Appendix N: GRADE evidence profiles for all intervention studies

Abbreviations

CBT	cognitive behavioural therapy
CI	confidence interval
RR	relative risk
SOTP	sex offender treatment program

N.1 Interventions for promoting health and well being

N.1.1 Parent training for parent-child attachment for women with sub-threshold symptoms

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Parent training versus treatment as usual	promoting mental health and wellbeing in adults in contact with the criminal justice system	Relative (95% CI)	Absolute (95% CI)		
Depression (CES-D) (Scale from 0 to 60; lower better)												
1	randomised trials	serious [1]	not serious	not serious	serious [2]	none	62	53	Mean 15.3 (SD 11.8)	MD 1.70 lower (5.65 lower to 2.25 higher)	⨁⨁◯◯ LOW	CRITICAL
Number of participants with symptoms of depression (CES-D=>16)												
1	randomised trials	serious [1]	not serious	not serious	serious [2,3]	none	23/62 (37.1%)	25/53 (47.2%)	**RR 0.79** (0.51 to 1.21)	**99 fewer per 1,000** (from 99 more to 231 fewer)	⨁⨁◯◯ LOW	CRITICAL
Mother-child attachment: Reflective functioning (PDI) (Scale from -1 to 9; higher better)												
1	randomised trials	serious [1]	not serious	not serious	serious [2]	none	57	52	Mean 3.15 (SD 1.33)	MD 0.39 higher (0.15 lower to 0.93 higher)	⨁⨁◯◯ LOW	CRITICAL
Mother-child interaction: Dyadic attunement (behavioural observation) (scale from 11 to 55; higher better)												
1	randomised trials	serious [1]	not serious	not serious	serious [2]	none	51	37	-	MD 3.08 lower (6.39 lower to 0.23 higher)	⨁⨁◯◯ LOW	IMPORTANT
Mother-child interaction: Parent positive engagement (behavioural observation; scale from 5 to 25; higher better)												
1	randomised trials	serious [1]	not serious	not serious	serious [2]	none	51	37	-	MD 0.17 lower (1.44 lower to 1.10 higher)	⨁⨁◯◯ LOW	IMPORTANT
Mother-child interaction: Child involvement (behavioural observation; scale from 6 to 30; higher better)												
1	randomised trials	serious [1]	not serious	not serious	serious [2]	none	51	52	-	MD 0.37 lower (2.19 lower to 1.45 higher)	⨁⨁◯◯ LOW	IMPORTANT
Maternal perceptions of child: Warmth (MORS)												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Parent training versus treatment as usual	promoting mental health and wellbeing in adults in contact with the criminal justice system	Relative (95% CI)	Absolute (95% CI)		
1	randomised trials	serious [1]	not serious	not serious	serious [2]	none	31	40	-	SMD **0.44 higher** (0.04 lower to 0.91 higher)	⨁⨁◯◯ LOW	IMPORTANT
Maternal perceptions of child: Invasion (MORS)												
1	randomised trials	serious [1]	not serious	not serious	serious [2]	none	31	40	-	SMD **0.12 lower** (0.58 lower to 0.35 higher)	⨁⨁◯◯ LOW	IMPORTANT
Maternal perceptions of child: Intensity of problem behaviour (ECBI)												
1	randomised trials	serious [4]	not serious	not serious	serious [2]	none	78	25	-	SMD **0.29 lower** (0.74 lower to 0.16 higher)	⨁⨁◯◯ LOW	IMPORTANT
Maternal perceptions of child: Frequency of problem behaviour (ECBI)												
1	randomised trials	serious [4]	not serious	not serious	serious [2]	none	78	25	-	SMD **0.04 higher** (0.41 lower to 0.49 higher)	⨁⨁◯◯ LOW	IMPORTANT
Maternal perceptions of parenting: Involvement (APQ)												
1	randomised trials	serious [4]	not serious	not serious	serious [2]	none	77	25	-	SMD **0.08 lower** (0.53 lower to 0.37 higher)	⨁⨁◯◯ LOW	IMPORTANT
Maternal perceptions of parenting: Positive parenting (APQ)												
1	randomised trials	serious [4]	not serious	not serious	serious [2]	none	78	25	-	SMD **0.66 lower** (1.12 lower to 0.2 lower)	⨁⨁◯◯ LOW	IMPORTANT
Maternal perceptions of parenting: Poor monitoring/supervision (APQ)												
1	randomised trials	serious [4]	not serious	not serious	serious [2]	none	77	25	-	SMD **0.33 higher** (0.13 lower to 0.78 higher)	⨁⨁◯◯ LOW	IMPORTANT
Maternal perceptions of parenting: Inconsistent discipline (APQ) (Scale from 6 to 30; lower better)												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Parent training versus treatment as usual	promoting mental health and wellbeing in adults in contact with the criminal justice system	Relative (95% CI)	Absolute (95% CI)		
1	randomised trials	serious [4]	not serious	not serious	serious [2]	none	78	25	-	MD 3.02 lower (4.72 to 1.33 lower)	⨁⨁◯◯ LOW	IMPORTANT
Maternal perceptions of parenting: Corporal punishment (APQ) (Scale from 3 to 15; lower better)												
1	randomised trials	serious [4]	not serious	not serious	serious [2]	none	78	25	-	MD 0.29 lower (1.21 lower to 0.63 higher)	⨁⨁◯◯ LOW	IMPORTANT
Drop-out (all cause)												
2	randomised trials	serious [1,4]	not serious	not serious	serious [2,3]	none	54/182 (29.7%)	31/126 (24.6%)	**RR 1.12** (0.76 to 1.64)	**30 more per 1,000** (from 59 fewer to 157 more)	⨁⨁◯◯ LOW	IMPORTANT

