
										118 per 1000	**142 more per 1000** (from 7 more to 419 more)
										Moderate	
										118 per 1000	**142 more per 1000** (from 7 more to 418 more)
Abdominal discomfort (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	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	9/76 (11.8%)	13/73 (17.8%)	**RR 1.5** (0.68 to 3.3)	**Study population**	
										118 per 1000	**59 more per 1000** (from 38 fewer to 272 more)
										Moderate	
										118 per 1000	**59 more per 1000** (from 38 fewer to 271 more)
Vomiting or nausea (assessed with: Safety Monitoring Uniform Report Form)											

149 (1 study) 12 weeks	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	6/76 (7.9%)	14/73 (19.2%)	**RR 2.43** (0.99 to 5.98)	**Study population**	
										79 per 1000	**113 more per 1000** (from 1 fewer to 393 more)
										Moderate	
										79 per 1000	**113 more per 1000** (from 1 fewer to 393 more)
Any insomnia (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	17/76 (22.4%)	28/73 (38.4%)	**RR 1.71** (1.03 to 2.86)	**Study population**	
										224 per 1000	**159 more per 1000** (from 7 more to 416 more)
										Moderate	
										224 per 1000	**159 more per 1000** (from 7 more to 417 more)
Initial insomnia or difficulty falling asleep (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	7/76 (9.2%)	17/73 (23.3%)	**RR 2.53** (1.11 to 5.74)	**Study population**	
										92 per 1000	**141 more per 1000** (from 10 more to 437 more)
										Moderate	
										92 per 1000	**141 more per 1000** (from 10 more to 436 more)
Midcycle or other insomnia (assessed with: Safety Monitoring Uniform Report Form)											

149 (1 study) 12 weeks	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	9/76 (11.8%)	13/73 (17.8%)	**RR 1.5** (0.68 to 3.3)	**Study population**	
										118 per 1000	**59 more per 1000** (from 38 fewer to 272 more)
										Moderate	
										118 per 1000	**59 more per 1000** (from 38 fewer to 271 more)
Cold, flu or other systemic infection (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	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	26/76 (34.2%)	31/73 (42.5%)	**RR 1.24** (0.82 to 1.87)	**Study population**	
										342 per 1000	**82 more per 1000** (from 62 fewer to 298 more)
										Moderate	
										342 per 1000	**82 more per 1000** (from 62 fewer to 298 more)
Decreased appetite (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	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	10/76 (13.2%)	11/73 (15.1%)	**RR 1.15** (0.52 to 2.53)	**Study population**	
										132 per 1000	**20 more per 1000** (from 63 fewer to 201 more)
										Moderate	
										132 per 1000	**20 more per 1000** (from 63 fewer to 202 more)
Increased appetite (assessed with: Safety Monitoring Uniform Report Form)											

149 (1 study) 12 weeks	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	8/76 (10.5%)	7/73 (9.6%)	**RR 0.91** (0.35 to 2.38)	**Study population**	
										105 per 1000	**9 fewer per 1000** (from 68 fewer to 145 more)
										Moderate	
										105 per 1000	**9 fewer per 1000** (from 68 fewer to 145 more)
Rash (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	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	8/76 (10.5%)	12/73 (16.4%)	**RR 1.56** (0.68 to 3.6)	**Study population**	
										105 per 1000	**59 more per 1000** (from 34 fewer to 274 more)
										Moderate	
										105 per 1000	**59 more per 1000** (from 34 fewer to 273 more)
Other skin or subcutaneous tissue disorder (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	1/76 (1.3%)	9/73 (12.3%)	**RR 9.37** (1.22 to 72.12)	**Study population**	
										13 per 1000	**110 more per 1000** (from 3 more to 936 more)
										Moderate	
										13 per 1000	**109 more per 1000** (from 3 more to 925 more)
Fatigue (assessed with: Safety Monitoring Uniform Report Form)											

149 (1 study) 12 weeks	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	10/76 (13.2%)	10/73 (13.7%)	**RR 1.04** (0.46 to 2.35)	**Study population**	
										132 per 1000	**5 more per 1000** (from 71 fewer to 178 more)
										Moderate	
										132 per 1000	**5 more per 1000** (from 71 fewer to 178 more)
Allergies (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	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	11/76 (14.5%)	15/73 (20.5%)	**RR 1.42** (0.7 to 2.88)	**Study population**	
										145 per 1000	**61 more per 1000** (from 43 fewer to 272 more)
										Moderate	
										145 per 1000	**61 more per 1000** (from 43 fewer to 273 more)
Cough (assessed with: Safety Monitoring Uniform Report Form)											
149 (1 study) 12 weeks	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	5/76 (6.6%)	10/73 (13.7%)	**RR 2.08** (0.75 to 5.8)	**Study population**	
										66 per 1000	**71 more per 1000** (from 16 fewer to 316 more)
										Moderate	
										66 per 1000	**71 more per 1000** (from 16 fewer to 317 more)
Any serious adverse event (assessed with: Safety Monitoring Uniform Report Form)											

149 (1 study) 12 weeks	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	0/76 (0%)	1/73 (1.4%)	**RR 3.12** (0.13 to 75.42)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A

[1] High risk of detection bias as unclear if follow-up duration (=<12 weeks) is sufficient to observe potential longer term adverse events
[2] Events<300
[3] Authors are consultants to pharmaceutical companies
[4] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)

1.33.3 Adverse events associated with antihistamines

Adverse events associated with cyproheptadine and haloperidol versus placebo and haloperidol

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 Adverse events associated with combined antihistamines and antipsychotics**		**Risk with Control**	**Risk difference with Adverse events associated with combined antihistamines and antipsychotics** (95% CI)
Extrapyramidal symptoms (assessed with: Extrapyramidal Symptoms Rating Scale (ESRS): Total)											
40 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	6/20 (30%)	2/20 (10%)	**RR 0.33** (0.08 to 1.46)	**Study population**	
										300 per 1000	**201 fewer per 1000** (from 276 fewer to 138 more)
										Moderate	
										300 per 1000	**201 fewer per 1000** (from 276 fewer to 138

										more)
Trouble swallowing (assessed with: Study-specific side effect checklist)										
40 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	4/20 (20%) 2/20 (10%)	**RR 0.5** (0.1 to 2.43)	**Study population**	
									200 per 1000	**100 fewer per 1000** (from 180 fewer to 286 more)
									Moderate	
									200 per 1000	**100 fewer per 1000** (from 180 fewer to 286 more)
Stiffness (assessed with: Study-specific side effect checklist)										
40 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	3/20 (15%) 1/20 (5%)	**RR 0.33** (0.04 to 2.94)	**Study population**	
									150 per 1000	**101 fewer per 1000** (from 144 fewer to 291 more)
									Moderate	
									150 per 1000	**101 fewer per 1000** (from 144 fewer to 291 more)
Slow movement (assessed with: Study-specific side effect checklist)										
40 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	3/20 (15%) 1/20 (5%)	**RR 0.33** (0.04 to 2.94)	**Study population**	
									150 per 1000	**101 fewer per 1000** (from 144 fewer to 291 more)
									Moderate	
									150 per	**101 fewer per 1000**

									1000	(from 144 fewer to 291 more)
Constipation (assessed with: Study-specific side effect checklist)										
40 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	2/20 (10%) 4/20 (20%)	**RR 2** (0.41 to 9.71)	**Study population**	
									100 per 1000	**100 more per 1000** (from 59 fewer to 871 more)
									Moderate	
									100 per 1000	**100 more per 1000** (from 59 fewer to 871 more)
Diarrhoea (assessed with: Study-specific side effect checklist)										
40 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	3/20 (15%) 2/20 (10%)	**RR 0.67** (0.12 to 3.57)	**Study population**	
									150 per 1000	**49 fewer per 1000** (from 132 fewer to 386 more)
									Moderate	
									150 per 1000	**49 fewer per 1000** (from 132 fewer to 386 more)
Increased appetite (assessed with: Study-specific side effect checklist)										
40 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	4/20 (20%) 9/20 (45%)	**RR 2.25** (0.83 to 6.13)	**Study population**	
									200 per 1000	**250 more per 1000** (from 34 fewer to 1000 more)
									Moderate	
									200 per	**250 more per 1000** (from 34 fewer to 1000

									1000	more)
Morning drowsiness (assessed with: Study-specific side effect checklist)										
40 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	2/20 (10%) 3/20 (15%)	**RR 1.5** (0.28 to 8.04)	**Study population**	
									100 per 1000	**50 more per 1000** (from 72 fewer to 704 more)
									Moderate	
									100 per 1000	**50 more per 1000** (from 72 fewer to 704 more)
Day time drowsiness (assessed with: Study-specific side effect checklist)										
40 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	2/20 (10%) 1/20 (5%)	**RR 0.5** (0.05 to 5.08)	**Study population**	
									100 per 1000	**50 fewer per 1000** (from 95 fewer to 408 more)
									Moderate	
									100 per 1000	**50 fewer per 1000** (from 95 fewer to 408 more)
Restlessness (assessed with: Study-specific side effect checklist)										
40 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	4/20 (20%) 1/20 (5%)	**RR 0.25** (0.03 to 2.05)	**Study population**	
									200 per 1000	**150 fewer per 1000** (from 194 fewer to 210 more)
									Moderate	
									200 per 1000	**150 fewer per 1000** (from 194 fewer to 210 more)
Fatigue (assessed with: Study-specific side effect checklist)										

40 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	2/20 (10%)	3/20 (15%)	**RR 1.5** (0.28 to 8.04)	**Study population** **100 per 1000** **50 more per 1000** (from 72 fewer to 704 more) **Moderate** **100 per 1000** **50 more per 1000** (from 72 fewer to 704 more)	

[1] High risk of detection bias as unclear if follow-up duration (=<12 weeks) is sufficient to observe potential longer term adverse events
[2] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)

1.33.4 Adverse events associated with antioxidants

Adverse events associated with N-acetylcysteine versus placebo

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 Adverse events associated with antioxidants**		**Risk with Control**	**Risk difference with Adverse events associated with antioxidants** (95% CI)
Any gastrointestinal side effect (assessed with: Dosage Record and Treatment Emergent Symptom Scale (DOTES))											
29 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	7/15 (46.7%)	11/14 (78.6%)	**RR 1.68** (0.92 to 3.09)	**Study population** **467 per 1000**	**317 more per 1000** (from 37 fewer to 975 more) **Moderate** **467 per 1000** **318 more per 1000** (from 37 fewer to 976 more)
Constipation (assessed with: Dosage Record and Treatment Emergent Symptom Scale (DOTES))											

29 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	2/15 (13.3%)	3/14 (21.4%)	**RR 1.61** (0.31 to 8.24)	**Study population**	
										133 per 1000	**81 more per 1000** (from 92 fewer to 965 more)
										Moderate	
										133 per 1000	**81 more per 1000** (from 92 fewer to 963 more)
Nausea (assessed with: Dosage Record and Treatment Emergent Symptom Scale (DOTES))											
29 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	3/15 (20%)	6/14 (42.9%)	**RR 2.14** (0.66 to 6.97)	**Study population**	
										200 per 1000	**228 more per 1000** (from 68 fewer to 1000 more)
										Moderate	
										200 per 1000	**228 more per 1000** (from 68 fewer to 1000 more)
Diarrhoea (assessed with: Dosage Record and Treatment Emergent Symptom Scale (DOTES))											
29 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	1/15 (6.7%)	3/14 (21.4%)	**RR 3.21** (0.38 to 27.4)	**Study population**	
										67 per 1000	**147 more per 1000** (from 41 fewer to 1000 more)
										Moderate	
										67 per 1000	**148 more per 1000** (from 42 fewer to 1000 more)
Increased appetite (assessed with: Dosage Record and Treatment Emergent Symptom Scale (DOTES))											

29 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	0/15 (0%)	2/14 (14.3%)	**RR 5.33** (0.28 to 102.26)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Decreased appetite (assessed with: Dosage Record and Treatment Emergent Symptom Scale (DOTES))											
29 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	3/15 (20%)	2/14 (14.3%)	**RR 0.71** (0.14 to 3.66)	**Study population**	
										200 per 1000	**58 fewer per 1000** (from 172 fewer to 532 more)
										Moderate	
										200 per 1000	**58 fewer per 1000** (from 172 fewer to 532 more)
Akathisia (assessed with: Dosage Record and Treatment Emergent Symptom Scale (DOTES))											
29 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	0/15 (0%)	1/14 (7.1%)	**RR 3.2** (0.14 to 72.62)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Increased motor activity (assessed with: Dosage Record and Treatment Emergent Symptom Scale (DOTES))											
29 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of	3/15 (20%)	2/14 (14.3%)	**RR 0.71** (0.14 to 3.66)	**Study population**	
										200 per	**58 fewer per 1000**

						bias, imprecision				**1000**	(from 172 fewer to 532 more)
										Moderate	
										200 per 1000	**58 fewer per 1000** (from 172 fewer to 532 more)
Tremor (assessed with: Dosage Record and Treatment Emergent Symptom Scale (DOTES))											
29 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	1/15 (6.7%)	0/14 (0%)	**RR 0.36** (0.02 to 8.07)	**Study population**	
										67 per 1000	**43 fewer per 1000** (from 65 fewer to 471 more)
										Moderate	
										67 per 1000	**43 fewer per 1000** (from 66 fewer to 474 more)
Dizziness (assessed with: Dosage Record and Treatment Emergent Symptom Scale (DOTES))											
29 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	1/15 (6.7%)	0/14 (0%)	**RR 0.36** (0.02 to 8.07)	**Study population**	
										67 per 1000	**43 fewer per 1000** (from 65 fewer to 471 more)
										Moderate	
										67 per 1000	**43 fewer per 1000** (from 66 fewer to 474 more)
Excitement/agitation (assessed with: Dosage Record and Treatment Emergent Symptom Scale (DOTES))											
29	serious[1]	no serious	no serious	very	undetected	⊕⊖⊖⊖	3/15	2/14	**RR 0.71**	**Study population**	

(1 study) 12 weeks		inconsistency	indirectness	serious[2]		**VERY LOW**[1,2] due to risk of bias, imprecision	(20%)	(14.3%)	(0.14 to 3.66)	**200 per 1000**	**58 fewer per 1000** (from 172 fewer to 532 more)
										Moderate	
										200 per 1000	**58 fewer per 1000** (from 172 fewer to 532 more)
Depressed affect (assessed with: Dosage Record and Treatment Emergent Symptom Scale (DOTES))											
29 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	0/15 (0%)	1/14 (7.1%)	**RR 3.2** (0.14 to 72.62)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Nasal congestion (assessed with: Dosage Record and Treatment Emergent Symptom Scale (DOTES))											
29 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious	undetected	⊕⊖⊖⊖ **VERY LOW**[1] due to risk of bias, imprecision	6/15 (40%)	4/14 (28.6%)	**RR 0.71** (0.25 to 2.01)	**Study population**	
										400 per 1000	**116 fewer per 1000** (from 300 fewer to 404 more)
										Moderate	
										400 per 1000	**116 fewer per 1000** (from 300 fewer to 404 more)
Increased salivation (assessed with: Dosage Record and Treatment Emergent Symptom Scale (DOTES))											
29 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of	2/15 (13.3%)	0/14 (0%)	**RR 0.21** (0.01 to 4.09)	**Study population**	
										133 per	**105 fewer per 1000**

						bias, imprecision			**1000**	(from 132 fewer to 412 more)
									Moderate	
									133 per 1000	**105 fewer per 1000** (from 132 fewer to 411 more)
Sweating (assessed with: Dosage Record and Treatment Emergent Symptom Scale (DOTES))										
29 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	1/15 (6.7%)	0/14 (0%)	**RR 0.36** (0.02 to 8.07)	**Study population**
									67 per 1000	**43 fewer per 1000** (from 65 fewer to 471 more)
									Moderate	
									67 per 1000	**43 fewer per 1000** (from 66 fewer to 474 more)

[1] High risk of detection bias as unclear if follow-up duration (=<12 weeks) is sufficient to observe potential longer term adverse events
[2] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)

1.33.5 Adverse events associated with antipsychotics

Adverse events associated with antipsychotics versus placebo

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	**With Adverse events associated with antipsychotics**		**Risk with**	**Risk difference with Adverse events associated with antipsychotics** (95% CI)
Any side effect (Aripiprazole, haloperidol or risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure or study-specific report)											

528 (5 studies) 6-12 weeks	serious[1]	serious[2]	no serious indirectness	no serious imprecision	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, inconsistency, publication bias	130/195 (66.7%)	283/333 (85%)	**RR 1.27** (1.14 to 1.42)	**Study population**	
										667 per 1000	**180 more per 1000** (from 93 more to 280 more)
										Moderate	
										720 per 1000	**194 more per 1000** (from 101 more to 302 more)
Any side effect (Aripiprazole) (assessed with: Study-specific report of adverse events)											
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	73/101 (72.3%)	188/212 (88.7%)	**RR 1.23** (1.08 to 1.41)	**Study population**	
										723 per 1000	**166 more per 1000** (from 58 more to 296 more)
										Moderate	
										723 per 1000	**166 more per 1000** (from 58 more to 296 more)
Any side effect (Haloperidol) (assessed with: Outcome measure not reported)											
40 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	5/20 (25%)	16/20 (80%)	**RR 3.2** (1.45 to 7.05)	**Study population**	
										250 per 1000	**550 more per 1000** (from 113 more to 1000 more)
										Moderate	
										250 per 1000	**550 more per 1000** (from 113 more to 1000 more)
Any side effect (Risperidone) (assessed with: Non-systematic assessment or study-specific outcome measure)											

175 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	52/74 (70.3%)	79/101 (78.2%)	**RR 1.17** (0.98 to 1.39)	**Study population**	
										703 per 1000	**119 more per 1000** (from 14 fewer to 274 more)
										Moderate	
										697 per 1000	**118 more per 1000** (from 14 fewer to 272 more)
Discontinuation due to adverse events (Aripiprazole)) (assessed with: Study-specific report)											
98 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	3/51 (5.9%)	5/47 (10.6%)	**RR 1.81** (0.46 to 7.16)	**Study population**	
										59 per 1000	**48 more per 1000** (from 32 fewer to 362 more)
										Moderate	
										59 per 1000	**48 more per 1000** (from 32 fewer to 363 more)
Discontinuation due to drooling (Aripiprazole) (assessed with: Study-specific report)											
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	0/51 (0%)	3/165 (1.8%)	**RR 2.19** (0.12 to 41.76)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Discontinuation due to sedation (Aripiprazole) (assessed with: Study-specific report)											

216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	0/51 (0%)	7/165 (4.2%)	**RR 4.7** (0.27 to 80.88)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Discontinuation due to tremor (Aripiprazole) (assessed with: Study-specific report)											
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	0/51 (0%)	4/165 (2.4%)	**RR 2.82** (0.15 to 51.5)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Clinically relevant (>=7%) weight gain (Aripiprazole) (assessed with: Weight assessment)											
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	7/101 (6.9%)	56/212 (26.4%)	**RR 3.80** (1.79 to 8.05)	**Study population**	
										69 per 1000	**194 more per 1000** (from 55 more to 489 more)
										Moderate	
										60 per 1000	**168 more per 1000** (from 47 more to 423 more)
Weight gain (Aripiprazole or risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure or study-specific report)											
391 (3 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias,	4/125 (3.2%)	18/266 (6.8%)	**RR 2.43** (0.85 to 6.98)	**Study population**	
										32	**46 more per 1000**

						imprecision, publication bias			**per 1000**	(from 5 fewer to 191 more)
									Moderate	
									26 per 1000	**37 more per 1000** (from 4 fewer to 155 more)
Weight gain (Aripiprazole) (assessed with: Study-specific report)										
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	1/51 (2%) 7/165 (4.2%)	**RR 2.16** (0.27 to 17.17)	**Study population**	
									20 per 1000	**23 more per 1000** (from 14 fewer to 317 more)
									Moderate	
									20 per 1000	**23 more per 1000** (from 15 fewer to 323 more)
Weight gain (Risperidone) (assessed with: Non-systematic assessment or study-specific outcome measure)										
175 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	3/74 (4.1%) 11/101 (10.9%)	**RR 2.55** (0.75 to 8.66)	**Study population**	
									41 per 1000	**63 more per 1000** (from 10 fewer to 311 more)
									Moderate	
									41 per 1000	**64 more per 1000** (from 10 fewer to 314 more)
Weight gain (Aripiprazole or risperidone) (measured with: Weight assessment (in kg); Better indicated by lower values)										
541 (6 studies)	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	reporting bias strongly	⊕⊕⊖⊖ **LOW**[1,3]	206 335	-		The mean weight gain (aripiprazole or risperidone)

6-26 weeks					suspected [3]	due to risk of bias, publication bias				in the intervention groups was **0.69 standard deviations higher** (0.51 to 0.88 higher)
Weight gain (Aripiprazole) (measured with: Weight gain (in kg); Better indicated by lower values)										
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[6]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,6] due to risk of bias, imprecision, publication bias	51	165	-		The mean weight gain (aripiprazole) in the intervention groups was **0.48 standard deviations higher** (0.16 to 0.8 higher)
Weight gain (Risperidone) (measured with: Weight gain (in kg); Better indicated by lower values)										
325 (5 studies) 6-26 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[6]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,6] due to risk of bias, imprecision, publication bias	155	170	-		The mean weight gain (risperidone) in the intervention groups was **0.8 standard deviations higher** (0.57 to 1.03 higher)
BMI change (Aripiprazole) (measured with: BMI change (kg/m-squared); Better indicated by lower values)										
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[7]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,7] due to risk of bias, imprecision, publication bias	51	165	-		The mean bmi change (aripiprazole) in the intervention groups was **0.31 standard deviations higher** (0 to 0.63 higher)
Clinically relevant prolactin elevation (above upper limit of normal for age & gender) (Aripiprazole) (assessed with: Laboratory assessment)										
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	5/101 (5%)	1/212 (0.5%)	**RR 0.19** (0.04 to 0.98)	**Study population** **50 per 1000**	**40 fewer per 1000** (from 1 fewer to 48 fewer)

										Moderate	
										50 per 1000	**41 fewer per 1000** (from 1 fewer to 48 fewer)
Prolactin concentration (ng/ml) (Risperidone) (measured with: Laboratory assessment; Better indicated by lower values)											
124 (2 studies) 8-24 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[6]	undetected	⊕⊕⊖⊖ **LOW**[1,6] due to risk of bias, imprecision	64	60	**-**		The mean prolactin concentration (ng/ml) (risperidone) in the intervention groups was **1.8 standard deviations higher** (1.38 to 2.22 higher)
Any treatment-emergent extrapyramidal symptom (Aripiprazole) (assessed with: Study-specific report of adverse event)											
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	10/101 (9.9%)	44/212 (20.8%)	**RR 1.89** (0.98 to 3.67)	**Study population**	
										99 per 1000	**88 more per 1000** (from 2 fewer to 264 more)
										Moderate	
										99 per 1000	**88 more per 1000** (from 2 fewer to 264 more)
Extrapyramidal symptoms (Risperidone) (measured with: Abnormal Involuntary Movements Scale (AIMS): Total; Better indicated by lower values)											
92 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[6]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,6] due to risk of bias, imprecision, publication bias	34	58	**-**		The mean extrapyramidal symptoms (risperidone) in the intervention groups was **0.46 standard deviations lower** (0.89 to 0.03 lower)
Extrapyramidal disorder (Aripirazole) (assessed with: Study-specific report of adverse event)											

313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	0/101 (0%)	13/212 (6.1%)	**RR 6.02** (0.7 to 51.91)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Fasting glucose (mg/dL) change score (Risperidone) (measured with: Laboratory assessment; Better indicated by lower values)											
68 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[7]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,7] due to risk of bias, imprecision, publication bias	22	46	N/A	N/A	The mean fasting glucose (mg/dl) change score (risperidone) in the intervention groups was **0.02 standard deviations higher** (0.49 lower to 0.53 higher)
Fasting glucose (=>115 mg/dL) - Aripiprazole (assessed with: Laboratory assessment)											
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	0/101 (0%)	2/212 (0.9%)	**RR 1.57** (0.08 to 32.11)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Fasting triglycerides (=>120 mg/dL for females or 160 mg/dL for males) (Aripiprazole) (assessed with: Laboratory assessment)											
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	7/101 (6.9%)	23/212 (10.8%)	**RR 1.8** (0.74 to 4.35)	**Study population**	
										69 per 1000	**55 more per 1000** (from 18 fewer to 232 more)
										Moderate	

									70 per 1000	**56 more per 1000** (from 18 fewer to 234 more)
Insulin resistance (HOMA-IR) change score (Risperidone) (measured with: Laboratory assessment; Better indicated by lower values)										
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[7]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,7] due to risk of bias, imprecision, publication bias	22	43	N/A	N/A	The mean insulin resistance (homa-ir) change score (risperidone) in the intervention groups was **0.12 standard deviations lower** (0.63 lower to 0.4 higher)
Leptin (mg/L) change score (Risperidone) (measured with: Laboratory assessment; Better indicated by lower values)										
104 (2 studies) 8-24 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[6]	undetected	⊕⊕⊖⊖ **LOW**[1,6] due to risk of bias, imprecision	52	52	N/A	N/A	The mean leptin (mg/l) change score (risperidone) in the intervention groups was **0.64 standard deviations higher** (0.24 to 1.04 higher)
Diastolic blood pressure (mm Hg) change scores (Risperidone) (measured with: Physical exam; Better indicated by lower values)										
78 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[7]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,7] due to risk of bias, imprecision, publication bias	38	40	N/A	N/A	The mean diastolic blood pressure (mm hg) change scores (risperidone) in the intervention groups was **0.15 standard deviations higher** (0.29 lower to 0.6 higher)
Systolic blood pressure (mm Hg) change scores (Risperidone) (measured with: Physical exam; Better indicated by lower values)										
78 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[7]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,7] due to risk of bias, imprecision, publication bias	38	40	N/A	N/A	The mean systolic blood pressure (mm hg) change scores (risperidone) in the intervention groups was **0.44 standard deviations**

										higher (0.01 lower to 0.89 higher)
Pulse (bpm) change score (Risperidone) (measured with: Physical exam; Better indicated by lower values)										
78 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[6]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,6] due to risk of bias, imprecision, publication bias	38 40	N/A	N/A	The mean pulse (bpm) change score (risperidone) in the intervention groups was **0.7 standard deviations higher** (0.24 to 1.15 higher)
Somnolence/Drowsiness (Aripiprazole or risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, study-specific report or study-specific side effect checklist)										
588 (5 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	14/226 (6.2%) 82/362 (22.7%)	**RR 4.81** (2.85 to 8.13)	**Study population**	
									62 per 1000	**236 more per 1000** (from 115 more to 442 more)
									Moderate	
									40 per 1000	**152 more per 1000** (from 74 more to 285 more)
Somnolence/Drowsiness (Aripiprazole) (assessed with: Study-specific report of adverse event)										
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	4/101 (4%) 22/212 (10.4%)	**RR 2.98** (1.07 to 8.31)	**Study population**	
									40 per 1000	**78 more per 1000** (from 3 more to 290 more)
									Moderate	
									40 per 1000	**79 more per 1000** (from 3 more to 292 more)

Somnolence/Drowsiness (Risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, or study-specific side effect checklist)										
275 (3 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	10/125 (8%)	60/150 (40%)	**RR 5.71** (3.08 to 10.6)	**Study population** **80 per 1000** | **377 more per 1000** (from 166 more to 768 more) **Moderate** **77 per 1000** | **363 more per 1000** (from 160 more to 739 more)
Fatigue (Aripirazole or risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, study-specific report or study-specific side effect checklist)										
588 (5 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	17/226 (7.5%)	69/362 (19.1%)	**RR 3.16** (1.95 to 5.13)	**Study population** **75 per 1000** | **162 more per 1000** (from 71 more to 311 more) **Moderate** **26 per 1000** | **56 more per 1000** (from 25 more to 107 more)
Fatigue (Aripiprazole) (assessed with: Study-specific report of adverse event)										
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	2/101 (2%)	35/212 (16.5%)	**RR 8.33** (2.11 to 32.9)	**Study population** **20 per 1000** | **145 more per 1000** (from 22 more to 632 more) **Moderate** **20 per** | **147 more per 1000** (from 22 more to 638 more)

									1000	
Fatigue (Risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, or study-specific side effect checklist)										
275 (3 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	15/125 (12%)	34/150 (22.7%)	**RR 2.25** (1.38 to 3.68)	**Study population**
									120 per 1000	**150 more per 1000** (from 46 more to 322 more)
									Moderate	
									26 per 1000	**32 more per 1000** (from 10 more to 70 more)
Lethargy (Aripiprazole) (assessed with: Study-specific report of adverse event)										
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	0/51 (0%)	10/165 (6.1%)	**RR 6.58** (0.39 to 110.35)	**Study population**
									0 per 1000	N/A
									Moderate	
									0 per 1000	N/A
Sedation (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report)										
409 (3 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	4/136 (2.9%)	53/273 (19.4%)	**RR 4.94** (1.94 to 12.58)	**Study population**
									29 per 1000	**116 more per 1000** (from 28 more to 341 more)
									Moderate	
									20 per	**79 more per 1000** (from 19 more to 232 more)

									1000	
Sedation (Aripirazole) (assessed with: Study-specific report of adverse event)										
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	4/101 (4%)	44/212 (20.8%)	**RR 4.25** (1.57 to 11.51)	**Study population** **40 per 1000** **129 more per 1000** (from 23 more to 416 more) **Moderate** **39 per 1000** **127 more per 1000** (from 22 more to 410 more)
Sedation (Risperidone) (assessed with: Non-systematic assessment)										
96 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	0/35 (0%)	9/61 (14.8%)	**RR 11.03** (0.66 to 183.98)	**Study population** **0 per 1000** N/A **Moderate** **0 per 1000** N/A
Upper respiratory tract infection (Aripiprazole or risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, study-specific report or study-specific side effect checklist)										
588 (5 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	14/226 (6.2%)	30/362 (8.3%)	**RR 1.78** (0.97 to 3.25)	**Study population** **62 per 1000** **48 more per 1000** (from 2 fewer to 139 more) **Moderate** **39 per** **30 more per 1000**

									1000	(from 1 fewer to 88 more)
Upper respiratory tract infection (Aripiprazole) (assessed with: Study-specific report of adverse event)										
313 (2 studies) 8 weeks	serious[1]	serious[2]	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,5] due to risk of bias, inconsistency, imprecision, publication bias	5/101 (5%)	6/212 (2.8%)	**RR 0.65** (0.16 to 2.58)	**Study population**
									50 per 1000	**17 fewer per 1000** (from 42 fewer to 78 more)
									Moderate	
									50 per 1000	**18 fewer per 1000** (from 42 fewer to 79 more)
Upper respiratory tract infection (Risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, or study-specific side effect checklist)										
275 (3 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[1,4] due to risk of bias, imprecision	9/125 (7.2%)	24/150 (16%)	**RR 2.45** (1.21 to 4.96)	**Study population**
									72 per 1000	**104 more per 1000** (from 15 more to 285 more)
									Moderate	
									39 per 1000	**57 more per 1000** (from 8 more to 154 more)
Rhinitis/rhinorrhea (Aripiprazole or risperidone) (assessed with: Study-specific outcome measure or study-specific report)										
295 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	5/90 (5.6%)	19/205 (9.3%)	**RR 2.62** (1.02 to 6.77)	**Study population**
									56 per 1000	**90 more per 1000** (from 1 more to 321 more)
									Moderate	
									61	**99 more per 1000**

									per 1000	(from 1 more to 352 more)
Rhinitis/rhinorrhea (Aripiprazole) (assessed with: Study-specific report of adverse event)										
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	1/51 (2%)	8/165 (4.8%)	**RR 2.47** (0.32 to 19.3)	**Study population**
									20 per 1000	**29 more per 1000** (from 13 fewer to 359 more)
									Moderate	
									20 per 1000	**29 more per 1000** (from 14 fewer to 366 more)
Rhinitis/rhinorrhea (Risperidone) (assessed with: Study-specific outcome measure)										
79 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	4/39 (10.3%)	11/40 (27.5%)	**RR 2.68** (0.93 to 7.71)	**Study population**
									103 per 1000	**172 more per 1000** (from 7 fewer to 688 more)
									Moderate	
									103 per 1000	**173 more per 1000** (from 7 fewer to 691 more)
Nasal congestion (Aripirazole or risperidone) (assessed with: Study-specific report or study-specific side effect checklist)										
413 (3 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	22/152 (14.5%)	34/261 (13%)	**RR 1.42** (0.92 to 2.19)	**Study population**
									145 per 1000	**61 more per 1000** (from 12 fewer to 172 more)
									Moderate	

									20 per 1000	**8 more per 1000** (from 2 fewer to 24 more)
Nasal congestion (Aripiprazole) (assessed with: Study-specific report of adverse event)										
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	2/101 (2%)	9/212 (4.2%)	**RR 2.37** (0.52 to 10.77)	**Study population**
									20 per 1000	**27 more per 1000** (from 10 fewer to 193 more)
									Moderate	
									20 per 1000	**27 more per 1000** (from 10 fewer to 195 more)
Nasal congestion (Risperidone) (assessed with: Study-specific side effect checklist)										
100 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	20/51 (39.2%)	25/49 (51%)	**RR 1.3** (0.84 to 2.02)	**Study population**
									392 per 1000	**118 more per 1000** (from 63 fewer to 400 more)
									Moderate	
									392 per 1000	**118 more per 1000** (from 63 fewer to 400 more)
Nasopharyngitis (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report)										
409 (3 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	7/136 (5.1%)	24/273 (8.8%)	**RR 1.65** (0.68 to 3.97)	**Study population**
									51 per 1000	**33 more per 1000** (from 16 fewer to 153 more)
									Moderate	

									57 per 1000	**37 more per 1000** (from 18 fewer to 169 more)
Nasopharyngitis (Aripiprazole) (assessed with: Study-specific report of adverse event)										
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	5/101 (5%)	18/212 (8.5%)	**RR 1.61** (0.55 to 4.71)	**Study population**
									50 per 1000	**30 more per 1000** (from 22 fewer to 184 more)
									Moderate	
									50 per 1000	**31 more per 1000** (from 22 fewer to 186 more)
Nasopharyngitis (Risperidone) (assessed with: Non-systematic assessment)										
96 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	2/35 (5.7%)	6/61 (9.8%)	**RR 1.72** (0.37 to 8.07)	**Study population**
									57 per 1000	**41 more per 1000** (from 36 fewer to 404 more)
									Moderate	
									57 per 1000	**41 more per 1000** (from 36 fewer to 403 more)
Nose bleed (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report)										
312 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	0/86 (0%)	7/226 (3.1%)	**RR 3.2** (0.4 to 25.77)	**Study population**
									0 per 1000	N/A
									Moderate	

									0 per 1000	N/A
Nose bleed (Aripiprazole) (assessed with: Study-specific report of adverse event)										
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	0/51 (0%)	5/165 (3%)	**RR 3.45** (0.19 to 61.28)	**Study population**
									0 per 1000	N/A
									Moderate	
									0 per 1000	N/A
Nose bleed (Risperidone) (assessed with: Non-systematic assessment)										
96 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	0/35 (0%)	2/61 (3.3%)	**RR 2.9** (0.14 to 58.81)	**Study population**
									0 per 1000	N/A
									Moderate	
									0 per 1000	N/A
Coughing (Aripiprazole or risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure or study-specific report)										
391 (3 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	6/125 (4.8%)	18/266 (6.8%)	**RR 1.63** (0.65 to 4.12)	**Study population**
									48 per 1000	**30 more per 1000** (from 17 fewer to 150 more)
									Moderate	
									39 per 1000	**25 more per 1000** (from 14 fewer to 122 more)

Coughing (Aripiprazole) (assessed with: Study-specific report of adverse event)										
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	2/51 (3.9%) 12/165 (7.3%)	**RR 1.85** (0.43 to 8.01)	**Study population**	
									39 per 1000	**33 more per 1000** (from 22 fewer to 275 more)
									Moderate	
									39 per 1000	**33 more per 1000** (from 22 fewer to 273 more)
Coughing (Risperidone) (assessed with: Non-systematic assessment or study-specific outcome measure)										
175 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	4/74 (5.4%) 6/101 (5.9%)	**RR 1.46** (0.45 to 4.79)	**Study population**	
									54 per 1000	**25 more per 1000** (from 30 fewer to 205 more)
									Moderate	
									51 per 1000	**23 more per 1000** (from 28 fewer to 193 more)
Increased appetite (Aripiprazole or risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, study-specific report or study-specific side effect checklist)										
588 (5 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	15/226 (6.6%) 64/362 (17.7%)	**RR 3.01** (1.73 to 5.24)	**Study population**	
									66 per 1000	**133 more per 1000** (from 48 more to 281 more)
									Moderate	
									57 per	**115 more per 1000** (from 42 more to 242 more)