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

1. Sleed (2013) - no blinding
2. Small sample size (N<400), no sample size calculation reported
3. 95% CI includes both no effect and clinically significant harm or benefit
4. Menting (2014) - unclear randomisation method and no blinding

N.1.2 Yoga for promoting mental health and wellbeing

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Yoga versus waitlist control	promoting mental health and wellbeing in adults in contact with the criminal justice system	Relative (95% CI)	Absolute (95% CI)		
Positive affect (PANAS) (Scale from 10 to 50; higher better)												
1	randomised trials	serious [1]	not serious	not serious	serious [2]	none	45	55	-	MD 5.94 higher (2.91 higher to 8.97 higher)	⨁⨁◯◯ LOW	CRITICAL
Negative affect (PANAS) (Scale from 10 to 50; lower better)												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Yoga versus waitlist control	promoting mental health and wellbeing in adults in contact with the criminal justice system	Relative (95% CI)	Absolute (95% CI)		
1	randomised trials	serious [1]	not serious	not serious	serious [2]	none	45	55	-	MD 4.13 lower (6.80 lower to 1.46lower)	⨁⨁◯◯ LOW	CRITICAL
Perceived stress (PSS) (Scale from 0 to 40; lower better)												
1	randomised trials	serious [1]	not serious	not serious	serious [2]	none	45	55	-	MD 4.67 lower (7.65 lower to 1.69 lower)	⨁⨁◯◯ LOW	CRITICAL
Psychological distress (BSI) (Scale from 0 to 212; lower better)												
1	randomised trials	serious [1]	not serious	not serious	serious [2]	none	45	55	-	MD 12.60 lower (22.82 lower to 2.38 lower)	⨁⨁◯◯ LOW	CRITICAL
Drop-out (all cause)												
1	randomised trials	serious [1]	not serious	not serious	serious [2,3]	none	42/87 (48.3%)	25/80 (31.3%)	**RR 1.54** (1.04 to 2.28)	**169 more per 1,000** (from 13 more to 400 more)	⨁⨁◯◯ LOW	IMPORTANT

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

1. Bilderbeck (2013) - no blinding, attrition bias (significantly higher dropout with yoga)
2. Study was an exploratory trial - without sample size calculation
3. 95% CI includes the possibility that the benefit is less than the minimum important difference

N.1.3 Meditation for promoting mental health and well-being

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Meditation	treatment as usual	Relative (95% CI)	Absolute (95% CI) [3]		
Desire to throw things or hit people within past month (study-specific measure)												
1	randomised trials	serious [1]	not serious	not serious	serious [2]	none	17	16	-	SMD **1.01 lower** (1.73 lower to 0.28 lower)	⨁⨁◯◯ LOW	IMPORTANT

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Meditation	treatment as usual	Relative (95% CI)	Absolute (95% CI) [3]		
Feelings of guilt within past month (study-specific measure)												
1	randomised trials	serious [1]	not serious	not serious	serious [2]	none	17	16	-	SMD **0.42 lower** (1.11 lower to 0.27 higher)	⨁⨁◯◯ LOW	IMPORTANT
Feelings of hopelessness within past month (study-specific measure)												
1	randomised trials	serious [1]	not serious	not serious	serious [2]	none	17	16	-	SMD **0.06 lower** (0.74 lower to 0.63 higher)	⨁⨁◯◯ LOW	IMPORTANT
Being bothered by nail biting within past month (study-specific measure)												
1	randomised trials	serious [1]	not serious	not serious	serious [2]	none	17	16	-	SMD **1.18 lower** (1.91 lower to 0.44 lower)	⨁⨁◯◯ LOW	IMPORTANT
Being bothered by sleeping difficulties within past month (study-specific measure)												
1	randomised trials	serious [1]	not serious	not serious	serious [2]	none	17	16	-	SMD **0.28 lower** (0.96 lower to 0.41 higher)	⨁⨁◯◯ LOW	IMPORTANT

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

1. Sumter (2009) - no blinding, unclear allocation concealment
2. Small sample size (N<400), no sample size calculation reported
3. It was not possible to calculate MD, so SMD is reported.