									1000	
Increased appetite (Aripiprazole) (assessed with: Study-specific report of adverse event)										
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	7/101 (6.9%)	27/212 (12.7%)	**RR 2.11** (0.89 to 5.01)	**Study population** **69 per 1000** | **77 more per 1000** (from 8 fewer to 278 more) **Moderate** **70 per 1000** | **78 more per 1000** (from 8 fewer to 281 more)
Increased appetite (Risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, or study-specific side effect checklist)										
275 (3 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	8/125 (6.4%)	37/150 (24.7%)	**RR 3.83** (1.84 to 8.01)	**Study population** **64 per 1000** | **181 more per 1000** (from 54 more to 449 more) **Moderate** **57 per 1000** | **161 more per 1000** (from 48 more to 400 more)
Decreased appetite (Aripiprazole or risperidone) (assessed with: Study-specific report or study-specific side effect checklist)										
316 (2 studies) 8 weeks	serious[1]	serious[2]	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,5] due to risk of bias, inconsistency, imprecision	6/102 (5.9%)	16/214 (7.5%)	**RR 1.43** (0.5 to 4.13)	**Study population** **59 per 1000** | **25 more per 1000** (from 29 fewer to 184 more) **Moderate** **59** | **25 more per 1000**

									per 1000	(from 30 fewer to 185 more)
Decreased appetite (Aripiprazole) (assessed with: Study-specific report of adverse event)										
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	1/51 (2%)	13/165 (7.9%)	**RR 4.02** (0.54 to 29.98)	**Study population**
									20 per 1000	**59 more per 1000** (from 9 fewer to 568 more)
									Moderate	
									20 per 1000	**60 more per 1000** (from 9 fewer to 580 more)
Decreased appetite (Risperidone) (assessed with: Study-specific side effect checklist)										
100 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	5/51 (9.8%)	3/49 (6.1%)	**RR 0.62** (0.16 to 2.47)	**Study population**
									98 per 1000	**37 fewer per 1000** (from 82 fewer to 144 more)
									Moderate	
									98 per 1000	**37 fewer per 1000** (from 82 fewer to 144 more)
Abdominal pain/Stomachache (Aripiprazole or risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, study-specific report or study-specific side effect checklist)										
491 (4 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	13/176 (7.4%)	23/315 (7.3%)	**RR 1.35** (0.69 to 2.64)	**Study population**
									74 per 1000	**26 more per 1000** (from 23 fewer to 121 more)
									Moderate	

									48 per 1000	**17 more per 1000** (from 15 fewer to 79 more)
Abdominal pain/Stomachache (Aripiprazole) (assessed with: Study-specific report of adverse event)										
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	1/51 (2%)	7/165 (4.2%)	**RR 2.16** (0.27 to 17.17)	**Study population**
									20 per 1000	**23 more per 1000** (from 14 fewer to 317 more)
									Moderate	
									20 per 1000	**23 more per 1000** (from 15 fewer to 323 more)
Abdominal pain/Stomachache (Risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, or study-specific side effect checklist)										
275 (3 studies) 6-8 weeks	serious[1]	serious[2]	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2,5] due to risk of bias, inconsistency, imprecision	12/125 (9.6%)	16/150 (10.7%)	**RR 1.25** (0.61 to 2.54)	**Study population**
									96 per 1000	**24 more per 1000** (from 37 fewer to 148 more)
									Moderate	
									77 per 1000	**19 more per 1000** (from 30 fewer to 119 more)
Abdominal discomfort (Risperidone) (assessed with: Non-systematic assessment)										
96 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	3/35 (8.6%)	0/61 (0%)	**RR 0.08** (0 to 1.56)	**Study population**
									86 per 1000	**79 fewer per 1000** (from 86 fewer to 48 more)
									Moderate	

									86 per 1000	**79 fewer per 1000** (from 86 fewer to 48 more)
Vomiting (Aripiprazole or risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, study-specific report or study-specific side effect checklist)										
588 (5 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	26/226 (11.5%)	55/362 (15.2%)	**RR 1.5** (0.97 to 2.34)	**Study population**
									115 per 1000	**58 more per 1000** (from 3 fewer to 154 more)
									Moderate	
									78 per 1000	**39 more per 1000** (from 2 fewer to 105 more)
Vomiting (Aripiprazole) (assessed with: Study-specific report of adverse event)										
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	6/101 (5.9%)	29/212 (13.7%)	**RR 2.19** (0.95 to 5.03)	**Study population**
									59 per 1000	**71 more per 1000** (from 3 fewer to 239 more)
									Moderate	
									59 per 1000	**70 more per 1000** (from 3 fewer to 238 more)
Vomiting (Risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, or study-specific side effect checklist)										
275 (3 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	20/125 (16%)	26/150 (17.3%)	**RR 1.23** (0.74 to 2.07)	**Study population**
									160 per 1000	**37 more per 1000** (from 42 fewer to 171 more)

										Moderate	
										154 per 1000	**35 more per 1000** (from 40 fewer to 165 more)
Nausea (Aripiprazole or risperidone) (assessed with: Non-systematic assessment, study-specific report or study-specific side effect checklist)											
412 (3 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	7/137 (5.1%)	15/275 (5.5%)	**RR 1.3** (0.51 to 3.37)	**Study population**	
										51 per 1000	**15 more per 1000** (from 25 fewer to 121 more)
										Moderate	
										29 per 1000	**9 more per 1000** (from 14 fewer to 69 more)
Nausea (Aripiprazole) (assessed with: Study-specific report of adverse event)											
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	1/51 (2%)	8/165 (4.8%)	**RR 2.47** (0.32 to 19.3)	**Study population**	
										20 per 1000	**29 more per 1000** (from 13 fewer to 359 more)
										Moderate	
										20 per 1000	**29 more per 1000** (from 14 fewer to 366 more)
Nausea (Risperidone) (assessed with: Non-systematic assessment or study-specific side effect checklist)											
196 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	6/86 (7%)	7/110 (6.4%)	**RR 1.02** (0.34 to 3)	**Study population**	
										70 per 1000	**1 more per 1000** (from 46 fewer to 140 more)

									Moderate	
									63 per 1000	**1 more per 1000** (from 42 fewer to 126 more)
Gastroenteritis viral (Aripiprazole) (assessed with: Study-specific report of adverse event)										
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	0/51 (0%)	5/165 (3%)	**RR 3.45** (0.19 to 61.28)	**Study population**
									0 per 1000	N/A
									Moderate	
									0 per 1000	N/A
Constipation (Risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, or study-specific side effect checklist)										
275 (3 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[1,4] due to risk of bias, imprecision	8/125 (6.4%)	21/150 (14%)	**RR 2.53** (1.19 to 5.39)	**Study population**
									64 per 1000	**98 more per 1000** (from 12 more to 281 more)
									Moderate	
									29 per 1000	**44 more per 1000** (from 6 more to 127 more)
Diarrhoea (Aripiprazole or risperidone) (assessed with: Non-systematic assessment, study-specific report or study-specific side effect checklist)										
293 (3 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	17/136 (12.5%)	14/157 (8.9%)	**RR 0.83** (0.43 to 1.59)	**Study population**
									125 per 1000	**21 fewer per 1000** (from 71 fewer to 74 more)

									Moderate	
									64 per 1000	**11 fewer per 1000** (from 36 fewer to 38 more)
Diarrhoea (Aripiprazole) (assessed with: Study-specific report of adverse event)										
97 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	5/50 (10%)	4/47 (8.5%)	**RR 0.85** (0.24 to 2.98)	**Study population**
									100 per 1000	**15 fewer per 1000** (from 76 fewer to 198 more)
									Moderate	
									100 per 1000	**15 fewer per 1000** (from 76 fewer to 198 more)
Diarrhoea (Risperidone) (assessed with: Non-systematic assessment or study-specific side effect checklist)										
196 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	12/86 (14%)	10/110 (9.1%)	**RR 0.82** (0.39 to 1.75)	**Study population**
									140 per 1000	**25 fewer per 1000** (from 85 fewer to 105 more)
									Moderate	
									29 per 1000	**5 fewer per 1000** (from 18 fewer to 22 more)
Fever (Aripiprazole or risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure or study-specific report)										
488 (4 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	8/175 (4.6%)	29/313 (9.3%)	**RR 2.25** (1.04 to 4.87)	**Study population**
									46 per 1000	**57 more per 1000** (from 2 more to 177 more)

										Moderate	
										10 per 1000	**12 more per 1000** (from 0 more to 39 more)
Fever (Aripiprazole) (assessed with: Study-specific report of adverse event)											
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	1/101 (1%)	19/212 (9%)	**RR 6.66** (1.13 to 39.2)	**Study population**	
										10 per 1000	**56 more per 1000** (from 1 more to 378 more)
										Moderate	
										10 per 1000	**57 more per 1000** (from 1 more to 382 more)
Fever (Risperidone) (assessed with: Non-systematic assessment or study-specific outcome measure)											
175 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	7/74 (9.5%)	10/101 (9.9%)	**RR 1.26** (0.53 to 3.02)	**Study population**	
										95 per 1000	**25 more per 1000** (from 44 fewer to 191 more)
										Moderate	
										90 per 1000	**23 more per 1000** (from 42 fewer to 182 more)
Influenza-like symptoms (Risperidone) (assessed with: Study-specific outcome measure)											
79 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	2/39 (5.1%)	4/40 (10%)	**RR 1.95** (0.38 to 10.04)	**Study population**	
										51 per 1000	**49 more per 1000** (from 32 fewer to 464 more)

									Moderate	
									51 per 1000	**48 more per 1000** (from 32 fewer to 461 more)
Insomnia (Aripiprazole or risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, study-specific report or study-specific side effect checklist)										
372 (4 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[1,4] due to risk of bias, imprecision	27/175 (15.4%)	16/197 (8.1%)	**RR 0.59** (0.34 to 1.04)	**Study population**
									154 per 1000	**63 fewer per 1000** (from 102 fewer to 6 more)
									Moderate	
									117 per 1000	**48 fewer per 1000** (from 77 fewer to 5 more)
Insomnia (Aripiprazole) (assessed with: Study-specific report of adverse event)										
97 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	4/50 (8%)	3/47 (6.4%)	**RR 0.8** (0.19 to 3.38)	**Study population**
									80 per 1000	**16 fewer per 1000** (from 65 fewer to 190 more)
									Moderate	
									80 per 1000	**16 fewer per 1000** (from 65 fewer to 190 more)
Insomnia (Risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, or study-specific side effect checklist)										
275 (3 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[1,4] due to risk of bias, imprecision	23/125 (18.4%)	13/150 (8.7%)	**RR 0.56** (0.31 to 1.03)	**Study population**
									184 per	**81 fewer per 1000** (from 127 fewer to 6 more)

										1000	
										Moderate	
										154 per 1000	**68 fewer per 1000** (from 106 fewer to 5 more)
Hypersomnia (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report)											
312 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	1/86 (1.2%)	7/226 (3.1%)	**RR 2.01** (0.33 to 12.16)	**Study population**	
										12 per 1000	**12 more per 1000** (from 8 fewer to 130 more)
										Moderate	
										14 per 1000	**14 more per 1000** (from 9 fewer to 156 more)
Hypersomnia (Aripiprazole) (assessed with: Study-specific report of adverse event)											
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	0/51 (0%)	5/165 (3%)	**RR 3.45** (0.19 to 61.28)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Hypersomnia (Risperidone) (assessed with: Non-systematic assessment)											
96 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision,	1/35 (2.9%)	2/61 (3.3%)	**RR 1.15** (0.11 to 12.2)	**Study population**	
										29 per	**4 more per 1000** (from 25 fewer to 320

						publication bias			**1000**	more)
									Moderate	
									29 per 1000	**4 more per 1000** (from 26 fewer to 325 more)
Sleep problems (Risperidone) (assessed with: Study-specific side effect checklist)										
100 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	9/51 (17.6%) 11/49 (22.4%)	**RR 1.27** (0.58 to 2.8)	**Study population**	
									176 per 1000	**48 more per 1000** (from 74 fewer to 318 more)
									Moderate	
									177 per 1000	**48 more per 1000** (from 74 fewer to 319 more)
Headache (Aripiprazole or risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, study-specific report or study-specific side effect checklist)										
588 (5 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	22/226 (9.7%) 34/362 (9.4%)	**RR 1.1** (0.65 to 1.88)	**Study population**	
									97 per 1000	**10 more per 1000** (from 34 fewer to 86 more)
									Moderate	
									114 per 1000	**11 more per 1000** (from 40 fewer to 100 more)
Headache (Aripiprazole) (assessed with: Study-specific report of adverse event)										
313	serious[1]	very serious[8]	no serious	very serious[5]	reporting bias	⊕⊖⊖⊖	10/101 16/212	**RR 0.85**	**Study population**	

(2 studies) 8 weeks			indirectness		strongly suspected [3]	**VERY LOW**[1,3,5,8] due to risk of bias, inconsistency, imprecision, publication bias	(9.9%)	(7.5%)	(0.35 to 2.07)	**99 per 1000**	**15 fewer per 1000** (from 64 fewer to 106 more)
										Moderate	
										100 per 1000	**15 fewer per 1000** (from 65 fewer to 107 more)
Headache (Risperidone) (assessed with: Non-systematic assessment, study-specific outcome measure, or study-specific side effect checklist)											
275 (3 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	12/125 (9.6%)	18/150 (12%)	**RR 1.31** (0.67 to 2.57)	**Study population**	
										96 per 1000	**30 more per 1000** (from 32 fewer to 151 more)
										Moderate	
										114 per 1000	**35 more per 1000** (from 38 fewer to 179 more)
Dizziness (Risperidone) (assessed with: Study-specific side effect checklist)											
100 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	2/51 (3.9%)	8/49 (16.3%)	**RR 4.16** (0.93 to 18.64)	**Study population**	
										39 per 1000	**124 more per 1000** (from 3 fewer to 692 more)
										Moderate	
										39 per 1000	**123 more per 1000** (from 3 fewer to 688 more)
Increased salivation (Aripiprazole or risperidone) (assessed with: Study-specific outcome measure or study-specific report)											
295	serious[1]	no serious	no serious	very serious[5]	reporting bias	⊕⊖⊖⊖	2/90	15/205	**RR 3.6**	**Study population**	

(2 studies) 8 weeks		inconsistency	indirectness		strongly suspected [3]	**VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	(2.2%)	(7.3%)	(0.82 to 15.82)	**22 per 1000**	**58 more per 1000** (from 4 fewer to 329 more)
										Moderate	
										23 per 1000	**60 more per 1000** (from 4 fewer to 341 more)
Increased salivation (Aripiprazole) (assessed with: Study-specific report of adverse event)											
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	1/51 (2%)	11/165 (6.7%)	**RR 3.4** (0.45 to 25.7)	**Study population**	
										20 per 1000	**47 more per 1000** (from 11 fewer to 484 more)
										Moderate	
										20 per 1000	**48 more per 1000** (from 11 fewer to 494 more)
Increased salivation (Risperidone) (assessed with: Study-specific outcome measure)											
79 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	1/39 (2.6%)	4/40 (10%)	**RR 3.9** (0.46 to 33.36)	**Study population**	
										26 per 1000	**74 more per 1000** (from 14 fewer to 830 more)
										Moderate	
										26 per 1000	**75 more per 1000** (from 14 fewer to 841 more)
Drooling (Aripiprazole or risperidone) (assessed with: Study-specific report or study-specific side effect checklist)											
413	serious[1]	no serious	no serious	serious[4]	undetected	⊕⊕⊖⊖	3/152	32/261	**RR 6.04**	**Study population**	

(3 studies) 8 weeks		inconsistency	indirectness			**LOW**[1,4] due to risk of bias, imprecision	(2%)	(12.3%)	(2.1 to 17.39)	**20 per 1000**	**99 more per 1000** (from 22 more to 323 more)
										Moderate	
										0 per 1000	N/A
Drooling (Aripiprazole) (assessed with: Study-specific report of adverse event)											
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/101 (0%)	19/212 (9%)	**RR 9.65** (1.24 to 74.91)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Drooling (Risperidone) (assessed with: Study-specific side effect checklist)											
100 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[1,4] due to risk of bias, imprecision	3/51 (5.9%)	13/49 (26.5%)	**RR 4.51** (1.37 to 14.86)	**Study population**	
										59 per 1000	**206 more per 1000** (from 22 more to 815 more)
										Moderate	
										59 per 1000	**207 more per 1000** (from 22 more to 818 more)
Dry mouth (Risperidone) (assessed with: Study-specific side effect checklist)											
100 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	5/51 (9.8%)	9/49 (18.4%)	**RR 1.87** (0.68 to 5.2)	**Study population**	
										98 per	**85 more per 1000** (from 31 fewer to 412

									1000	more)
									Moderate	
									98 per 1000	**85 more per 1000** (from 31 fewer to 412 more)
Increased thirst (Aripiprazole or risperidone) (assessed with: Non-systematic assessment, study-specific report or study-specific side effect checklist)										
412 (3 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	6/137 (4.4%)	13/275 (4.7%)	**RR 1.46** (0.57 to 3.74)	**Study population**
									44 per 1000	**20 more per 1000** (from 19 fewer to 120 more)
									Moderate	
									20 per 1000	**9 more per 1000** (from 9 fewer to 55 more)
Increased thirst (Aripiprazole) (assessed with: Study-specific report of adverse event)										
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	1/51 (2%)	5/165 (3%)	**RR 1.55** (0.18 to 12.93)	**Study population**
									20 per 1000	**11 more per 1000** (from 16 fewer to 234 more)
									Moderate	
									20 per 1000	**11 more per 1000** (from 16 fewer to 239 more)
Increased thirst (Risperidone) (assessed with: Non-systematic assessment or study-specific side effect checklist)										
196 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias,	5/86 (5.8%)	8/110 (7.3%)	**RR 1.44** (0.51 to 4.09)	**Study population**
									58	**26 more per 1000**

						imprecision				**per 1000**	(from 28 fewer to 180 more)
										Moderate	
										49 per 1000	**22 more per 1000** (from 24 fewer to 151 more)
Tachycardia (Risperidone) (assessed with: Study-specific outcome measure or study-specific side effect checklist)											
179 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	undetected	⊕⊕⊖⊖ **LOW**[1,4] due to risk of bias, imprecision	1/90 (1.1%)	11/89 (12.4%)	**RR 7.77** (1.45 to 41.72)	**Study population**	
										11 per 1000	**75 more per 1000** (from 5 more to 452 more)
										Moderate	
										10 per 1000	**68 more per 1000** (from 5 more to 407 more)
Anorexia (Risperidone) (assessed with: Study-specific outcome measure)											
79 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	1/39 (2.6%)	4/40 (10%)	**RR 3.9** (0.46 to 33.36)	**Study population**	
										26 per 1000	**74 more per 1000** (from 14 fewer to 830 more)
										Moderate	
										26 per 1000	**75 more per 1000** (from 14 fewer to 841 more)
Anxiety (Risperidone) (assessed with: Study-specific side effect checklist)											
100	serious[1]	no serious	no serious	very serious[5]	undetected	⊕⊖⊖⊖	10/51	12/49	**RR 1.25**	**Study population**	

(1 study) 8 weeks		inconsistency	indirectness			**VERY LOW**[1,5] due to risk of bias, imprecision	(19.6%)	(24.5%)	(0.59 to 2.62)	**196 per 1000**	**49 more per 1000** (from 80 fewer to 318 more)
										Moderate	
										196 per 1000	**49 more per 1000** (from 80 fewer to 318 more)
Depression (Risperidone) (assessed with: Non-systematic assessment)											
96 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	0/35 (0%)	2/61 (3.3%)	**RR 2.9** (0.14 to 58.81)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Apathy (Risperidone) (assessed with: Study-specific outcome measure)											
79 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	0/39 (0%)	5/40 (12.5%)	**RR 10.73** (0.61 to 187.79)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Aggression (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report)											
193 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision,	6/85 (7.1%)	1/108 (0.9%)	**RR 0.2** (0.04 to 1.11)	**Study population**	
										71 per	**56 fewer per 1000** (from 68 fewer to 8 more)

						publication bias			**1000**	
									Moderate	
									69 per 1000	**55 fewer per 1000** (from 66 fewer to 8 more)
Aggression (Aripiprazole) (assessed with: Study-specific report of adverse event)										
97 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	4/50 (8%)	1/47 (2.1%)	**RR 0.27** (0.03 to 2.29)	**Study population**
									80 per 1000	**58 fewer per 1000** (from 78 fewer to 103 more)
									Moderate	
									80 per 1000	**58 fewer per 1000** (from 78 fewer to 103 more)
Aggression (Risperidone) (assessed with: Non-systematic assessment)										
96 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	2/35 (5.7%)	0/61 (0%)	**RR 0.12** (0.01 to 2.35)	**Study population**
									57 per 1000	**50 fewer per 1000** (from 57 fewer to 77 more)
									Moderate	
									57 per 1000	**50 fewer per 1000** (from 56 fewer to 77 more)
Agitation (Risperidone) (assessed with: Non-systematic assessment)										
96 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias,	2/35 (5.7%)	1/61 (1.6%)	**RR 0.29** (0.03 to 3.05)	**Study population**
									57	**41 fewer per 1000**

						imprecision, publication bias			**per 1000**	(from 55 fewer to 117 more)
									Moderate	
									57 per 1000	**40 fewer per 1000** (from 55 fewer to 117 more)
Restlessness (Aripiprazole or risperidone) (assessed with: Non-systematic assessment, study-specific report or study-specific side effect checklist)										
509 (4 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	8/187 (4.3%)	8/322 (2.5%)	**RR 0.63** (0.25 to 1.57)	**Study population**
									43 per 1000	**16 fewer per 1000** (from 32 fewer to 24 more)
									Moderate	
									44 per 1000	**16 fewer per 1000** (from 33 fewer to 25 more)
Restlessness (Aripiprazole) (assessed with: Study-specific report of adverse event)										
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	4/101 (4%)	3/212 (1.4%)	**RR 0.32** (0.08 to 1.32)	**Study population**
									40 per 1000	**27 fewer per 1000** (from 36 fewer to 13 more)
									Moderate	
									39 per 1000	**27 fewer per 1000** (from 36 fewer to 12 more)
Restlessness (Risperidone) (assessed with: Non-systematic assessment or study-specific side effect checklist)										
196	serious[1]	no serious	no serious	very serious[5]	undetected	⊕⊖⊖⊖	4/86	5/110	**RR 1.07**	**Study population**

(2 studies) 6-8 weeks		inconsistency	indirectness			**VERY LOW**[1,5] due to risk of bias, imprecision	(4.7%)	(4.5%)	(0.29 to 3.93)	**47 per 1000**	**3 more per 1000** (from 33 fewer to 136 more)
										Moderate	
										44 per 1000	**3 more per 1000** (from 31 fewer to 129 more)
Psychomotor hyperactivity (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report)											
193 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	4/85 (4.7%)	3/108 (2.8%)	**RR 0.56** (0.13 to 2.47)	**Study population**	
										47 per 1000	**21 fewer per 1000** (from 41 fewer to 69 more)
										Moderate	
										49 per 1000	**22 fewer per 1000** (from 43 fewer to 72 more)
Psychomotor hyperactivity (Aripiprazole) (assessed with: Study-specific report of adverse event)											
97 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	2/50 (4%)	1/47 (2.1%)	**RR 0.53** (0.05 to 5.67)	**Study population**	
										40 per 1000	**19 fewer per 1000** (from 38 fewer to 187 more)
										Moderate	
										40 per 1000	**19 fewer per 1000** (from 38 fewer to 187 more)
Psychomotor hyperactivity (Risperidone) (assessed with: Non-systematic assessment)											
96	serious[1]	no serious	no serious	very serious[5]	reporting bias	⊕⊖⊖⊖	2/35	2/61	**RR 0.57**	**Study population**	

(1 study) 6 weeks		inconsistency	indirectness		strongly suspected [3]	**VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	(5.7%)	(3.3%)	(0.08 to 3.9)	**57 per 1000**	**25 fewer per 1000** (from 53 fewer to 166 more)
										Moderate	
										57 per 1000	**25 fewer per 1000** (from 52 fewer to 165 more)
Tremor (Aripiprazole or risperidone) (assessed with: Study-specific outcome measure, study-specific report or study-specific side effect checklist)											
492 (4 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/191 (0.5%)	32/301 (10.6%)	**RR 8.99** (2.4 to 33.64)	**Study population**	
										5 per 1000	**42 more per 1000** (from 7 more to 171 more)
										Moderate	
										0 per 1000	N/A
Tremor (Aripiprazole) (assessed with: Study-specific report of adverse event)											
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/101 (0%)	21/212 (9.9%)	**RR 10.42** (1.33 to 81.48)	**Study population**	
										0 per 1000	N/A
										Moderate	
										0 per 1000	N/A
Tremor (Risperidone) (assessed with: Study-specific outcome measure or study-specific side effect checklist)											
179 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	undetected	⊕⊕⊖⊖ **LOW**[1,5] due to risk of bias, imprecision	1/90 (1.1%)	11/89 (12.4%)	**RR 7.79** (1.46 to 41.7)	**Study population**	
										11 per	**75 more per 1000** (from 5 more to 452 more)

									1000	
									Moderate	
									10 per 1000	**68 more per 1000** (from 5 more to 407 more)
Dyskinesia/Hyperkinesia (Aripiprazole or risperidone) (assessed with: Study-specific report or study-specific side effect checklist)										
197 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	4/101 (4%)	6/96 (6.3%)	**RR 1.51** (0.47 to 4.82)	**Study population**
									40 per 1000	**20 more per 1000** (from 21 fewer to 151 more)
									Moderate	
									39 per 1000	**20 more per 1000** (from 21 fewer to 149 more)
Dyskinesia/Hyperkinesia (Aripiprazole) (assessed with: Study-specific report of adverse event)										
97 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	1/50 (2%)	0/47 (0%)	**RR 0.35** (0.01 to 8.48)	**Study population**
									20 per 1000	**13 fewer per 1000** (from 20 fewer to 150 more)
									Moderate	
									20 per 1000	**13 fewer per 1000** (from 20 fewer to 150 more)
Dyskinesia/Hyperkinesia (Risperidone) (assessed with: Study-specific side effect checklist)										
100 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias,	3/51 (5.9%)	6/49 (12.2%)	**RR 2.08** (0.55 to 7.87)	**Study population**
									59	**64 more per 1000**

						imprecision			**per 1000**	(from 26 fewer to 404 more)
									Moderate	
									59 per 1000	**64 more per 1000** (from 27 fewer to 405 more)
Hypokinesia (Aripiprazole) (assessed with: Study-specific report of adverse event)										
97 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	0/50 (0%) 1/47 (2.1%)	**RR 3.19** (0.13 to 76.36)	**Study population**	
									0 per 1000	N/A
									Moderate	
									0 per 1000	N/A
Muscle rigidity (Aripiprazole or risperidone) (assessed with: Study-specific report or study-specific side effect checklist)										
197 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	1/101 (1%) 6/96 (6.3%)	**RR 4.54** (0.79 to 26.12)	**Study population**	
									10 per 1000	**35 more per 1000** (from 2 fewer to 249 more)
									Moderate	
									10 per 1000	**35 more per 1000** (from 2 fewer to 251 more)
Muscle rigidity (Aripiprazole) (assessed with: Study-specific report of adverse event)										
97 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias,	0/50 (0%) 1/47 (2.1%)	**RR 3.19** (0.13 to 76.36)	**Study population**	
									0 per	N/A

						imprecision, publication bias				**1000**	
										Moderate	
										0 per 1000	N/A
Muscle rigidity (Risperidone) (assessed with: Study-specific side effect checklist)											
100 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	1/51 (2%)	5/49 (10.2%)	**RR 5.2** (0.63 to 42.96)	**Study population**	
										20 per 1000	**82 more per 1000** (from 7 fewer to 823 more)
										Moderate	
										20 per 1000	**84 more per 1000** (from 7 fewer to 839 more)
Muscle spasms (Aripiprazole) (assessed with: Study-specific report of adverse event)											
97 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	1/50 (2%)	0/47 (0%)	**RR 0.35** (0.01 to 8.48)	**Study population**	
										20 per 1000	**13 fewer per 1000** (from 20 fewer to 150 more)
										Moderate	
										20 per 1000	**13 fewer per 1000** (from 20 fewer to 150 more)
Enuresis (Aripiprazole or risperidone) (assessed with: Non-systematic assessment, study-specific report or study-specific side effect checklist)											
509 (4 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	20/187 (10.7%)	26/322 (8.1%)	**RR 1.14** (0.67 to 1.93)	**Study population**	
										107 per	**15 more per 1000**

									1000	(from 35 fewer to 99 more)
									Moderate	
									50 per 1000	**7 more per 1000** (from 16 fewer to 46 more)
Enuresis (Aripiprazole) (assessed with: Study-specific report of adverse event)										
313 (2 studies) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	5/101 (5%)	7/212 (3.3%)	**RR 0.92** (0.28 to 3.05)	**Study population**
									50 per 1000	**4 fewer per 1000** (from 36 fewer to 101 more)
									Moderate	
									50 per 1000	**4 fewer per 1000** (from 36 fewer to 102 more)
Enuresis (Risperidone) (assessed with: Non-systematic assessment or study-specific side effect checklist)										
196 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	15/86 (17.4%)	19/110 (17.3%)	**RR 1.21** (0.68 to 2.18)	**Study population**
									174 per 1000	**37 more per 1000** (from 56 fewer to 206 more)
									Moderate	
									147 per 1000	**31 more per 1000** (from 47 fewer to 173 more)
Skin irritation/Rash (Aripiprazole or risperidone) (assessed with: Non-systematic assessment, study-specific report or study-specific side effect checklist)										
412 (3 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias,	8/137 (5.8%)	17/275 (6.2%)	**RR 1.66** (0.76 to 3.6)	**Study population**
									58	**39 more per 1000**

						imprecision			**per 1000**	(from 14 fewer to 152 more)
									Moderate	
									20 per 1000	**13 more per 1000** (from 5 fewer to 52 more)
Skin irritation/Rash (Aripiprazole) (assessed with: Study-specific report of adverse event)										
216 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,5] due to risk of bias, imprecision, publication bias	1/51 (2%) 4/165 (2.4%)	**RR 1.24** (0.14 to 10.81)	**Study population**	
									20 per 1000	**5 more per 1000** (from 17 fewer to 192 more)
									Moderate	
									20 per 1000	**5 more per 1000** (from 17 fewer to 196 more)
Skin irritation/Rash (Risperidone) (assessed with: Non-systematic assessment or study-specific side effect checklist)										
196 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	7/86 (8.1%) 13/110 (11.8%)	**RR 1.74** (0.76 to 4.01)	**Study population**	
									81 per 1000	**60 more per 1000** (from 20 fewer to 245 more)
									Moderate	
									69 per 1000	**51 more per 1000** (from 17 fewer to 208 more)
Earache/Ear infection (Risperidone) (assessed with: Non-systematic assessment or study-specific side effect checklist)										
196	serious[1]	no serious	no serious	very serious[5]	undetected	⊕⊖⊖⊖	4/86 4/110	**RR 0.85**	**Study population**	

(2 studies) 6-8 weeks		inconsistency	indirectness			**VERY LOW**[1,5] due to risk of bias, imprecision	(4.7%)	(3.6%)	(0.22 to 3.3)	**47 per 1000**	**7 fewer per 1000** (from 36 fewer to 107 more)
										Moderate	
										39 per 1000	**6 fewer per 1000** (from 30 fewer to 90 more)
Sore throat (Risperidone) (assessed with: Study-specific side effect checklist)											
100 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[5]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,5] due to risk of bias, imprecision	1/51 (2%)	5/49 (10.2%)	**RR 5.2** (0.63 to 42.96)	**Study population**	
										20 per 1000	**82 more per 1000** (from 7 fewer to 823 more)
										Moderate	
										20 per 1000	**84 more per 1000** (from 7 fewer to 839 more)

[1] High risk of detection bias as unclear if follow-up duration (=<12 weeks) is sufficient to observe potential longer term adverse eventsand reliability/validity of some outcome measures unclear
[2] I-squared value indicates moderate heterogeneity
[3] Trial funded by pharmaceutical company and/or study drugs were provided by pharmaceutical company and/or authors are consultants to pharmaceutical companies
[4] Events<300
[5] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)
[6] N<400
[7] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[8] I-squared value indicates substantial heterogeneity

Adverse events associated with low dose antipsychotics versus placebo

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 Adverse events associated with low dose antipsychotics versus placebo		Risk with Control	Risk difference with Adverse events associated with low dose antipsychotics versus placebo (95% CI)
Any side effect (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)											
168 (2 studies) 6-8 weeks	serious[1]	very serious[2]	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,2,3,4] due to risk of bias, inconsistency, imprecision, publication bias	58/86 (67.4%)	58/82 (70.7%)	**RR 1.03** (0.84 to 1.26)	**Study population**	
										674 per 1000	**20 more per 1000** (from 108 fewer to 175 more)
										Moderate	
										663 per 1000	**20 more per 1000** (from 106 fewer to 172 more)
Any side effect (Aripiprazole 5mg/day) (assessed with: study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,4,5] due to risk of bias, imprecision, publication bias	37/51 (72.5%)	46/52 (88.5%)	**RR 1.22** (1 to 1.48)	**Study population**	
										725 per 1000	**160 more per 1000** (from 0 more to 348 more)
										Moderate	
										726 per 1000	**160 more per 1000** (from 0 more to 348 more)
Any side effect (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)											
65	serious[1]	no serious	no serious	very	reporting bias	⊕⊖⊖⊖	21/35	12/30	**RR 0.67**	**Study population**	

(1 study) 6 weeks		inconsistency	indirectness	serious[3]	strongly suspected [4]	**VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	(60%)	(40%)	(0.4 to 1.12)	**600 per 1000**	**198 fewer per 1000** (from 360 fewer to 72 more)
										Moderate	
										600 per 1000	**198 fewer per 1000** (from 360 fewer to 72 more)
Discontinuation due to sedation (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/51 (0%)	1/52 (1.9%)	**RR 2.94** (0.12 to 70.61)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**
Discontinuation due to drooling (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/51 (0%)	1/52 (1.9%)	**RR 2.94** (0.12 to 70.61)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**
Discontinuation due to tremor (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103	serious[1]	no serious	no serious	very	reporting bias	⊕⊖⊖⊖	0/51	2/52	**RR 4.91**	**Study population**	

(1 study) 8 weeks		inconsistency	indirectness	serious[3]	strongly suspected [4]	**VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	(0%)	(3.8%)	(0.24 to 99.74)	**0 per 1000**	-
										Moderate	
										0 per 1000	-
Any treatment-emergent extrapyramidal symptoms (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	6/51 (11.8%)	12/52 (23.1%)	**RR 1.96** (0.8 to 4.83)	**Study population**	
										118 per 1000	**113 more per 1000** (from 24 fewer to 451 more)
										Moderate	
										118 per 1000	**113 more per 1000** (from 24 fewer to 452 more)
Extrapyramidal symptoms (Risperidone 0.125-0.175mg/day) (measured with: Non-systematic assessment; Better indicated by lower values)											
63 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,4,6] due to risk of bias, imprecision, publication bias	34	29	-		The mean extrapyramidal symptoms (risperidone 0.125-0.175mg/day) in the intervention groups was **0.37 standard deviations lower** (0.87 lower to 0.13 higher)
Extrapyramidal disorder (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study)	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias,	0/51 (0%)	2/52 (3.8%)	**RR 4.91** (0.24 to	**Study population**	
										0 per	-

8 weeks					suspected [4]	imprecision, publication bias			99.74)	**1000**	
										Moderate	
										0 per 1000	**-**
Tremor (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/51 (0%)	4/52 (7.7%)	**RR 8.83** (0.49 to 159.93)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**
Clinically relevant (>=7%) weight gain (Aripiprazole 5mg/day) (assessed with: Weight assessment)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,4,5] due to risk of bias, imprecision, publication bias	4/51 (7.8%)	17/52 (32.7%)	**RR 4.17** (1.51 to 11.54)	**Study population**	
										78 per 1000	**249 more per 1000** (from 40 more to 827 more)
										Moderate	
										78 per 1000	**247 more per 1000** (from 40 more to 822 more)
Weight gain (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)											
168 (2 studies)	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias,	3/86 (3.5%)	7/82 (8.5%)	**RR 2.52** (0.67 to	**Study population**	
										35 per	**53 more per 1000**

6-8 weeks					suspected [4]	imprecision, publication bias			9.51)	**1000**	(from 12 fewer to 297 more)
										Moderate	
										38 per 1000	**58 more per 1000** (from 13 fewer to 323 more)
Weight gain (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/51 (2%)	4/52 (7.7%)	**RR 3.92** (0.45 to 33.92)	**Study population**	
										20 per 1000	**57 more per 1000** (from 11 fewer to 645 more)
										Moderate	
										20 per 1000	**58 more per 1000** (from 11 fewer to 658 more)
Weight gain (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)											
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	2/35 (5.7%)	3/30 (10%)	**RR 1.75** (0.31 to 9.79)	**Study population**	
										57 per 1000	**43 more per 1000** (from 39 fewer to 502 more)
										Moderate	
										57 per 1000	**43 more per 1000** (from 39 fewer to 501 more)