N.1.4 Physical exercise programmes versus exercise as usual for promoting mental health and well-being

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Physical exercise programme	exercise as usual	Relative (95% CI)	Absolute (95% CI)		
Change in Symptom Checklist-90-Revised (SCL-90-R) Global Severity Index (GSI) - CRT or HIST exercise programme versus exercise as usual (follow up: 39 weeks) (Scale from 0 to 4; lower better)												
1	randomised trials	very serious [1]	not serious	not serious	not serious	none	44	20	-	MD **0.17 lower** (0.21 lower to 0.12 lower)	⨁⨁◯◯ LOW	CRITICAL
Change in Symptom Checklist-90-Revised (SCL-90-R) Positive Symptom Total (PST) - CRT or HIST exercise programme versus exercise as usual (follow up: 39 weeks) (Scale from 0 to 90; lower better)												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Physical exercise programme	exercise as usual	Relative (95% CI)	Absolute (95% CI)		
1	randomised trials	very serious [1]	not serious	not serious	not serious	none	44	20	-	MD **7.08 lower** (9.15 lower to 5 lower)	⊕⊕◯◯ LOW	CRITICAL
Change in Symptom Checklist-90-Revised (SCL-90-R) Positive Symptom Distress Index (PSDI) - CRT or HIST exercise programme versus exercise as usual (follow up: 39 weeks) (Scale from 0 to 4; lower better)												
1	randomised trials	very serious [1]	not serious	not serious	not serious	none	44	20	-	MD **0.33 lower** (0.41 lower to 0.25 lower)	⊕⊕◯◯ LOW	CRITICAL

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

1. Battaglia 2015 - unclear allocation concealment, no blinding, per-protocol analysis

N.2 Interventions for substance misuse

N.2.1 Psychological interventions

CBT versus active intervention

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	CBT versus active intervention	Control	Relative (95% CI)	Absolute		
Days using cannabis (during treatment) - Self-report (Better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	68	27	-	MD 10.15 days higher (6.63 lower to 26.93 higher)	⊕⊕OO LOW	CRITICAL
Days using cannabis (during treatment) - Urine test (Better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	68	27	-	MD 17.13 days higher (0.92 to 33.34 higher)	⊕⊕OO LOW	CRITICAL
Days with positive urine test (during treatment) (Better indicated by lower values)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	none	38	37	-	MD 0.3 days higher (2.23 lower to 2.15 higher)	⊕OOO VERY LOW	CRITICAL
Days with positive breathalyzer test (during treatment) (Better indicated by lower values)												

1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[1]	none	38	37	-	MD 0.04 lower (0.46 lower to 0.44 higher)	⊕OOO VERY LOW	CRITICAL
Days abstinent (during treatment) - Alcohol (Better indicated by lower values)												
1	randomised trials	Very serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	36	35	-	MD 10.40 higher (1.53 to 19.27 higher)	⊕OOO VERY LOW	CRITICAL
Days abstinent (during treatment) - Drugs (Better indicated by lower values)												
1	randomised trials	Very serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	36	35	-	MD 0.70 higher (0.41 lower to 6.12 higher)	⊕OOO VERY LOW	CRITICAL
Addiction Severity Index (ASI-6): alcohol composite score (follow-up 26-38 weeks; Better indicated by lower values)												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	serious[1]	none	23	21	-	MD 0.10 lower (0.22 lower to 0.02 higher)	⊕OOO VERY LOW	CRITICAL
Addiction Severity Index (ASI-6): drug composite score (follow-up 26-38 weeks; Better indicated by lower values)												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	none	23	21	-	MD 0.02 lower (0.09 lower to 0.05 higher)	⊕OOO VERY LOW	CRITICAL
Weeks abstinent (follow-up 26-38 weeks; Better indicated by higher values)												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	serious[1]	none	23	21	-	MD 1.30 lower (4.4 lower to 1.8 higher)	⊕OOO VERY LOW	CRITICAL
Reincarceration (follow-up 26-38 weeks)												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	serious[1]	none	5/23 (21.7%)	9/21 (42.9%)	RR 0.51 (0.2 to 1.27)	210 fewer per 1000 (from 343 fewer to 116 more)	⊕OOO VERY LOW	CRITICAL

[1] 95% CI includes both no effect and the minimal important difference
[2] 95% CI includes the minimal important difference
[3] high risk of performance bias. Unclear risk for allocation concealment, detection, attrition, reporting and other bias
[4] high risk of concealment bias, unclear risk on all other dimensions

CBT versus control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**CBT versus control/TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Addiction Severity Index (ASI-6): alcohol composite score (Better indicated by lower values)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	39	32	-	SMD 0.37 lower (0.85 lower to 0.1 higher)	⊕OOO VERY	CRITICAL

											LOW	
Addiction Severity Index (ASI-6): drug composite score (Better indicated by lower values)												
2	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	none	39	32	-	SMD 0.28 lower (0.75 lower to 0.2 higher)	⊕OOO VERY LOW	CRITICAL
Abstinent in previous 3 months (6 month follow-up)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[5]	none	4/16 (25%)	2/11 (18.2%)	RR 1.38 (0.3 to 6.25)	69 more per 1000 (from 127 fewer to 955 more)	⊕⊕OO LOW	CRITICAL
Weeks abstinent (3 month follow-up) (Better indicated by lower values)												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[7]	none	23	21	-	SMD 0.24 lower (0.84 lower to 0.35 higher)	⊕OOO VERY LOW	CRITICAL
Reincarceration												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[8]	none	5/23 (21.7%)	9/21 (42.9%)	RR 0.51 (0.2 to 1.27)	210 fewer per 1000 (from 343 fewer to 116 more)	⊕OOO VERY LOW	CRITICAL
								42.9%		210 fewer per 1000 (from 343 fewer to 116 more)		