Weight gain (in kg) (Aripiprazole or risperidone) (measured with: Weight assessment; Better indicated by lower values)										
160 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[7]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,4,7] due to risk of bias, imprecision, publication bias	84 76	-		The mean weight gain (in kg) (aripiprazole or risperidone) in the intervention groups was **0.45 standard deviations higher** (0.13 to 0.76 higher)
Weight gain (in kg) - Aripiprazole (5mg/day) (measured with: Weight assessment; Better indicated by lower values)										
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[7]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,4,7] due to risk of bias, imprecision, publication bias	51 52	-		The mean weight gain (in kg) - aripiprazole (5mg/day) in the intervention groups was **0.46 standard deviations higher** (0.07 to 0.85 higher)
Weight gain (in kg) - Risperidone (0.125-0.175mg/day) (measured with: Weight assessment; Better indicated by lower values)										
57 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,4,6] due to risk of bias, imprecision, publication bias	33 24	-		The mean weight gain (in kg) - risperidone (0.125-0.175mg/day) in the intervention groups was **0.42 standard deviations higher** (0.11 lower to 0.96 higher)
BMI change (kg/m-squared) - Aripiprazole (5mg/day) (measured with: Weight assessment; Better indicated by lower values)										
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,4,6] due to risk of bias, imprecision,	51 52	-		The mean bmi change (kg/m-squared) - aripiprazole (5mg/day) in the intervention groups was

						publication bias				**0.28 standard deviations higher** (0.11 lower to 0.66 higher)
Increased appetite (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)										
168 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,4,5] due to risk of bias, imprecision, publication bias	4/86 (4.7%)	15/82 (18.3%)	**RR 3.95** (1.36 to 11.51)	**Study population**
										47 per 1000 / **137 more per 1000** (from 17 more to 489 more)
										Moderate
										48 per 1000 / **142 more per 1000** (from 17 more to 504 more)
Increased appetite (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)										
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,4,5] due to risk of bias, imprecision, publication bias	2/51 (3.9%)	10/52 (19.2%)	**RR 4.9** (1.13 to 21.29)	**Study population**
										39 per 1000 / **153 more per 1000** (from 5 more to 796 more)
										Moderate
										39 per 1000 / **152 more per 1000** (from 5 more to 791 more)
Increased appetite (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)										
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision,	2/35 (5.7%)	5/30 (16.7%)	**RR 2.92** (0.61 to 13.96)	**Study population**
										57 per 1000 / **110 more per 1000** (from 22 fewer to 741 more)

						publication bias				**Moderate**	
									57 per 1000	**109 more per 1000** (from 22 fewer to 739 more)	
Decreased appetite (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/51 (2%)	5/52 (9.6%)	**RR 4.9** (0.59 to 40.53)	**Study population**	
										20 per 1000	**76 more per 1000** (from 8 fewer to 775 more)
										Moderate	
										20 per 1000	**78 more per 1000** (from 8 fewer to 791 more)
Fasting Glucose (mg/dL) (Change Score) - Risperidone (0.125-0.175mg/day) (measured with: Laboratory assessment; Better indicated by lower values)											
45 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,4,6] due to risk of bias, imprecision, publication bias	22	23	**-**		The mean fasting glucose (mg/dl) (change score) - risperidone (0.125-0.175mg/day) in the intervention groups was **0.03 standard deviations higher** (0.55 lower to 0.62 higher)
Fasting glucose (=>115 mg/dL) - Aripiprazole (5mg/day) (assessed with: Laboratory assessment)											
103 (1 study) 8 weeks						See comment	0/51 (0%)	0/52 (0%)	**not pooled**	See comment	See comment

Fasting triglycerides (=>120 mg/dL for females or 160 mg/dL for males) - Aripiprazole (5mg/day) (assessed with: Laboratory assessment)										
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	2/51 (3.9%)	6/52 (11.5%)	**RR 2.94** (0.62 to 13.9)	**Study population** **39 per 1000** / **76 more per 1000** (from 15 fewer to 506 more) **Moderate** **39 per 1000** / **76 more per 1000** (from 15 fewer to 503 more)
Insulin Resistance (HOMA-IR) (Change Score) - Risperidone (0.125-0.175mg/day) (measured with: Laboratory assessment; Better indicated by lower values)										
43 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,4,6] due to risk of bias, imprecision, publication bias	22	21	-	The mean insulin resistance (homa-ir) (change score) - risperidone (0.125-0.175mg/day) in the intervention groups was **0.3 standard deviations lower** (0.9 lower to 0.3 higher)
Aggression (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)										
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	2/35 (5.7%)	0/30 (0%)	**RR 0.23** (0.01 to 4.66)	**Study population** **57 per 1000** / **44 fewer per 1000** (from 57 fewer to 209 more) **Moderate**

									57 per 1000	**44 fewer per 1000** (from 56 fewer to 209 more)
Agitation (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)										
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	2/35 (5.7%)	0/30 (0%)	**RR 0.23** (0.01 to 4.66)	**Study population**
									57 per 1000	**44 fewer per 1000** (from 57 fewer to 209 more)
									Moderate	
									57 per 1000	**44 fewer per 1000** (from 56 fewer to 209 more)
Depression (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)										
65 (1 study) 6 weeks	[1]					See comment	0/35 (0%)	0/30 (0%)	**not pooled**	See comment	See comment
Abdominal discomfort (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)										
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	3/35 (8.6%)	0/30 (0%)	**RR 0.17** (0.01 to 3.09)	**Study population**
									86 per 1000	**71 fewer per 1000** (from 85 fewer to 179 more)
									Moderate	
									86 per 1000	**71 fewer per 1000** (from 85 fewer to 180 more)

Abdominal pain (upper) (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)											
168 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/86 (1.2%)	3/82 (3.7%)	**RR 2.44** (0.37 to 15.99)	**Study population**	
										12 per 1000	**17 more per 1000** (from 7 fewer to 174 more)
										Moderate	
										10 per 1000	**14 more per 1000** (from 6 fewer to 150 more)
Abdominal pain (upper) - Aripiprazole (5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/51 (2%)	2/52 (3.8%)	**RR 1.96** (0.18 to 20.97)	**Study population**	
										20 per 1000	**19 more per 1000** (from 16 fewer to 392 more)
										Moderate	
										20 per 1000	**19 more per 1000** (from 16 fewer to 399 more)
Abdominal pain (upper) - Risperidone (0.125-0.175mg/day) (assessed with: Non-systematic assessment)											
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/35 (0%)	1/30 (3.3%)	**RR 3.48** (0.15 to 82.48)	**Study population**	
										0 per 1000	**-**
										Moderate	

									0 per 1000	**-**
Constipation (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)										
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/35 (2.9%)	0/30 (0%)	**RR 0.39** (0.02 to 9.16)	**Study population**
									29 per 1000	**17 fewer per 1000** (from 28 fewer to 233 more)
									Moderate	
									29 per 1000	**18 fewer per 1000** (from 28 fewer to 237 more)
Nausea (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)										
168 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	2/86 (2.3%)	2/82 (2.4%)	**RR 1.07** (0.15 to 7.39)	**Study population**
									23 per 1000	**2 more per 1000** (from 20 fewer to 149 more)
									Moderate	
									24 per 1000	**2 more per 1000** (from 20 fewer to 153 more)
Nausea (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)										
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision,	1/51 (2%)	1/52 (1.9%)	**RR 0.98** (0.06 to 15.26)	**Study population**
									20 per 1000	**0 fewer per 1000** (from 18 fewer to 280

						publication bias					more)
										Moderate	
										20 per 1000	**0 fewer per 1000** (from 19 fewer to 285 more)
Nausea (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)											
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/35 (2.9%)	1/30 (3.3%)	**RR 1.17** (0.08 to 17.86)	**Study population**	
										29 per 1000	**5 more per 1000** (from 26 fewer to 482 more)
										Moderate	
										29 per 1000	**5 more per 1000** (from 27 fewer to 489 more)
Vomiting (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)											
168 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	6/86 (7%)	7/82 (8.5%)	**RR 1.21** (0.42 to 3.44)	**Study population**	
										70 per 1000	**15 more per 1000** (from 40 fewer to 170 more)
										Moderate	
										68 per 1000	**14 more per 1000** (from 39 fewer to 166 more)
Vomiting (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											

103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	4/51 (7.8%)	5/52 (9.6%)	**RR 1.23** (0.35 to 4.31)	**Study population**	
										78 per 1000	**18 more per 1000** (from 51 fewer to 260 more)
										Moderate	
										78 per 1000	**18 more per 1000** (from 51 fewer to 258 more)
Vomiting (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)											
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	2/35 (5.7%)	2/30 (6.7%)	**RR 1.17** (0.17 to 7.79)	**Study population**	
										57 per 1000	**10 more per 1000** (from 47 fewer to 388 more)
										Moderate	
										57 per 1000	**10 more per 1000** (from 47 fewer to 387 more)
Gastroenteritis viral (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/51 (0%)	1/52 (1.9%)	**RR 2.94** (0.12 to 70.61)	**Study population**	
										0 per 1000	-
										Moderate	
										0 per 1000	-

Diarrhoea (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)											
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/35 (2.9%)	1/30 (3.3%)	**RR 1.17** (0.08 to 17.86)	**Study population**	
										29 per 1000	**5 more per 1000** (from 26 fewer to 482 more)
										Moderate	
										29 per 1000	**5 more per 1000** (from 27 fewer to 489 more)
Pyrexia (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)											
168 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/86 (0%)	3/82 (3.7%)	**RR 6.87** (0.36 to 129.7)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**
Pyrexia (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/51 (0%)	3/52 (5.8%)	**RR 6.87** (0.36 to 129.7)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**

Pyrexia (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)											
65 (1 study) 6 weeks						See comment	0/35 (0%)	0/30 (0%)	**not pooled**	See comment	See comment
Drooling (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/51 (0%)	2/52 (3.8%)	**RR 4.91** (0.24 to 99.74)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**
Increased salivation (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/51 (2%)	1/52 (1.9%)	**RR 0.98** (0.06 to 15.26)	**Study population**	
										20 per 1000	**0 fewer per 1000** (from 18 fewer to 280 more)
										Moderate	
										20 per 1000	**0 fewer per 1000** (from 19 fewer to 285 more)
Thirst (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)											
168 (2 studies)	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias,	1/86 (1.2%)	3/82 (3.7%)	**RR 2.94** (0.32 to	**Study population**	
										12 per	**23 more per 1000**

6-8 weeks					suspected [4]	imprecision, publication bias		27.36)	**1000**	(from 8 fewer to 307 more)
									Moderate	
									10 per 1000	**19 more per 1000** (from 7 fewer to 264 more)
Thirst (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)										
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/51 (2%) 3/52 (5.8%)	**RR 2.94** (0.32 to 27.36)	**Study population**	
									20 per 1000	**38 more per 1000** (from 13 fewer to 517 more)
									Moderate	
									20 per 1000	**39 more per 1000** (from 14 fewer to 527 more)
Thirst (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)										
65 (1 study) 6 weeks						See comment	0/35 (0%) 0/30 (0%)	**not pooled**	See comment	See comment
Fatigue (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)										
168 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/86 (0%) 2/82 (2.4%)	**RR 4.91** (0.24 to 99.74)	**Study population**	
									0 per 1000	-
									Moderate	

										0 per 1000	**-**
Fatigue (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/51 (0%)	2/52 (3.8%)	**RR 4.91** (0.24 to 99.74)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**
Fatigue (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)											
65 (1 study) 6 weeks	no serious risk of bias					See comment	0/35 (0%)	0/30 (0%)	**not pooled**	See comment	See comment
Lethargy (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/51 (0%)	4/52 (7.7%)	**RR 8.83** (0.49 to 159.93)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**
Somnolence (Aripirazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)											

168 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	3/86 (3.5%)	4/82 (4.9%)	**RR 1.32** (0.33 to 5.26)	**Study population**	
										35 per 1000	**11 more per 1000** (from 23 fewer to 149 more)
										Moderate	
										34 per 1000	**11 more per 1000** (from 23 fewer to 145 more)
Somnolence (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	2/51 (3.9%)	4/52 (7.7%)	**RR 1.96** (0.38 to 10.24)	**Study population**	
										39 per 1000	**38 more per 1000** (from 24 fewer to 362 more)
										Moderate	
										39 per 1000	**37 more per 1000** (from 24 fewer to 360 more)
Somnolence (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)											
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/35 (2.9%)	0/30 (0%)	**RR 0.39** (0.02 to 9.16)	**Study population**	
										29 per 1000	**17 fewer per 1000** (from 28 fewer to 233 more)
										Moderate	
										29 per 1000	**18 fewer per 1000** (from 28 fewer to 237

										more)
Sedation (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse)										
168 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	3/86 (3.5%)	10/82 (12.2%)	**RR 3.01** (0.94 to 9.62)	**Study population** **35 per 1000** | **70 more per 1000** (from 2 fewer to 301 more) **Moderate** **29 per 1000** | **58 more per 1000** (from 2 fewer to 250 more)
Sedation (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)										
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	3/51 (5.9%)	9/52 (17.3%)	**RR 2.94** (0.84 to 10.25)	**Study population** **59 per 1000** | **114 more per 1000** (from 9 fewer to 544 more) **Moderate** **59 per 1000** | **114 more per 1000** (from 9 fewer to 546 more)
Sedation (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)										
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision,	0/35 (0%)	1/30 (3.3%)	**RR 3.48** (0.15 to 82.48)	**Study population** **0 per 1000** | **-**

						publication bias			**Moderate**	
									0 per 1000	**-**
Headache (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)										
168 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	6/86 (7%)	5/82 (6.1%)	**RR 0.9** (0.28 to 2.86)	**Study population**
									70 per 1000	**7 fewer per 1000** (from 50 fewer to 130 more)
									Moderate	
									77 per 1000	**8 fewer per 1000** (from 55 fewer to 143 more)
Headache (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)										
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	2/51 (3.9%)	3/52 (5.8%)	**RR 1.47** (0.26 to 8.44)	**Study population**
									39 per 1000	**18 more per 1000** (from 29 fewer to 292 more)
									Moderate	
									39 per 1000	**18 more per 1000** (from 29 fewer to 290 more)
Headache (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)										
65 (1 study)	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias,	4/35 (11.4%)	2/30 (6.7%)	**RR 0.58** (0.11 to	**Study population**
									114 per	**48 fewer per 1000**

6 weeks					suspected [4]	imprecision, publication bias			2.96)	**1000**	(from 102 fewer to 224 more)
										Moderate	
										114 per 1000	**48 fewer per 1000** (from 101 fewer to 223 more)
Ear infection (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)											
65 (1 study) 6 weeks						See comment	0/35 (0%)	0/30 (0%)	**not pooled**	See comment	See comment
Upper respiratory tract infection (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)											
168 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/86 (1.2%)	3/82 (3.7%)	**RR 2.49** (0.36 to 17.01)	**Study population**	
										12 per 1000	**17 more per 1000** (from 7 fewer to 186 more)
										Moderate	
										14 per 1000	**21 more per 1000** (from 9 fewer to 224 more)
Upper respiratory tract infection (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/51 (0%)	2/52 (3.8%)	**RR 4.91** (0.24 to 99.74)	**Study population**	
										0 per 1000	-
										Moderate	

									0 per 1000	**-**
Upper respiratory tract infection (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)										
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/35 (2.9%)	1/30 (3.3%)	**RR 1.17** (0.08 to 17.86)	**Study population**
									29 per 1000	**5 more per 1000** (from 26 fewer to 482 more)
									Moderate	
									29 per 1000	**5 more per 1000** (from 27 fewer to 489 more)
Cough (Aripirazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)										
168 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	2/86 (2.3%)	8/82 (9.8%)	**RR 3.92** (0.87 to 17.59)	**Study population**
									23 per 1000	**68 more per 1000** (from 3 fewer to 386 more)
									Moderate	
									20 per 1000	**58 more per 1000** (from 3 fewer to 332 more)
Cough (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)										
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision,	2/51 (3.9%)	8/52 (15.4%)	**RR 3.92** (0.87 to 17.59)	**Study population**
									39 per 1000	**115 more per 1000** (from 5 fewer to 651

						publication bias				more)	
									Moderate		
									39 per 1000	**114 more per 1000** (from 5 fewer to 647 more)	
Cough (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)											
65 (1 study) 6 weeks						See comment	0/35 (0%)	0/30 (0%)	**not pooled**	See comment	See comment
Rhinorrhea (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/51 (2%)	2/52 (3.8%)	**RR 1.96** (0.18 to 20.97)	**Study population**	
										20 per 1000	**19 more per 1000** (from 16 fewer to 392 more)
										Moderate	
										20 per 1000	**19 more per 1000** (from 16 fewer to 399 more)
Nasal congestion (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/51 (2%)	1/52 (1.9%)	**RR 0.98** (0.06 to 15.26)	**Study population**	
										20 per 1000	**0 fewer per 1000** (from 18 fewer to 280 more)
										Moderate	

									20 per 1000	**0 fewer per 1000** (from 19 fewer to 285 more)
Nasopharyngitis (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)										
168 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	4/86 (4.7%)	8/82 (9.8%)	**RR 2.09** (0.65 to 6.79)	**Study population**
									47 per 1000	**51 more per 1000** (from 16 fewer to 269 more)
									Moderate	
									48 per 1000	**52 more per 1000** (from 17 fewer to 278 more)
Nasopharyngitis (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)										
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	2/51 (3.9%)	6/52 (11.5%)	**RR 2.94** (0.62 to 13.9)	**Study population**
									39 per 1000	**76 more per 1000** (from 15 fewer to 506 more)
									Moderate	
									39 per 1000	**76 more per 1000** (from 15 fewer to 503 more)
Nasopharyngitis (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)										
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision,	2/35 (5.7%)	2/30 (6.7%)	**RR 1.17** (0.17 to 7.79)	**Study population**
									57 per	**10 more per 1000** (from 47 fewer to 388

						publication bias				**1000**	more)
										Moderate	
										57 per 1000	**10 more per 1000** (from 47 fewer to 387 more)
Nose bleed (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)											
168 (2 studies) 6-8 weeks						See comment	0/86 (0%)	0/82 (0%)	**not pooled**	See comment	See comment
Nose bleed (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks						See comment	0/51 (0%)	0/52 (0%)	**not pooled**	See comment	See comment
Nose bleed (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)											
65 (1 study) 6 weeks						See comment	0/35 (0%)	0/30 (0%)	**not pooled**	See comment	See comment
Akathisia (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)											
168 (2 studies) 6-8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	4/86 (4.7%)	1/82 (1.2%)	**RR 0.35** (0.06 to 2.14)	**Study population**	
										47 per 1000	**30 fewer per 1000** (from 44 fewer to 53 more)
										Moderate	

									44 per 1000	**29 fewer per 1000** (from 41 fewer to 50 more)
Akathisia (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)										
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	3/51 (5.9%)	1/52 (1.9%)	**RR 0.33** (0.04 to 3.04)	**Study population**
									59 per 1000	**39 fewer per 1000** (from 56 fewer to 120 more)
									Moderate	
									59 per 1000	**40 fewer per 1000** (from 57 fewer to 120 more)
Akathisia (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)										
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/35 (2.9%)	0/30 (0%)	**RR 0.39** (0.02 to 9.16)	**Study population**
									29 per 1000	**17 fewer per 1000** (from 28 fewer to 233 more)
									Moderate	
									29 per 1000	**18 fewer per 1000** (from 28 fewer to 237 more)
Insomnia (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)										
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision,	2/35 (5.7%)	0/30 (0%)	**RR 0.23** (0.01 to 4.66)	**Study population**
									57 per	**44 fewer per 1000** (from 57 fewer to 209

						publication bias				**1000**	more)
										Moderate	
										57 per 1000	**44 fewer per 1000** (from 56 fewer to 209 more)
Hypersomnia (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)											
168 (2 studies) 6-8 weeks	serious[1]	serious[8]	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4,8] due to risk of bias, inconsistency, imprecision, publication bias	1/86 (1.2%)	3/82 (3.7%)	**RR 2.12** (0.38 to 11.88)	**Study population**	
										12 per 1000	**13 more per 1000** (from 7 fewer to 127 more)
										Moderate	
										14 per 1000	**16 more per 1000** (from 9 fewer to 152 more)
Hypersomnia (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/51 (0%)	3/52 (5.8%)	**RR 6.87** (0.36 to 129.7)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**
Hypersomnia (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)											
65	serious[1]	no serious	no serious	very	reporting bias	⊕⊖⊖⊖	1/35	0/30	**RR 0.39**	**Study population**	

(1 study) 6 weeks		inconsistency	indirectness	serious[3]	strongly suspected [4]	**VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	(2.9%)	(0%)	(0.02 to 9.16)	**29 per 1000**	**17 fewer per 1000** (from 28 fewer to 233 more)
										Moderate	
										29 per 1000	**18 fewer per 1000** (from 28 fewer to 237 more)
Psychomotor hyperactivity (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)											
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	2/35 (5.7%)	1/30 (3.3%)	**RR 0.58** (0.06 to 6.12)	**Study population**	
										57 per 1000	**24 fewer per 1000** (from 54 fewer to 293 more)
										Moderate	
										57 per 1000	**24 fewer per 1000** (from 54 fewer to 292 more)
Enuresis (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)											
168 (2 studies) 6-8 weeks	serious[1]	serious[8]	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4,8] due to risk of bias, inconsistency, imprecision, publication bias	1/86 (1.2%)	2/82 (2.4%)	**RR 1.61** (0.29 to 9.04)	**Study population**	
										12 per 1000	**7 more per 1000** (from 8 fewer to 93 more)
										Moderate	
										10 per 1000	**6 more per 1000** (from 7 fewer to 80 more)
Enuresis (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											

103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/51 (2%)	0/52 (0%)	**RR 0.33** (0.01 to 7.85)	**Study population**	
										20 per 1000	**13 fewer per 1000** (from 19 fewer to 134 more)
										Moderate	
										20 per 1000	**13 fewer per 1000** (from 20 fewer to 137 more)
Enuresis (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)											
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/35 (0%)	2/30 (6.7%)	**RR 5.81** (0.29 to 116.41)	**Study population**	
										0 per 1000	-
										Moderate	
										0 per 1000	-
Rash (Aripiprazole or risperidone) (assessed with: Non-systematic assessment or study-specific report of adverse event)											
168 (2 studies) 6-8 weeks	serious[1]	serious[8]	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4,8] due to risk of bias, inconsistency, imprecision, publication bias	1/86 (1.2%)	2/82 (2.4%)	**RR 1.61** (0.29 to 9.04)	**Study population**	
										12 per 1000	**7 more per 1000** (from 8 fewer to 93 more)
										Moderate	
										10 per 1000	**6 more per 1000** (from 7 fewer to 80 more)
Rash (Aripiprazole 5mg/day) (assessed with: Study-specific report of adverse event)											

103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	1/51 (2%)	0/52 (0%)	**RR 0.33** (0.01 to 7.85)	**Study population**	
										20 per 1000	**13 fewer per 1000** (from 19 fewer to 134 more)
										Moderate	
										20 per 1000	**13 fewer per 1000** (from 20 fewer to 137 more)
Rash (Risperidone 0.125-0.175mg/day) (assessed with: Non-systematic assessment)											
65 (1 study) 6 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	0/35 (0%)	2/30 (6.7%)	**RR 5.81** (0.29 to 116.41)	**Study population**	
										0 per 1000	-
										Moderate	
										0 per 1000	-
Clinically relevant prolactin elevation (above upper limit of normal) - Aripiprazole (5mg/day) (assessed with: Study-specific report of adverse event)											
103 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	reporting bias strongly suspected [4]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	2/51 (3.9%)	0/52 (0%)	**RR 0.2** (0.01 to 3.99)	**Study population**	
										39 per 1000	**31 fewer per 1000** (from 39 fewer to 117 more)
										Moderate	
										39 per 1000	**31 fewer per 1000** (from 39 fewer to 117 more)

[1] High risk of detection bias as unclear if follow-up duration (=<12 weeks) is sufficient to observe potential longer term adverse eventsand reliability/validity of some outcome measures unclear
[2] I-squared value indicates substantial to considerable heterogeneity
[3] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)
[4] Trial funded by pharmaceutical company and/or study drugs were provided by pharmaceutical company and/or authors are consultants to pharmaceutical companies
[5] Events<300
[6] N<400 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)
[7] N<400
[8] I-squared value indicates moderate heterogeneity

Adverse events associated with risperidone versus placebo

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 Adverse events associated with risperidone versus haloperidol**		**Risk with Control**	**Risk difference with Adverse events associated with risperidone versus haloperidol** (95% CI)
Treatment-emergent extrapyramidal symptoms (measured with: Chouinard Extrapyramidal Symptoms Rating Scale (ESRS): Section I; Better indicated by lower values)											
28 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	15	13	**-**		The mean treatment-emergent extrapyramidal symptoms in the intervention groups was **0.83 standard deviations lower** (1.61 to 0.05 lower)
Prolactin concentration (ng/ml) Change Scores (measured with: Laboratory assessment; Better indicated by lower values)											
28 (1 study)	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of	15	13	**-**		The mean prolactin concentration (ng/ml) change scores in the

12 weeks					suspected [3]	bias, imprecision, publication bias				intervention groups was **1.01 standard deviations lower** (1.80 to 0.22 lower)
Liver problems (change in alanine transaminase [ALT]) (measured with: Laboratory assessment; Better indicated by lower values)										
28 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	15	13	-	The mean liver problems (change in alanine transaminase [alt]) in the intervention groups was **0.83 standard deviations lower** (1.60 to 0.05 lower)

[1] High risk of detection bias as unclear if 12 weeks is sufficient follow-up duration to observe potential longer term adverse effects
[2] N<400
[3] Study was partly funded by the pharmaceutical company that manufactured the drug tested

1.33.6 Adverse events associated with antivirals

Adverse events associated with amantadine hydrochloride versus placebo

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 Adverse events associated with antivirals**		**Risk with Control**	**Risk difference with Adverse events associated with antivirals** (95% CI)
Any adverse event (assessed with: Study-specific report of adverse event)											
39	serious[1]	no serious	no serious	very	reporting bias	⊕⊖⊖⊖	14/20	14/19	**RR 1.05**	**Study population**	

(1 study) 5 weeks		inconsistency	indirectness	serious[2]	strongly suspected [3]	**VERY LOW**[1,2,3] due to risk of bias, imprecision, publication bias	(70%)	(73.7%)	(0.71 to 1.56)	**700 per 1000**	**35 more per 1000** (from 203 fewer to 392 more)
										Moderate	
										700 per 1000	**35 more per 1000** (from 203 fewer to 392 more)
Insomnia (assessed with: Study-specific report of adverse event)											
39 (1 study) 5 weeks	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	2/20 (10%)	4/19 (21.1%)	**RR 2.11** (0.43 to 10.19)	**Study population**	
										100 per 1000	**111 more per 1000** (from 57 fewer to 919 more)
										Moderate	
										100 per 1000	**111 more per 1000** (from 57 fewer to 919 more)
Antisocial behaviour (assessed with: Study-specific report of adverse event)											
39 (1 study) 5 weeks	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	4/20 (20%)	2/19 (10.5%)	**RR 0.53** (0.11 to 2.55)	**Study population**	
										200 per 1000	**94 fewer per 1000** (from 178 fewer to 310 more)
										Moderate	
										200 per 1000	**94 fewer per 1000** (from 178 fewer to 310 more)

[1] High risk of detection bias as unclear if 5 weeks is sufficient follow-up duration to observe longer-term adverse events and reliability/validity of measure is unclear

[2] Events<300 and 95% CI crosses both line of no effect and measure of significant benefit or harm (RR 0.75/1.25)
[3] Trial funded by pharmaceutical company

1.33.7 Adverse events associated with cognitive enhancers

Adverse events associated with piracetam and risperidone versus placebo and risperidone

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 Adverse events associated with combined piracetam and risperidone versus combined placebo and risperidone**		**Risk with Control**	**Risk difference with Adverse events associated with combined piracetam and risperidone versus combined placebo and risperidone** (95% CI)
Any treatment-emergent extrapyramidal symptom (assessed with: Extrapyramidal Symptoms Rating Scale (ESRS))											
40 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	8/20 (40%)	6/20 (30%)	**RR 0.75** (0.32 to 1.77)	**Study population**	
										400 per 1000	**100 fewer per 1000** (from 272 fewer to 308 more)
										Moderate	
										400 per 1000	**100 fewer per 1000** (from 272 fewer to 308 more)
Constipation (assessed with: Study-specific side effect checklist)											
40 (1 study)	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of	3/20 (15%)	4/20 (20%)	**RR 1.33** (0.34 to	**Study population**	
										150 per	**50 more per 1000**

10 weeks						bias, imprecision			5.21)	**1000**	(from 99 fewer to 632 more)
										Moderate	
										150 per 1000	**50 more per 1000** (from 99 fewer to 632 more)
Nervousness (assessed with: Study-specific side effect checklist)											
40 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	2/20 (10%)	1/20 (5%)	**RR 0.5** (0.05 to 5.08)	**Study population**	
										100 per 1000	**50 fewer per 1000** (from 95 fewer to 408 more)
										Moderate	
										100 per 1000	**50 fewer per 1000** (from 95 fewer to 408 more)
Day time drowsiness (assessed with: Study-specific side effect checklist)											
40 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	9/20 (45%)	7/20 (35%)	**RR 0.78** (0.36 to 1.68)	**Study population**	
										450 per 1000	**99 fewer per 1000** (from 288 fewer to 306 more)
										Moderate	
										450 per 1000	**99 fewer per 1000** (from 288 fewer to 306 more)
Morning drowsiness (assessed with: Study-specific side effect checklist)											
40 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias,	8/20 (40%)	11/20 (55%)	**RR 1.38** (0.71 to 2.68)	**Study population**	
										400 per 1000	**152 more per 1000** (from 116 fewer to 672 more)

						imprecision				**Moderate**	
										400 per 1000	**152 more per 1000** (from 116 fewer to 672 more)
Increased appetite (assessed with: Study-specific side effect checklist)											
40 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	6/20 (30%)	7/20 (35%)	**RR 1.17** (0.48 to 2.86)	**Study population**	
										300 per 1000	**51 more per 1000** (from 156 fewer to 558 more)
										Moderate	
										300 per 1000	**51 more per 1000** (from 156 fewer to 558 more)
Loss of appetite (assessed with: Study-specific side effect checklist)											
40 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	1/20 (5%)	1/20 (5%)	**RR 1** (0.07 to 14.9)	**Study population**	
										50 per 1000	**0 fewer per 1000** (from 47 fewer to 695 more)
										Moderate	
										50 per 1000	**0 fewer per 1000** (from 47 fewer to 695 more)
Dry mouth (assessed with: Study-specific side effect checklist)											
40 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	3/20 (15%)	4/20 (20%)	**RR 1.33** (0.34 to 5.21)	**Study population**	
										150 per 1000	**50 more per 1000** (from 99 fewer to 632 more)
										Moderate	

									150 per 1000	**50 more per 1000** (from 99 fewer to 632 more)
Fatigue (assessed with: Study-specific side effect checklist)										
40 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	3/20 (15%) 5/20 (25%)	**RR 1.67** (0.46 to 6.06)	**Study population**	
									150 per 1000	**100 more per 1000** (from 81 fewer to 759 more)
									Moderate	
									150 per 1000	**100 more per 1000** (from 81 fewer to 759 more)

[1] High risk of detection bias as not clear if 10 weeks a sufficient follow-up duration to observe potential longer-term adverse events
[2] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)

1.33.8 Adverse events associated with melatonin

Adverse events associated with melatonin versus placebo

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 Adverse events associated with melatonin**		**Risk with Control**	**Risk difference with Adverse events associated with melatonin** (95% CI)
Coughing (assessed with: Study-specific report of adverse event)											
63 (1 study)	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of	13/33 (39.4%)	6/30 (20%)	**RR 0.51** (0.22 to	**Study population**	
										394 per	**193 fewer per 1000**

12 weeks						bias, imprecision			1.17)	**1000**	(from 307 fewer to 67 more)
										Moderate	
										394 per 1000	**193 fewer per 1000** (from 307 fewer to 67 more)
Mood swings (assessed with: Study-specific report of adverse event)											
63 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	6/33 (18.2%)	7/30 (23.3%)	**RR 1.28** (0.49 to 3.39)	**Study population**	
										182 per 1000	**51 more per 1000** (from 93 fewer to 435 more)
										Moderate	
										182 per 1000	**51 more per 1000** (from 93 fewer to 435 more)
Vomiting (assessed with: Study-specific report of adverse event)											
63 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	7/33 (21.2%)	7/30 (23.3%)	**RR 1.1** (0.44 to 2.77)	**Study population**	
										212 per 1000	**21 more per 1000** (from 119 fewer to 375 more)
										Moderate	
										212 per 1000	**21 more per 1000** (from 119 fewer to 375 more)

Increased excitability (assessed with: Study-specific report of adverse event)											
63 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	6/33 (18.2%)	5/30 (16.7%)	**RR 0.92** (0.31 to 2.7)	**Study population**	
										182 per 1000	**15 fewer per 1000** (from 125 fewer to 309 more)
										Moderate	
										182 per 1000	**15 fewer per 1000** (from 126 fewer to 309 more)
Headache (assessed with: Study-specific report of adverse event)											
63 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	2/33 (6.1%)	2/30 (6.7%)	**RR 1.1** (0.17 to 7.33)	**Study population**	
										61 per 1000	**6 more per 1000** (from 50 fewer to 384 more)
										Moderate	
										61 per 1000	**6 more per 1000** (from 51 fewer to 386 more)
Rash (assessed with: Study-specific report of adverse event)											
63 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	3/33 (9.1%)	4/30 (13.3%)	**RR 1.47** (0.36 to 6.03)	**Study population**	
										91 per 1000	**43 more per 1000** (from 58 fewer to 457 more)
										Moderate	

									91 per 1000	**43 more per 1000** (from 58 fewer to 458 more)
Somnolence (assessed with: Study-specific report of adverse event)										
63 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	5/33 (15.2%)	3/30 (10%)	**RR 0.66** (0.17 to 2.53)	**Study population**
									152 per 1000	**52 fewer per 1000** (from 126 fewer to 232 more)
									Moderate	
									152 per 1000	**52 fewer per 1000** (from 126 fewer to 233 more)
Fatigue (assessed with: Study-specific report of adverse event)										
63 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	6/33 (18.2%)	1/30 (3.3%)	**RR 0.18** (0.02 to 1.44)	**Study population**
									182 per 1000	**149 fewer per 1000** (from 178 fewer to 80 more)
									Moderate	
									182 per 1000	**149 fewer per 1000** (from 178 fewer to 80 more)
Hypothermia (assessed with: Study-specific report of adverse event)										
63 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of	2/33 (6.1%)	1/30 (3.3%)	**RR 0.55** (0.05 to 5.76)	**Study population**
									61 per	**27 fewer per 1000** (from 58 fewer to 288

						bias, imprecision				**1000**	more)
										Moderate	
										61 per 1000	**27 fewer per 1000** (from 58 fewer to 290 more)
Increased activity (assessed with: Study-specific report of adverse event)											
63 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	3/33 (9.1%)	3/30 (10%)	**RR 1.1** (0.24 to 5.04)	**Study population**	
										91 per 1000	**9 more per 1000** (from 69 fewer to 367 more)
										Moderate	
										91 per 1000	**9 more per 1000** (from 69 fewer to 368 more)
Nausea (assessed with: Study-specific report of adverse event)											
63 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	2/33 (6.1%)	1/30 (3.3%)	**RR 0.55** (0.05 to 5.76)	**Study population**	
										61 per 1000	**27 fewer per 1000** (from 58 fewer to 288 more)
										Moderate	
										61 per 1000	**27 fewer per 1000** (from 58 fewer to 290 more)
Dizziness (assessed with: Study-specific report of adverse event)											

63 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	2/33 (6.1%)	0/30 (0%)	**RR 0.22** (0.01 to 4.39)	**Study population**	
										61 per 1000	**47 fewer per 1000** (from 60 fewer to 205 more)
										Moderate	
										61 per 1000	**48 fewer per 1000** (from 60 fewer to 207 more)
Breathlessness (assessed with: Study-specific report of adverse event)											
63 (1 study) 12 weeks						See comment	0/33 (0%)	0/30 (0%)	**not pooled**	See comment	See comment
Hung-over feeling (assessed with: Study-specific report of adverse event)											
63 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	0/33 (0%)	1/30 (3.3%)	**RR 3.29** (0.14 to 77.82)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**
Tremor (assessed with: Study-specific report of adverse event)											
63 (1 study) 12 weeks						See comment	0/33 (0%)	0/30 (0%)	**not pooled**	See comment	See comment

Seizures (assessed with: Study-specific report of adverse event)										
63 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	1/33 (3%) 0/30 (0%)	**RR 0.37** (0.02 to 8.65)	**Study population**	
									30 per 1000	**19 fewer per 1000** (from 30 fewer to 232 more)
									Moderate	
									30 per 1000	**19 fewer per 1000** (from 29 fewer to 229 more)
Other (assessed with: Study-specific report of adverse event)										
63 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	20/33 (60.6%) 15/30 (50%)	**RR 0.82** (0.53 to 1.3)	**Study population**	
									606 per 1000	**109 fewer per 1000** (from 285 fewer to 182 more)
									Moderate	
									606 per 1000	**109 fewer per 1000** (from 285 fewer to 182 more)

[1] High risk of detection bias as unclear if 12 weeks is sufficient duration to observe potential longer-term adverse events
[2] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)

1.33.9 Adverse events associated with opioid antagonists

Adverse events associated with naltrexone versus placebo

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 Adverse events associated with opioid antagonists		Risk with Control	Risk difference with Adverse events associated with opioid antagonists (95% CI)
Any side effect (assessed with: Study-specific side effect checklist)											
41 (1 study) 6 weeks	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	7/18 (38.9%)	13/23 (56.5%)	**RR 1.45** (0.74 to 2.87)	**Study population**	
										389 per 1000	**175 more per 1000** (from 101 fewer to 727 more)
										Moderate	
										389 per 1000	**175 more per 1000** (from 101 fewer to 727 more)
Aggressiveness (assessed with: Study-specific side effect checklist)											
41 (1 study) 6 weeks	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	5/18 (27.8%)	4/23 (17.4%)	**RR 0.63** (0.2 to 2)	**Study population**	
										278 per 1000	**103 fewer per 1000** (from 222 fewer to 278 more)
										Moderate	
										278 per 1000	**103 fewer per 1000** (from 222 fewer to 278 more)
Self-injurious behaviour (assessed with: Study-specific side effect checklist)											
41 (1 study) 6 weeks	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	2/18 (11.1%)	1/23 (4.3%)	**RR 0.39** (0.04 to 3.98)	**Study population**	
										111 per 1000	**68 fewer per 1000** (from 107 fewer to 331 more)