[1] one study high risk for performance bias. Remaining study high risk for 'other bias' and unclear risk for all other categories
[2] N<100 & CI -0.85-0.1
[3] No explanation was provided
[4] N<100 & CI -0.75-0.2
[5] very small number of events & CI 0.3-6.25
[6] high risk for 'other bias' and unclear risk for all other categories
[7] N<100 & CI -0.84-0.35
[8] very small number of events and CI 0.2-1.27

ACT versus CBT

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Psychoeducation versus active intervention**	**Control**	**Relative (95% CI)**	**Absolute**		
Addiction Severity Index (ASI-6): alcohol composite score (follow-up mean 42 weeks; Scale from 0 to 9; lower better)												
1	randomised trials	serious risk of bias[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	14	16	-	MD 0.04 lower (0.07 to 0.01 lower)	⊕⊕OO LOW	CRITICAL
Addicition Severity Index (ASI-6): drug composite score (Scale from 0 to 9; lower better)												
1	randomised	serious	no serious	no serious	very serious[3]	none	14	16	-	MD 0.01 lower	⊕OOO	CRITICAL

	trials	risk of bias[1]	inconsistency	indirectness						(0.05 lower to 0.03 higher)	VERY LOW	
Abstinent from drugs in previous 3 months												
1	randomised trials	serious risk of bias[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	6/14 (42.9%)	4/16 (25%)	RR 1.71 (0.6 to 4.86)	178 more per 1000 (from 100 fewer to 965 more)	⊕OOO VERY LOW	CRITICAL

1 High risk of performance and detection bias, all other domains low risk
2 optimal information size criterion not met
3 confidence interval includes both clinically significant benefit and harm

ACT versus waitlist

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**ACT**	**Waitlist**	**Relative (95% CI)**	**Absolute**		
Addiction Severity Index (ASI-6): alcohol composite score (follow-up mean 42 weeks; Better indicated by lower values)												
2	randomised trials	very serious[1]	very serious[2]	no serious indirectness	serious[3]	none	32	24	-	SMD 0.60 lower (1.72 lower to 0.53 higher)	⊕OOO VERY LOW	CRITICAL
Addiction Severity Index (ASI-6): drug composite score (follow-up mean 42 weeks; Better indicated by lower values)												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	30	22	-	SMD 0.44 lower (1.19 lower to 0.3 higher)	⊕OOO VERY LOW	CRITICAL
Abstinent from drugs in previous 3 months (follow-up mean 42 weeks)												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[3]	none	6/14 (42.9%)	2/11 (18.2%)	RR 2.36 (0.59 to 9.48)	247 more per 1000 (from 75 fewer to 1000 more)	⊕⊕OO LOW	CRITICAL

1 high risk of performance bias, unclear or mixed risk on three other facets
2 I^2 =75%, random effects model used and outcome downgraded for inconsistency
3 confidence interval includes both clinically significant benefit and harm
4 high risk of performance bias, unclear or mixed risk on two other facets

Mindfulness-based relapse prevention versus active intervention

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Mindfulness-based relapse prevention**	**Active intervention**	**Relative (95%**	**Absolute**		

									CI)			
Drug-use days (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	28	26	-	MD 0.46 lower (1.16 lower to 0.24 higher)	⊕OOO VERY LOW	CRITICAL
Short Inventory of Problems (SIP) follow-up (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	28	26	-	MD 7.30 lower (15.81 lower to 1.21 higher)	⊕OOO VERY LOW	CRITICAL
Addiction Severity Index: family-social composite score (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	28	26	-	MD 0.01 lower (0.09 lower to 0.07 higher)	⊕OOO VERY LOW	CRITICAL
Addiction Severity Index: legal composite score (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	28	26	-	MD 0.31 lower (0.45 to 0.17 lower)	⊕OOO VERY LOW	CRITICAL
Addiction Severity Index: medical composite score (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	28	26	-	MD 0.20 lower (0.37 to 0.03 lower)	⊕OOO VERY LOW	CRITICAL
Addiction Severity index: psychiatric compose score (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	28	26	-	MD 0.11 lower (0.22 lower to 0 higher)	⊕OOO VERY LOW	CRITICAL

[1] high risk of bias from blinding and other factors, unclear risk of bias on 5 other domains
[2] optimal information size criterion not met

Contingency management versus active intervention

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Contingency management versus active intervention**	**Control**	**Relative (95% CI)**	**Absolute**		
Days using cannabis (during treatment) - Self-report (Better indicated by lower values)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	158	105	-	SMD 0.01 higher (0.24 lower to 0.26 higher)	⊕⊕OO LOW	CRITICAL
Days using cannabis (during treatment) - Urine test (Better indicated by lower values)												
2	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	none	67	69	-	SMD 0.23 lower (0.57 lower to 0.11 higher)	⊕OOO VERY LOW	CRITICAL