									Moderate	
									111 per 1000	**68 fewer per 1000** (from 107 fewer to 331 more)
Hyperactivity (assessed with: Study-specific side effect checklist)										
41 (1 study) 6 weeks	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	3/18 (16.7%)	2/23 (8.7%)	**RR 0.52** (0.1 to 2.8)	**Study population**
									167 per 1000	**80 fewer per 1000** (from 150 fewer to 300 more)
									Moderate	
									167 per 1000	**80 fewer per 1000** (from 150 fewer to 301 more)
Temper tantrums (assessed with: Study-specific side effect checklist)										
41 (1 study) 6 weeks	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	1/18 (5.6%)	2/23 (8.7%)	**RR 1.57** (0.15 to 15.92)	**Study population**
									56 per 1000	**32 more per 1000** (from 47 fewer to 829 more)
									Moderate	
									56 per 1000	**32 more per 1000** (from 48 fewer to 836 more)
Stereotypies (assessed with: Study-specific side effect checklist)										
41 (1 study) 6 weeks	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	3/18 (16.7%)	2/23 (8.7%)	**RR 0.52** (0.1 to 2.8)	**Study population**
									167 per 1000	**80 fewer per 1000** (from 150 fewer to 300 more)

									Moderate	
									167 per 1000	**80 fewer per 1000** (from 150 fewer to 301 more)
Irritability (assessed with: Study-specific side effect checklist)										
41 (1 study) 6 weeks	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	2/18 (11.1%)	3/23 (13%)	**RR 1.17** (0.22 to 6.3)	**Study population**
									111 per 1000	**19 more per 1000** (from 87 fewer to 589 more)
									Moderate	
									111 per 1000	**19 more per 1000** (from 87 fewer to 588 more)
Decreased transient verbal production (assessed with: Study-specific side effect checklist)										
41 (1 study) 6 weeks	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	0/18 (0%)	1/23 (4.3%)	**RR 2.38** (0.1 to 55.06)	**Study population**
									0 per 1000	-
									Moderate	
									0 per 1000	-
Slight sleepiness (assessed with: Study-specific side effect checklist)										
41 (1 study) 6 weeks	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	0/18 (0%)	1/23 (4.3%)	**RR 2.38** (0.1 to 55.06)	**Study population**
									0 per 1000	-
									Moderate	

									0 per 1000	-
Falling asleep (assessed with: Study-specific side effect checklist)										
41 (1 study) 6 weeks	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	0/18 (0%)	2/23 (8.7%)	**RR 3.96** (0.2 to 77.63)	**Study population**
									0 per 1000	-
									Moderate	
									0 per 1000	-
Decreased appetite (assessed with: Study-specific side effect checklist)										
41 (1 study) 6 weeks	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	0/18 (0%)	2/23 (8.7%)	**RR 3.96** (0.2 to 77.63)	**Study population**
									0 per 1000	-
									Moderate	
									0 per 1000	-
Vomiting (assessed with: Study-specific side effect checklist)										
41 (1 study) 6 weeks	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	0/18 (0%)	3/23 (13%)	**RR 5.54** (0.3 to 100.86)	**Study population**
									0 per 1000	-
									Moderate	
									0 per 1000	-

[1] High risk of detection bias as outcome measure designed specifically for the study with no independent reliability or validity ratings, and it is unclear if 6 weeks is a sufficient follow-up duration to observe potential longer-term side effects

[2] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)
[3] Potential conflict of interest as drug and placebo were supplied by the manufacturer

1.33.10 Adverse events associated with SNRIs

Adverse events associated with atomoxetine versus placebo

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 Adverse events associated with selective noradrenaline reuptake inhibitors		Risk with Control	Risk difference with Adverse events associated with selective noradrenaline reuptake inhibitors (95% CI)
Any adverse event (assessed with: Study-specific open-ended questionning for adverse events)											
97 (1 study) 8 weeks	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	32/49 (65.3%)	39/48 (81.3%)	**RR 1.24** (0.97 to 1.59)	**Study population**	
										653 per 1000	**157 more per 1000** (from 20 fewer to 385 more)
										Moderate	
										653 per 1000	**157 more per 1000** (from 20 fewer to 385 more)
Discontinuation due to adverse events (assessed with: Study-specific open-ended questionning for adverse events)											
97 (1 study) 8 weeks	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	0/49 (0%)	1/48 (2.1%)	**OR 3.13** (0.12 to 78.66)	**Study population**	
										0 per 1000	-
										Moderate	
										0 per	-

									1000	
Abdominal pain (assessed with: Study-specific open-ended questionning for adverse events)										
97 (1 study) 8 weeks	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	3/49 (6.1%)	4/48 (8.3%)	**RR 1.36** (0.32 to 5.76)	**Study population** **61 per 1000** | **22 more per 1000** (from 42 fewer to 291 more) **Moderate** **61 per 1000** | **22 more per 1000** (from 41 fewer to 290 more)
Upper abdominal pain (assessed with: Study-specific open-ended questionning for adverse events)										
97 (1 study) 8 weeks	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	3/49 (6.1%)	9/48 (18.8%)	**RR 3.06** (0.88 to 10.63)	**Study population** **61 per 1000** | **126 more per 1000** (from 7 fewer to 590 more) **Moderate** **61 per 1000** | **126 more per 1000** (from 7 fewer to 587 more)
Diarrhoea (assessed with: Study-specific open-ended questionning for adverse events)										
97 (1 study) 8 weeks	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	3/49 (6.1%)	1/48 (2.1%)	**RR 0.34** (0.04 to 3.16)	**Study population** **61 per 1000** | **40 fewer per 1000** (from 59 fewer to 132 more) **Moderate** **61 per** | **40 fewer per 1000**

										1000	(from 59 fewer to 132 more)
Nausea (assessed with: Study-specific open-ended questionning for adverse events)											
97 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	4/49 (8.2%)	14/48 (29.2%)	**RR 3.57** (1.27 to 10.08)	**Study population**	
										82 per 1000	**210 more per 1000** (from 22 more to 741 more)
										Moderate	
										82 per 1000	**211 more per 1000** (from 22 more to 745 more)
Vomiting (assessed with: Study-specific open-ended questionning for adverse events)											
97 (1 study) 8 weeks	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	5/49 (10.2%)	7/48 (14.6%)	**RR 1.43** (0.49 to 4.19)	**Study population**	
										102 per 1000	**44 more per 1000** (from 52 fewer to 326 more)
										Moderate	
										102 per 1000	**44 more per 1000** (from 52 fewer to 325 more)
Fatigue (assessed with: Study-specific open-ended questionning for adverse events)											
97 (1 study) 8 weeks	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	4/49 (8.2%)	11/48 (22.9%)	**RR 2.81** (0.96 to 8.21)	**Study population**	
										82 per 1000	**148 more per 1000** (from 3 fewer to 589 more)
										Moderate	

										82 per 1000	**148 more per 1000** (from 3 fewer to 591 more)
Pyrexia (assessed with: Study-specific open-ended questionning for adverse events)											
97 (1 study) 8 weeks	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	3/49 (6.1%)	0/48 (0%)	**RR 0.15** (0.01 to 2.75)	**Study population**	
										61 per 1000	**52 fewer per 1000** (from 61 fewer to 107 more)
										Moderate	
										61 per 1000	**52 fewer per 1000** (from 60 fewer to 107 more)
Influenza (assessed with: Study-specific open-ended questionning for adverse events)											
97 (1 study) 8 weeks	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	0/49 (0%)	3/48 (6.3%)	**RR 7.14** (0.38 to 134.69)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**
Deceased appetite (assessed with: Study-specific open-ended questionning for adverse events)											
97 (1 study) 8 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	reporting bias strongly suspected [3]	⊕⊖⊖⊖ **VERY LOW**[1,3,4] due to risk of bias, imprecision, publication bias	3/49 (6.1%)	13/48 (27.1%)	**RR 4.42** (1.34 to 14.55)	**Study population**	
										61 per 1000	**209 more per 1000** (from 21 more to 830 more)
										Moderate	
										61 per	**209 more per 1000**

										1000	(from 21 more to 827 more)
Myalgia (assessed with: Study-specific open-ended questionning for adverse events)											
97 (1 study) 8 weeks	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	0/49 (0%)	3/48 (6.3%)	**RR 7.14** (0.38 to 134.69)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**
Dizziness (assessed with: Study-specific open-ended questionning for adverse events)											
97 (1 study) 8 weeks	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	1/49 (2%)	3/48 (6.3%)	**RR 3.06** (0.33 to 28.42)	**Study population**	
										20 per 1000	**42 more per 1000** (from 14 fewer to 560 more)
										Moderate	
										20 per 1000	**41 more per 1000** (from 13 fewer to 548 more)
Headache (assessed with: Study-specific open-ended questionning for adverse events)											
97 (1 study) 8 weeks	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	9/49 (18.4%)	12/48 (25%)	**RR 1.36** (0.63 to 2.93)	**Study population**	
										184 per 1000	**66 more per 1000** (from 68 fewer to 354 more)
										Moderate	
										184 per	**66 more per 1000** (from 68 fewer to 355

										1000	more)
Psychomotor hyperactivity (assessed with: Study-specific open-ended questionning for adverse events)											
97 (1 study) 8 weeks	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	4/49 (8.2%)	1/48 (2.1%)	**RR 0.26** (0.03 to 2.2)	**Study population**	
										82 per 1000	**60 fewer per 1000** (from 79 fewer to 98 more)
										Moderate	
										82 per 1000	**61 fewer per 1000** (from 80 fewer to 98 more)
Aggression (assessed with: Study-specific open-ended questionning for adverse events)											
97 (1 study) 8 weeks	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	3/49 (6.1%)	2/48 (4.2%)	**RR 0.68** (0.12 to 3.89)	**Study population**	
										61 per 1000	**20 fewer per 1000** (from 54 fewer to 177 more)
										Moderate	
										61 per 1000	**20 fewer per 1000** (from 54 fewer to 176 more)
Early morning awakening (assessed with: Study-specific open-ended questionning for adverse events)											
97 (1 study) 8 weeks	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	0/49 (0%)	5/48 (10.4%)	**RR 11.22** (0.64 to 197.6)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per	**-**

									1000	
Initial insomnia (assessed with: Study-specific open-ended questionning for adverse events)										
97 (1 study) 8 weeks	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	5/49 (10.2%) 3/48 (6.3%)	**RR 0.61** (0.15 to 2.42)	**Study population**	
									102 per 1000	**40 fewer per 1000** (from 87 fewer to 145 more)
									Moderate	
									102 per 1000	**40 fewer per 1000** (from 87 fewer to 145 more)

[1] High risk of detection bias as unclear if 8 weeks is sufficient follow-up duration to observe potential longer-term adverse events
[2] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)
[3] Trial run and reported by pharmaceutical company
[4] Events<300

1.34 ADVERSE EVENTS ASSOCIATED WITH BIOMEDICAL INTERVENTIONS

1.34.1 Adverse events associated with medical procedures

Adverse events associated with HBOT versus attention-placebo

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 Adverse events associated with hyperbaric oxygen treatment (HBOT) versus attention-placebo control**		**Risk with Control**	**Risk difference with Adverse events associated with hyperbaric oxygen treatment (HBOT) versus attention-placebo control** (95% CI)
Any adverse event (assessed with: Study-specific daily treatment logbooks)											

62 (1 study) 4 weeks	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	2/29 (6.9%)	3/33 (9.1%)	**RR 1.32** (0.24 to 7.35)	**Study population**	
										69 per 1000	**22 more per 1000** (from 52 fewer to 438 more)
										Moderate	
										69 per 1000	**22 more per 1000** (from 52 fewer to 438 more)
Minor-grade ear barotrauma (assessed with: Not reported)											
58 (1 study) 4 weeks	serious[4]	no serious inconsistency	no serious indirectness	serious[5]	undetected	⊕⊕⊖⊖ **LOW**[4,5] due to risk of bias, imprecision	3/29 (10.3%)	11/29 (37.9%)	**RR 3.67** (1.14 to 11.79)	**Study population**	
										103 per 1000	**276 more per 1000** (from 14 more to 1000 more)
										Moderate	
										103 per 1000	**275 more per 1000** (from 14 more to 1000 more)

[1] High risk of detection bias as unclear if 4 weeks sufficient follow-up duration to detect potential longer-term adverse events and adverse events were recorded by the intervention administrator who was non-blind to treatment assignment and to other potentially confounding factors
[2] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)
[3] Potential conflict of interest as study funded by the International Hyperbarics Association and authors profit from the use of hyperbaric treatment in their clinical practices
[4] High risk of detection bias as unclear if 4 weeks was a sufficient follow-up duration to observe potential longer-term adverse events and outcome measure and outcome assessor/s not reported so blinding, and reliability and validity unclear
[5] Events<300

1.34.2 Adverse events associated with nutritional interventions

Adverse events associated with multivitamin/mineral supplement versus placebo

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 Adverse events associated with multivitamin and mineral supplement		Risk with Control	Risk difference with Adverse events associated with multivitamin and mineral supplement (95% CI)
Discontinuation due to adverse events (assessed with: Number of participants who discontinued due to adverse events)											
141 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	5/69 (7.2%)	3/72 (4.2%)	**RR 0.57** (0.14 to 2.31)	**Study population**	
										72 per 1000	**31 fewer per 1000** (from 62 fewer to 95 more)
										Moderate	
										73 per 1000	**31 fewer per 1000** (from 63 fewer to 96 more)
Discontinuation due to diarrhoea (assessed with: Number of participants who discontinued due to diarrhoea)											
141 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	3/69 (4.3%)	1/72 (1.4%)	**RR 0.32** (0.03 to 3)	**Study population**	
										43 per 1000	**30 fewer per 1000** (from 42 fewer to 87 more)
										Moderate	
										44 per 1000	**30 fewer per 1000** (from 43 fewer to 88 more)
Discontinuation due to increased stimming (assessed with: Number of participants who discontinued due to increased stimming)											
141 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	1/69 (1.4%)	0/72 (0%)	**RR 0.32** (0.01 to 7.72)	**Study population**	
										14 per 1000	**10 fewer per 1000** (from 14 fewer to 97 more)
										Moderate	

									15 per 1000	**10 fewer per 1000** (from 15 fewer to 101 more)
Discontinuation due to behaviour problems (assessed with: Number of participants who discontinued due to behaviour problems)										
141 (1 study) 13 weeks	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	undetected	⊕⊕⊖⊖ **LOW**[1] due to imprecision	1/69 (1.4%) 2/72 (2.8%)	**RR 1.92** (0.18 to 20.66)	**Study population**	
									14 per 1000	**13 more per 1000** (from 12 fewer to 285 more)
									Moderate	
									15 per 1000	**14 more per 1000** (from 12 fewer to 295 more)

[1] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)

Adverse events associated with omega-3 fatty acids versus placebo

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 Adverse events associated with omega-3 fatty acids**		**Risk with Control**	**Risk difference with Adverse events associated with omega-3 fatty acids** (95% CI)
Any adverse event (assessed with: Study-specific report of adverse event)											
27 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	4/13 (30.8%)	5/14 (35.7%)	**RR 1.16** (0.4 to 3.41)	**Study population**	
										308 per 1000	**49 more per 1000** (from 185 fewer to 742 more)
										Moderate	
										308 per	**49 more per 1000**

									1000	(from 185 fewer to 742 more)
Rashes (assessed with: Study-specific report of adverse event)										
27 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	0/13 (0%)	2/14 (14.3%)	**RR 4.67** (0.24 to 88.96)	**Study population**
									0 per 1000	-
									Moderate	
									0 per 1000	-
Upper respiratory infection (assessed with: Study-specific report of adverse event)										
27 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	0/13 (0%)	1/14 (7.1%)	**RR 2.8** (0.12 to 63.2)	**Study population**
									0 per 1000	-
									Moderate	
									0 per 1000	-
Nose bleeds (assessed with: Study-specific report of adverse event)										
27 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	0/13 (0%)	1/14 (7.1%)	**RR 2.8** (0.12 to 63.2)	**Study population**
									0 per 1000	-
									Moderate	
									0 per 1000	-
GI symptoms (assessed with: Study-specific report of adverse event)										

27 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	0/13 (0%)	1/14 (7.1%)	**RR 2.8** (0.12 to 63.2)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**
Hyperactivity (assessed with: Study-specific report of adverse event)											
27 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	3/13 (23.1%)	0/14 (0%)	**RR 0.13** (0.01 to 2.36)	**Study population**	
										231 per 1000	**201 fewer per 1000** (from 228 fewer to 314 more)
										Moderate	
										231 per 1000	**201 fewer per 1000** (from 229 fewer to 314 more)
Self-stimulatory behaviour (assessed with: Study-specific report of adverse event)											
27 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	1/13 (7.7%)	0/14 (0%)	**RR 0.31** (0.01 to 7.02)	**Study population**	
										77 per 1000	**53 fewer per 1000** (from 76 fewer to 463 more)
										Moderate	
										77 per 1000	**53 fewer per 1000** (from 76 fewer to 464 more)

[1] High risk of detection bias as unclear if 12 weeks is sufficient follow-up duration to observe potential longer-term adverse effects and reliability/validity of outcome measure is unclear
[2] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)