Addiction Severity Index (ASI): marijuana composite score - Post-treatment (Better indicated by lower values)												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	very serious[6]	none	37	28	-	SMD 0.18 higher (0.32 lower to 0.67 higher)	⊕OOO VERY LOW	CRITICAL
Addiction Severity Index (ASI): marijuana composite score - Follow-up (Better indicated by lower values)												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	very serious[6]	none	37	28	-	SMD 0.11 higher (0.38 lower to 0.6 higher)	⊕OOO VERY LOW	CRITICAL
Days cannabis use per month - Post-treatment (Better indicated by lower values)												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	very serious[6]	none	37	28	-	SMD 0.5 higher (0 to 1 higher)	⊕OOO VERY LOW	CRITICAL
Days cannabis use per month - Follow-up (Better indicated by lower values)												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	very serious[6]	none	58	28	-	SMD 0.22 higher (0.24 lower to 0.67 higher)	⊕OOO VERY LOW	CRITICAL
Participants still in treatment at follow-up (follow-up mean 52 weeks)												
1	randomised trials	very serious[7]	no serious inconsistency	no serious indirectness	serious[6]	none	18/83 (21.7%)	22/82 (26.8%)	RR 0.81 (0.47 to 1.39)	51 fewer per 1000 (from 142 fewer to 105 more)	⊕OOO VERY LOW	CRITICAL
No. of days in treatment (follow-up mean 52 weeks; Better indicated by higher values)												
1	randomised trials	very serious[7]	no serious inconsistency	no serious indirectness	serious[6]	none	83	82	-	MD 3.00 lower (21.01 lower to 15.01 higher)	⊕OOO VERY LOW	CRITICAL

[1] One study high risk for performance and attrition bias, unclear for selection and reporting bias. Other study high risk for performance and unclear for allocation concealment and reporting bias
[2] Optimal information size criterion not met (N<400)
[3] high risk of bias, unclear for selection and reporting bias
[4] Optimal information size criterion not met (N<200) & CI includes both clinically significant harm and no effect
[5] performance bias is high risk, all other categories (except other) are unclear risk
[6] CI includes both clinically significant or harm and no effect
[7] high risk of blinding and outcome reporting bias, unclear risk of performance and concealment bias

Contingency management versus treatment as usual

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Contingency management**	**TAU**	**Relative (95% CI)**	**Absolute**		
Arrests for public drunkenness (Better indicated by lower values)												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	10	10	-	MD 1.70 fewer arrests (5.65 fewer to 2.25 more)	⊕⊕OO LOW	CRITICAL

[1] Optimal information size criterion not met (N<200); 95% CI of effect includes both clinically significant benefit and no effect

Motivational enhancement therapy versus active intervention

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Motivational enhancement therapy versus active intervention	Control	Relative (95% CI)	Absolute		
Percentage of days abstinent from alcohol (self-report) - 3 month follow-up (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	119	119	-	MD 9.5 % more (2.51 to 16.49 % more)	⊕OOO VERY LOW	CRITICAL
Percentage of days abstinent from alcohol (self-report) - 6 month follow-up (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	107	107	-	MD 4.8 % more (2.50 % fewer to 12.10 % more)	⊕OOO VERY LOW	CRITICAL
Percentage of days abstinent from alcohol (self-report) - 12 month follow-up (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	95	95	-	MD 0.8 % more (8.37 % fewer to 6.77 % more)	⊕OOO VERY LOW	CRITICAL
Percentage of days abstinent from alcohol and drugs - 3 month follow-up (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	119	119	-	MD 9.7 % more (0.7 % more to 18.63 % more)	⊕OOO VERY LOW	CRITICAL
Percentage of days abstinent from alcohol and drugs - 6 month follow-up (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	107	107	-	MD 5.2 % more (4.05 % fewer to 14.45 % more)	⊕OOO VERY LOW	CRITICAL
Percentage of days abstinent from alcohol and drugs - 12 month follow-up (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	95	95	-	MD 9.7 % more (0.7 % more to 18.63 % more)	⊕OOO VERY LOW	CRITICAL
Drinks per drinking days - 3 month follow-up (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	119	119	-	MD 1.7 drinks fewer (3.75 fewer to 0.35 more)	⊕OOO VERY LOW	CRITICAL

Drinks per drinking days - 6 month follow-up (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	107	107	-	MD 0.70 drinks more (0.93 fewer to 2.33 more)	⊕OOO VERY LOW	CRITICAL
Drinks per drinking days - 12 month follow-up (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	97	95	-	MD 0.30 drinks fewer (1.90 fewer to 1.33 more)	⊕OOO VERY LOW	CRITICAL
Percentage of days with cannabis use (during treatment) (Better indicated by lower values)												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	very serious[3]	none	69	67	-	SMD 0.1 lower (0.44 lower to 0.24 higher)	⊕OOO VERY LOW	CRITICAL
Percentage of urine tests positive for cannabis use (during treatment) (Better indicated by lower values)												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	very serious[3]	none	69	67	-	SMD 0.91 lower (1.27 to 0.56 lower)	⊕OOO VERY LOW	CRITICAL
Self-reported motivation to take steps to change substance abuse scores (Better indicated by higher values)												
1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	serious[2]	none	18	9	-	MD 4.10 higher (5.77 lower to 13.97 higher)	⊕OOO VERY LOW	CRITICAL

[1] High performance bias + unclear for 4 other bias types.
[2] Optimal information size criterion not met (N < 400)
[3] Attrition bias (more than 50% of sample)
[4] High performance bias + high attrition bias + unclear on 3 other types of bias.
[5] High risk of performance, detection and other bias, unclear selection and attrition bias

Motivational interviewing or feedback versus active intervention

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Motivational interviewing/Motivational feedback versus control/TAU**	**Control**	**Relative (95% CI)**	**Absolute**		
Self-reported drug use - 1 month follow-up												
1	randomised trials	very serious[1]	no serious inconsistency	very serious[2]	very serious[3]	none	24/39 (61.5%)	19/40 (47.5%)	RR 1.3 (0.86 to 1.95)	142 more per 1000 (from 66 fewer to 451 more)	⊕OOO VERY LOW	CRITICAL
Self-reported days with drug use in past 30 days (10 month follow-up) (Better indicated by lower values)												
1	randomised	very	no serious	no serious	very	none	90	24	-	SMD 0.04 higher	⊕OOO	CRITICAL

	trials	serious[4]	inconsistency	indirectness	serious[5]					(0.41 lower to 0.49 higher)	VERY LOW	
Urine test positive for drug use (during study period)												
1	randomised trials	very serious[6]	no serious inconsistency	very serious[2]	very serious[3]	none	15/39 (38.5%)	14/40 (35%)	RR 1.1 (0.62 to 1.96)	35 more per 1000 (from 133 fewer to 336 more)	⊕OOO VERY LOW	CRITICAL
Self-reported alcohol use - 1 month follow-up												
1	randomised trials	very serious[1]	no serious inconsistency	very serious[2]	very serious[3]	none	24/39 (61.5%)	19/40 (47.5%)	RR 1.3 (0.86 to 1.95)	142 more per 1000 (from 66 fewer to 451 more)	⊕OOO VERY LOW	CRITICAL
Days with illegal activity in past 30 days (10 month follow-up) (Better indicated by lower values)												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	very serious[5]	none	80	23	-	SMD 0.07 higher (0.4 lower to 0.53 higher)	⊕OOO VERY LOW	CRITICAL
Drop-out from subsequent treatment - binge drinking group (follow-up mean 26 weeks)												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[8]	none	2/11 (18.2%)	8/12 (66.7%)	RR 0.27 (0.07 to 1.02)	487 fewer per 1000 (from 620 fewer to 13 more)	⊕⊕OO LOW	CRITICAL
Drop-out from subsequent treatment - no binge drinking group (follow-up mean 26 weeks)												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[8]	none	5/20 (25%)	4/15 (26.7%)	RR 0.94 (0.3 to 2.91)	16 fewer per 1000 (from 187 fewer to 509 more)	⊕⊕OO LOW	CRITICAL
Number of subsequent treatment sessions attended - binge drinking group (follow-up mean 26 weeks; Better indicated by higher values)												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[8]	none	10	9	-	MD 11.16 higher (3.86 to 18.46 higher)	⊕⊕OO LOW	CRITICAL
Number of subsequent treatment sessions attended - no binge drinking group (follow-up mean 26 weeks; Better indicated by higher values)												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[8]	none	20	15	-	MD 1.65 lower (8.28 lower to 4.98 higher)	⊕⊕OO LOW	CRITICAL
Speciality addiction clinic attendance												
1	randomised trials	very serious[9]	no serious inconsistency	no serious indirectness	serious[8]	none	8/17 (47.1%)	4/13 (30.8%)	RR 1.53 (0.59 to 3.99)	163 more per 1000 (from 126 fewer to 920 more)	⊕OOO VERY LOW	CRITICAL

1 high performance bias + high other bias + 3 unclear;
2 very serious limitations (outcome)

3 Optimal information size criterion not met (n = 79)
4 high performance and detection bias.
5 Optimal information size criterion not met (n = 114)
6 high performance bias + high other bias + 3 unclear
7 High risk of performance bias, unclear selection and detection bias
8 Optimal information size criterion not met
9 High risk of blinding, performance and detection bias, unclear selection and concealment bias

Group counselling versus treatment as usual

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Group counselling	TAU	Relative (95% CI)	Absolute		
Rearrest (12 month follow-up)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	46/85 (54.1%)	24/43 (55.8%)	RR 0.97 (0.7 to 1.35)	17 fewer per 1000 (from 167 fewer to 195 more)	⊕OOO VERY LOW	CRITICAL
Number of reconvictions (12 month follow-up) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	99	50	-	MD 0.10 fewer reconvictions (0.68 fewer to 0.48 more)	⊕OOO VERY LOW	CRITICAL
Reincarceration (12 month follow-up)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	24/85 (28.2%)	14/43 (32.6%)	RR 0.87 (0.5 to 1.5)	42 fewer per 1000 (from 163 fewer to 163 more)	⊕OOO VERY LOW	CRITICAL
Days incarcerated (12 month follow-up) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	99	50	-	MD 0.30 days more (28.9 fewer to 29.5 more)	⊕OOO VERY LOW	CRITICAL
Self-reported drug use (12 month follow-up) - Marijuana												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	31/85 (36.5%)	24/43 (55.8%)	RR 0.65 (0.44 to 0.96)	195 fewer per 1000 (from 22 fewer to 313 fewer)	⊕OOO VERY LOW	CRITICAL