Adverse events associated with immunoglobulin (dosages combined) versus placebo

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 Adverse events associated with immunoglobulin (dosages combined)		Risk with Control	Risk difference with Adverse events associated with immunoglobulin (dosages combined) (95% CI)
Any adverse event (assessed with: Study-specific report of adverse event)											
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	undetected	⊕⊕⊖⊖ **LOW**[1,2] due to risk of bias, imprecision	25/31 (80.6%)	71/94 (75.5%)	**RR 0.94** (0.76 to 1.15)	**Study population**	
										806 per 1000	**48 fewer per 1000** (from 194 fewer to 121 more)
										Moderate	
										807 per 1000	**48 fewer per 1000** (from 194 fewer to 121 more)
Discontinuation due to adverse events (assessed with: Study-specific report of adverse event)											
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	1/31 (3.2%)	7/94 (7.4%)	**RR 2.31** (0.3 to 18.03)	**Study population**	
										32 per 1000	**42 more per 1000** (from 23 fewer to 549 more)
										Moderate	
										32 per 1000	**42 more per 1000** (from 22 fewer to 545 more)
Infections/Infestations (assessed with: Study-specific report of adverse event)											
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias,	16/31 (51.6%)	46/94 (48.9%)	**RR 0.95** (0.64 to 1.41)	**Study population**	
										516 per	**26 fewer per 1000** (from 186 fewer to 212

						imprecision			**1000**	more)
									Moderate	
									516 per 1000	**26 fewer per 1000** (from 186 fewer to 212 more)
Gastrointestinal disorders (assessed with: Study-specific report of adverse event)										
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	9/31 (29%)	36/94 (38.3%)	**RR 1.32** (0.72 to 2.42)	**Study population**
									290 per 1000	**93 more per 1000** (from 81 fewer to 412 more)
									Moderate	
									290 per 1000	**93 more per 1000** (from 81 fewer to 412 more)
Psychiatric disorders (assessed with: Study-specific report of adverse event)										
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	6/31 (19.4%)	17/94 (18.1%)	**RR 0.93** (0.4 to 2.16)	**Study population**
									194 per 1000	**14 fewer per 1000** (from 116 fewer to 225 more)
									Moderate	
									194 per 1000	**14 fewer per 1000** (from 116 fewer to 225 more)
Respiratory/Thoracic/Mediastinal disorders (assessed with: Study-specific report of adverse event)										
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias,	4/31 (12.9%)	15/94 (16%)	**RR 1.24** (0.44 to 3.45)	**Study population**
									129 per 1000	**31 more per 1000** (from 72 fewer to 316 more)

						imprecision				**Moderate**	
										129 per 1000	**31 more per 1000** (from 72 fewer to 316 more)
Skin/Subcutaneous tissue disorders (assessed with: Study-specific report of adverse event)											
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	3/31 (9.7%)	12/94 (12.8%)	**RR 1.32** (0.4 to 4.37)	**Study population**	
										97 per 1000	**31 more per 1000** (from 58 fewer to 326 more)
										Moderate	
										97 per 1000	**31 more per 1000** (from 58 fewer to 327 more)
General disorders/Administration site conditions (assessed with: Study-specific report of adverse event)											
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	2/31 (6.5%)	9/94 (9.6%)	**RR 1.48** (0.34 to 6.5)	**Study population**	
										65 per 1000	**31 more per 1000** (from 43 fewer to 355 more)
										Moderate	
										65 per 1000	**31 more per 1000** (from 43 fewer to 357 more)
Nervous system disorders (assessed with: Study-specific report of adverse event)											
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	0/31 (0%)	7/94 (7.4%)	**RR 5.05** (0.3 to 86.01)	**Study population**	
										0 per 1000	-
										Moderate	
										0 per 1000	-

Injury/Poisoning/Procedural complications (assessed with: Study-specific report of adverse event)											
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	1/31 (3.2%)	5/94 (5.3%)	**RR 1.65** (0.2 to 13.58)	**Study population**	
										32 per 1000	**21 more per 1000** (from 26 fewer to 406 more)
										Moderate	
										32 per 1000	**21 more per 1000** (from 26 fewer to 403 more)
Investigations (assessed with: Study-specific report of adverse event)											
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	1/31 (3.2%)	3/94 (3.2%)	**RR 0.99** (0.11 to 9.17)	**Study population**	
										32 per 1000	**0 fewer per 1000** (from 29 fewer to 264 more)
										Moderate	
										32 per 1000	**0 fewer per 1000** (from 28 fewer to 261 more)
Metabolism/Nutrition disorders (assessed with: Study-specific report of adverse event)											
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	1/31 (3.2%)	3/94 (3.2%)	**RR 0.99** (0.11 to 9.17)	**Study population**	
										32 per 1000	**0 fewer per 1000** (from 29 fewer to 264 more)
										Moderate	
										32 per 1000	**0 fewer per 1000** (from 28 fewer to 261 more)
Eye disorders (assessed with: Study-specific report of adverse event)											
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of	0/31 (0%)	3/94 (3.2%)	**RR 2.36** (0.13 to 44.42)	**Study population**	
										0 per	**-**

						bias, imprecision				**1000**	
										Moderate	
										0 per 1000	**-**
Blood/Lymphatic system disorders (assessed with: Study-specific report of adverse event)											
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	1/31 (3.2%)	1/94 (1.1%)	**RR 0.33** (0.02 to 5.12)	**Study population**	
										32 per 1000	**22 fewer per 1000** (from 32 fewer to 133 more)
										Moderate	
										32 per 1000	**21 fewer per 1000** (from 31 fewer to 132 more)
Renal/Urinary disorders (assessed with: Study-specific report of adverse event)											
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	2/31 (6.5%)	0/94 (0%)	**RR 0.07** (0 to 1.37)	**Study population**	
										65 per 1000	**60 fewer per 1000** (from 65 fewer to 24 more)
										Moderate	
										65 per 1000	**60 fewer per 1000** (from 65 fewer to 24 more)
Ear/Labyrinth disorders (assessed with: Study-specific report of adverse event)											
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	0/31 (0%)	1/94 (1.1%)	**RR 1.01** (0.04 to 24.19)	**Study population**	
										0 per 1000	**-**
										Moderate	

									0 per 1000	**-**
Immune system disorders (assessed with: Study-specific report of adverse event)										
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	0/31 (0%) 1/94 (1.1%)	**RR 1.01** (0.04 to 24.19)	**Study population**	
									0 per 1000	**-**
									Moderate	
									0 per 1000	**-**
Vascular disorders (assessed with: Study-specific report of adverse event)										
125 (1 study) 12 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,3] due to risk of bias, imprecision	0/31 (0%) 1/94 (1.1%)	**RR 1.01** (0.04 to 24.19)	**Study population**	
									0 per 1000	**-**
									Moderate	
									0 per 1000	**-**

[1] High risk of detection bias as unclear if 12 weeks is sufficient follow-up duration to observe potential longer-term adverse events
[2] Events<300
[3] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (RR 0.75/1.25)

Adverse events associated with ginkgo biloba and risperidone versus placebo and risperidone

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 Adverse events associated with combined ginkgo biloba and risperidone versus combined placebo and**		**Risk with Control**	**Risk difference with Adverse events associated with combined ginkgo biloba and risperidone versus combined placebo and risperidone**

							risperidone		(95% CI)	
Daytime drowsiness (assessed with: Study-specific side effect checklist)										
47 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	7/24 (29.2%)	6/23 (26.1%)	**RR 0.89** (0.35 to 2.26)	**Study population** **292 per 1000** **32 fewer per 1000** (from 190 fewer to 368 more) **Moderate** **292 per 1000** **32 fewer per 1000** (from 190 fewer to 368 more)
Morning drowsiness (assessed with: Study-specific side effect checklist)										
47 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	0/24 (0%)	2/23 (8.7%)	**RR 5.21** (0.26 to 102.98)	**Study population** **0 per 1000** **-** **Moderate** **0 per 1000** **-**
Constipation (assessed with: Study-specific side effect checklist)										
47 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⨁⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	3/24 (12.5%)	3/23 (13%)	**RR 1.04** (0.23 to 4.65)	**Study population** **125 per 1000** **5 more per 1000** (from 96 fewer to 456 more) **Moderate** **125 per 1000** **5 more per 1000** (from 96 fewer to 456 more)
Dizziness (assessed with: Study-specific side effect checklist)										

47 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	3/24 (12.5%)	1/23 (4.3%)	**RR 0.35** (0.04 to 3.11)	**Study population**	
										125 per 1000	**81 fewer per 1000** (from 120 fewer to 264 more)
										Moderate	
										125 per 1000	**81 fewer per 1000** (from 120 fewer to 264 more)
Slow movement (assessed with: Study-specific side effect checklist)											
47 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	1/24 (4.2%)	2/23 (8.7%)	**RR 2.09** (0.2 to 21.48)	**Study population**	
										42 per 1000	**45 more per 1000** (from 33 fewer to 853 more)
										Moderate	
										42 per 1000	**46 more per 1000** (from 34 fewer to 860 more)
Nervousness (assessed with: Study-specific side effect checklist)											
47 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	1/24 (4.2%)	5/23 (21.7%)	**RR 5.22** (0.66 to 41.32)	**Study population**	
										42 per 1000	**176 more per 1000** (from 14 fewer to 1000 more)
										Moderate	
										42 per 1000	**177 more per 1000** (from 14 fewer to 1000 more)
Restlessness (assessed with: Study-specific side effect checklist)											

47 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	5/24 (20.8%)	3/23 (13%)	**RR 0.63** (0.17 to 2.33)	**Study population**	
										208 per 1000	**77 fewer per 1000** (from 173 fewer to 277 more)
										Moderate	
										208 per 1000	**77 fewer per 1000** (from 173 fewer to 277 more)
Increased appetite (assessed with: Study-specific side effect checklist)											
47 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	10/24 (41.7%)	6/23 (26.1%)	**RR 0.63** (0.27 to 1.44)	**Study population**	
										417 per 1000	**154 fewer per 1000** (from 304 fewer to 183 more)
										Moderate	
										417 per 1000	**154 fewer per 1000** (from 304 fewer to 183 more)
Loss of appetite (assessed with: Study-specific side effect checklist)											
47 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	4/24 (16.7%)	3/23 (13%)	**RR 0.78** (0.2 to 3.12)	**Study population**	
										167 per 1000	**37 fewer per 1000** (from 133 fewer to 353 more)
										Moderate	
										167 per 1000	**37 fewer per 1000** (from 134 fewer to 354 more)
Fatigue (assessed with: Study-specific side effect checklist)											

47 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	2/24 (8.3%)	5/23 (21.7%)	**RR 2.61** (0.56 to 12.13)	**Study population**	
										83 per 1000	**134 more per 1000** (from 37 fewer to 927 more)
										Moderate	
										83 per 1000	**134 more per 1000** (from 37 fewer to 924 more)
Diarrhoea (assessed with: Study-specific side effect checklist)											
47 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	3/24 (12.5%)	3/23 (13%)	**RR 1.04** (0.23 to 4.65)	**Study population**	
										125 per 1000	**5 more per 1000** (from 96 fewer to 456 more)
										Moderate	
										125 per 1000	**5 more per 1000** (from 96 fewer to 456 more)
Twitches (assessed with: Study-specific side effect checklist)											
47 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	0/24 (0%)	3/23 (13%)	**RR 7.29** (0.4 to 133.82)	**Study population**	
										0 per 1000	**-**
										Moderate	
										0 per 1000	**-**
Dry mouth (assessed with: Study-specific side effect checklist)											
47 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias,	1/24 (4.2%)	1/23 (4.3%)	**RR 1.04** (0.07 to 15.72)	**Study population**	
										42 per 1000	**2 more per 1000** (from 39 fewer to 613 more)

						imprecision			**Moderate**	
									42 per 1000	**2 more per 1000** (from 39 fewer to 618 more)
Trouble swallowing (assessed with: Study-specific side effect checklist)										
47 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	3/24 (12.5%) 1/23 (4.3%)	**RR 0.35** (0.04 to 3.11)	**Study population**	
									125 per 1000	**81 fewer per 1000** (from 120 fewer to 264 more)
									Moderate	
									125 per 1000	**81 fewer per 1000** (from 120 fewer to 264 more)
Sore throat/tongue (assessed with: Study-specific side effect checklist)										
47 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	5/24 (20.8%) 1/23 (4.3%)	**RR 0.21** (0.03 to 1.65)	**Study population**	
									208 per 1000	**165 fewer per 1000** (from 202 fewer to 135 more)
									Moderate	
									208 per 1000	**164 fewer per 1000** (from 202 fewer to 135 more)
Abdominal pain (assessed with: Study-specific side effect checklist)										
47 (1 study) 10 weeks	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	undetected	⊕⊖⊖⊖ **VERY LOW**[1,2] due to risk of bias, imprecision	3/24 (12.5%) 2/23 (8.7%)	**RR 0.7** (0.13 to 3.79)	**Study population**	
									125 per 1000	**38 fewer per 1000** (from 109 fewer to 349 more)

									Moderate	
									125 per 1000	**38 fewer per 1000** (from 109 fewer to 349 more)

[1] Risk of detection bias is unclear/unknown for adverse event outcomes as it is unclear if 10 weeks is a sufficient follow-up duration to observe potential longer-term adverse events, the reliability and validity of the checklist used to record adverse events is unclear, and the checklist is based on parental report and parents will be non-blind to other potentially confounding factors
[2] Events<300 and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)

Adverse events associated with gluten-free and casein-free diet versus treatment as usual

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 Adverse events associated with gluten-free and casein-free diet**		**Risk with Control**	**Risk difference with Adverse events associated with gluten-free and casein-free diet** (95% CI)
Any adverse event (assessed with: Outcome measure not reported)											
72 (1 study) 35 weeks						See comment	0/34 (0%)	0/38 (0%)	**not pooled**	Effect size not estimable	Effect size not estimable

2 ECONOMIC EVIDENCE PROFILES

2.1 CLINICAL / ECONOMIC QUESTION: RECIPROCAL-SOCIAL COMMUNICATION ADDED TO STANDARD CARE VERSUS STANDARD CARE ALONE FOR PRESCHOOL CHILDREN WITH AUTISM

Evidence profile - economic evidence

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1]	Incremental effect	ICER (£/effect)	Uncertainty[1]
Byford et al., Unpublished UK	Minor limitations[2]	Partially applicable[3]	• Measure of outcome: proportion of children with clinically meaningful improvement expressed by an ADOS-G score improvement ≥ 4 points • Time horizon: 13 months • ICER and probabilistic analysis based on bootstrapping techniques	£5,121 (£4,109 to £6,135)	12%	£297	Compared with standard care alone, intervention plus standard care has greater probability of being cost-effective above willingness to pay of £293 per 1% increase in % of children with clinically meaningful improvement

1. Costs uplifted to 2012 UK pounds using the hospital and community health services pay and prices inflation index (Curtis, 2012).
2. Economic analysis conducted alongside an RCT, all relevant costs included, unit costs based mostly on national sources, HRQoL not considered, sensitivity analysis undertaken including probabilistic sensitivity analysis, time horizon of 13 months
3. Conducted in the UK, perspective including statutory and non-statutory health and social services, no QALYs estimated

2.2 Clinical / economic question: antipsychotics versus placebo for the management of behaviour that challenges in children and young people with autism

Evidence profile - economic evidence

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1]	Incremental effect (QALY)	ICER (£/QALY)	Uncertainty[1]
Guideline model	Potentially serious limitations [2]	Partially applicable[3]	• Time horizon: 32 weeks	Risperidone: • tablets: £8.47 • oral solution: £144 • orodispersible tablets: £205 Aripiprazole tablets: £510	All antipsychotics: 0.008	Risperidone: • tablets: £1,004 • oral solution: £17,083 • orodispersible tablets: £24,267 Aripiprazole tablets: £60,527	PSA: probability of antipsychotics being cost-effective at £20,000/QALY: Risperidone tablets: 0.63 Risperidone oral solution: 0.47 Risperidone orodispersible tablets: 0.40 Aripiprazole tablets: 0.10

1. Costs expressed in 2012 UK pounds
2. Only intervention costs considered consisting of drug acquisition costs, efficacy data from 4 trials, PSA performed
3. NHS & PSS perspective, QALYs based on HUI3 (valuations elicited from Canadian population)

2.3 CLINICAL / ECONOMIC QUESTION: EARLY INTENSIVE BEHAVIOURAL INTERVENTION VERSUS STANDARD EDUCATIONAL SERVICE (SPECIAL EDUCATION) FOR CHILDREN WITH AUTISM

Evidence profile - economic evidence

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1]	Incremental effect	ICER (£/effect)[1]	Uncertainty[1]
Chasson *et al.*, 2007 US	Potentially serious limitations [2]	Partially applicable[3]	• Cost analysis • Time horizon: 18 years	-£99,039	NA	NA	Not estimated

1. Costs converted and uplifted to 2012 UK pounds – converted using PPP exchange rates (http://www.oecd.org/std/ppp) and UK PPS local authorities adults and children's services pay and prices inflation index (Curtis, 2012).
2. Simple economic model including education costs only, cost estimates based on personal communication and further assumptions, clinical model parameters based on published literature and further assumptions; local state costs, no sensitivity analysis
3. Conducted in the US, public perspective including state, local, federal and private costs, no discounting although time horizon was 18 years

2.4 CLINICAL / ECONOMIC QUESTION: EARLY INTENSIVE BEHAVIOURAL INTERVENTION VERSUS NO INTERVENTION FOR PRESCHOOL CHILDREN WITH AUTISM

Evidence profile - economic evidence

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1]	Incremental effect	ICER (£/effect)	Uncertainty[1]
Motiwala et al., 2006 US	Potentially serious limitations[2]	Partially applicable[3]	• Measure of outcome: number of dependency-free years • Time horizon: up to 65 years of age	-£37,450	4.4	Intervention dominant	Findings sensitive to discount rate and EIBI efficacy (net costs and not savings, with discount rate of 5%)

1. Costs converted and uplifted to 2012 UK pounds – converted using PPP exchange rates (http://www.oecd.org/std/ppp) and UK PPS local authorities adults and children's services pay and prices inflation index (Curtis, 2012).
2. Economic model over lifetime, provincial government resource use estimates and prices, all relevant costs included, but efficacy estimates were judgements based on literature review
3. Conducted in Canada, public perspective, discounting 3%, no QALYs but intervention dominant

2.5 CLINICAL / ECONOMIC QUESTION: EARLY INTENSIVE BEHAVIOURAL INTERVENTION VERSUS TREATMENT AS USUAL FOR PRESCHOOL CHILDREN WITH AUTISM

Evidence profile - economic evidence

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1]	Incremental effect	ICER (£/effect)	Uncertainty[1]
Peters-Scheffer et al., 2012 Netherlands	Potentially serious limitations[2]	Partially applicable[3]	• Time horizon: up to 65 years of age	-£925,338	NA	NA	Using more optimistic TAU efficacy data: -£210,358

1. Costs converted and uplifted to 2012 UK pounds – converted using PPP exchange rates (http://www.oecd.org/std/ppp) and UK PPS local authorities adults and children's services pay and prices inflation index (Curtis, 2012).
2. Economic model over lifetime, resource use and unit cost data based on national sources and assumptions, all relevant costs included, efficacy estimates based on review of meta-analyses, selection of studies based on their applicability to the Dutch context, and naïve addition of meta-analytic data across same treatment arms
3. Conducted in the Netherlands, public sector perspective, no discounting

2.6 CLINICAL / ECONOMIC QUESTION: CBT VERSUS WAIT LIST FOR THE MANAGEMENT OF ANXIETY IN CHILDREN AND YOUNG PEOPLE WITH AUTISM

Evidence profile - economic evidence

Study & country	Limitations	Applicability	Other comments	Incremental cost (£)[1]	Incremental effect (QALY)	ICER (£/QALY)	Uncertainty[1]
Guideline model	Potentially serious limitations[2]	Partially applicable[3]	• Time horizon: 38 weeks	Group-CBT: £387 Individual CBT: £2712	0.028	Group-CBT: £13,910 Individual CBT: £97,367	PSA: probability of CBT being cost-effective at £20,000/QALY: group-CBT: 0.53; individual CBT: 0

1. Costs expressed in 2012 UK pounds
2. Only intervention costs considered, resource use from RCTs included in guideline systematic review, efficacy data from 2 trials, PSA performed
3. NHS & PSS perspective, QALYs based on HUI3 (valuations elicited from Canadian population)