Self-reported drug use (12 month follow-up) - LSD												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	14/85 (16.5%)	9/43 (20.9%)	RR 0.79 (0.37 to 1.67)	44 fewer per 1000 (from 132 fewer to 140 more)	⊕OOO VERY LOW	CRITICAL
Self-reported drug use (12 month follow-up) - Speed												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	14/85 (16.5%)	4/43 (9.3%)	RR 1.77 (0.62 to 5.05)	72 more per 1000 (from 35 fewer to 377 more)	⊕OOO VERY LOW	CRITICAL
Self-reported drug use (12 month follow-up) - Heroin												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2,3]	none	7/85 (8.2%)	3/43 (7%)	RR 1.18 (0.32 to 4.34)	13 more per 1000 (from 47 fewer to 233 more)	⊕OOO VERY LOW	CRITICAL

1 high risk of performance and detection bias. Unclear risk of remaining categories (other than 'other' bias)
2 Imprecision: optimal information size criterion not met
3 Confidence interval of effect includes both clinically significant benefit and harm

Self-help versus control for substance misuse

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Self-help versus control**	**Control**	**Relative (95% CI)**	**Absolute**		
Subsequent bookings (12 month follow-up)												
1	randomised trials	serious risk of bias[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	49/98 (50%)	56/85 (65.9%)	RR 0.76 (0.59 to 0.97)	158 fewer per 1000 (from 20 fewer to 270 fewer)	⊕⊕OO LOW	CRITICAL

[1] Sample size not reported. 183 participants were randomised but is unclear how many were assessed for eligibility

N.2.2 Pharmacological interventions

Naloxone versus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Naloxone versus placebo	Control	Relative (95% CI)	Absolute		
Discontinued medication												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	16/55 (29.1%)	8/42 (19%)	RR 1.53 (0.72 to 3.23)	101 more per 1000 (from 53 fewer to 425 more)	⊕OOO VERY LOW	CRITICAL
Number of urine tests positive during treatment												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	5/73 (6.8%)	10/90 (11.1%)	RR 0.62 (0.22 to 1.72)	42 fewer per 1000 (from 87 fewer to 80 more)	⊕OOO VERY LOW	CRITICAL

1 unclear risk of bias for detection and performance bias.
2 optimal information size criterion not met; confidence interval for the effect includes clinically significant benefit

Naltrexone versus active intervention for drug misuse

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Naltrexone versus active intervention	Active intervention	Relative (95% CI)	Absolute		
Retained in treatment												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	17/34 (50%)	5/17 (29.4%)	RR 1.7 (0.76 to 3.82)	206 more per 1000 (from 71 fewer to 829 more)	⊕OOO VERY LOW	CRITICAL
Urine test positive for drugs (during treatment) - Alcohol												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	1/34 (2.9%)	1/17 (5.9%)	RR 0.5 (0.03 to 7.51)	29 fewer per 1000 (from 57 fewer to 383 more)	⊕OOO VERY LOW	CRITICAL
Urine test positive for drugs (during treatment) - Amphetamine												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	0/34 (0%)	0/17 (0%)	not estimable	not estimable[7]	⊕OOO VERY LOW	CRITICAL
Urine test positive for drugs (during treatment) - Benzodiazepine												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	1/34 (2.9%)	1/17 (5.9%)	RR 0.5 (0.03 to 7.51)	29 fewer per 1000 (from 57 fewer to 383 more)	⊕OOO VERY LOW	CRITICAL
Urine test positive for drugs (during treatment) - Cocaine												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	11/34 (32.4%)	8/17 (47.1%)	RR 0.69 (0.34 to 1.38)	146 fewer per 1000 (from 311 fewer to 179 more)	⊕OOO VERY LOW	CRITICAL
Urine test positive for drugs (during treatment) - Marijuana												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	4/34 (11.8%)	3/17 (17.6%)	RR 0.67 (0.17 to 2.65)	58 fewer per 1000 (from 146 fewer to 291 more)	⊕OOO VERY LOW	CRITICAL
Urine test positive for drugs (during treatment) - Opiates												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	3/34 (8.8%)	5/17 (29.4%)	RR 0.3 (0.08 to 1.11)	206 fewer per 1000 (from 271 fewer to 32 more)	⊕OOO VERY LOW	CRITICAL
Cocaine use (post-treatment)												
2	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	none	14/47 (29.8%)	10/49 (20.4%)	RR 1.34 (0.73 to 2.48)	69 more per 1000 (from 55 fewer to 302 more)	⊕OOO VERY LOW	CRITICAL

Opioid use (post-treatment)												
2	randomised trials	very serious[3,4]	no serious inconsistency	no serious indirectness	very serious[2]	none	13/47 (27.7%)	24/49 (49%)	RR 0.55 (0.35 to 0.87)	220 fewer per 1000 (from 64 fewer to 318 fewer)	⊕OOO VERY LOW	CRITICAL
Injection drug use (post-treatment)												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	very serious[2]	none	4/16 (25%)	1/17 (5.9%)	RR 4.25 (0.53 to 34.1)	191 more per 1000 (from 28 fewer to 1000 more)	⊕OOO VERY LOW	CRITICAL
Days of drug use per month (6 month follow-up) - Amphetamine (Better indicated by lower values)												
1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	very serious[6]	none	23	21	-	MD 2.50 higher (3.86 lower to 8.86 higher)	⊕OOO VERY LOW	CRITICAL
Days of drug use per month (6 month follow-up) - Benzodiazepine (Better indicated by lower values)												
1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	very serious[6]	none	23	21	-	MD 2.0 higher (4.49 lower to 8.49 higher)	⊕OOO VERY LOW	CRITICAL
Days of drug use per month (6 month follow-up) - Heroin (Better indicated by lower values)												
1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	very serious[6]	none	23	21	-	MD 4.60 lower (12.74 lower to 3.54 higher)	⊕OOO VERY LOW	CRITICAL
Reincarceration												
3	randomised trials	very serious[1,5]	no serious inconsistency	no serious indirectness	very serious[6]	none	19/73 (26%)	21/55 (38.2%)	RR 0.64 (0.39 to 1.06)	137 fewer per 1000 (from 233 fewer to 23 more)	⊕OOO VERY LOW	CRITICAL
Reincarceration - During treatment												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	9/34 (26.5%)	9/17 (52.9%)	RR 0.5 (0.24 to 1.02)	265 fewer per 1000 (from 402 fewer to 11 more)	⊕OOO VERY LOW	CRITICAL
Reincarceration - Post-treatment												

1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	very serious[2]	none	5/16 (31.3%)	7/17 (41.2%)	RR 0.76 (0.3 to 1.91)	99 fewer per 1000 (from 288 fewer to 375 more)	⊕OOO VERY LOW	CRITICAL
Reincarceration - 6 month follow-up												
1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	very serious[2]	none	5/23 (21.7%)	5/21 (23.8%)	RR 0.91 (0.31 to 2.71)	21 fewer per 1000 (from 164 fewer to 407 more)	⊕OOO VERY LOW	CRITICAL
Parole violations (post-treatment)												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	none	2/31 (6.5%)	9/32 (28.1%)	RR 0.23 (0.05 to 0.98)	217 fewer per 1000 (from 6 fewer to 267 fewer)	⊕OOO VERY LOW	CRITICAL
Drug charges (post-treatment)												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	none	3/31 (9.7%)	1/32 (3.1%)	RR 3.1 (0.34 to 28.19)	66 more per 1000 (from 21 fewer to 850 more)	⊕OOO VERY LOW	CRITICAL
Days of criminal activity per month (6 month follow-up) (Better indicated by lower values)												
1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	very serious[6]	none	23	21	-	Mean 14.4 days (SD 13.11)	⊕OOO VERY LOW	CRITICAL

1 Cornish 1997 - unclear randomisation and allocation concealment; unclear blinding; ITT analysis

2 Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or two boundaries of the defined minimally important difference (MID) for the outcome (imprecision) respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.

3 Caviello 2010 - Unclear randomisation and allocation concealment; unclear blinding; available case analysis

4 Lee 2016 - Appropriate randomisation and unclear allocation concealment; No blinding to participants; ITT analysis

5 Lobmaier 2010 - appropriate randomisation and allocation concealment; no blinding; ITT analysis

6 Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or two boundaries of the defined minimally important difference (MID) for the outcome (imprecision) respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.

7 No event in either arm of the trial.

Methadone versus waitlist control

Quality assessment							No of patients		Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Methadone versus waitlist control	Control	Relative (95% CI)	Absolute		
Drop-out												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	152/191 (79.6%)	123/191 (64.4%)	RR 1.24 (1.09 to 1.4)	155 more per 1000 (from 58 more to 258 more)	⊕OOO VERY LOW	CRITICAL
Positive for opioids - Post-treatment												
2	randomised trials	very serious[1,3]	serious[4]	no serious indirectness	no serious imprecision	none	82/277 (29.6%)	90/270 (33.3%)	RR 0.86 (0.61 to 1.23)	47 fewer per 1000 (from 130 fewer to 77 more)	⊕OOO VERY LOW	CRITICAL
Positive for opioids - 2 month follow-up												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	10/106 (9.4%)	12/101 (11.9%)	RR 0.79 (0.36 to 1.76)	25 fewer per 1000 (from 76 fewer to 90 more)	⊕⊕OO LOW	CRITICAL
Positive for opioids - 3 month follow-up												
2	randomised trials	very serious[1,3]	no serious inconsistency	no serious indirectness	serious[5]	none	40/233 (17.2%)	51/211 (24.2%)	RR 0.7 (0.5 to 0.99)	73 fewer per 1000 (from 2 fewer to 121 fewer)	⊕OOO VERY LOW	CRITICAL
Positive for opioids - 4 month follow-up												
2	randomised trials	very serious[1,3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	38/280 (13.6%)	39/258 (15.1%)	RR 0.91 (0.62 to 1.35)	14 fewer per 1000 (from 57 fewer to 53 more)	⊕⊕OO LOW	CRITICAL
Reincarceration (4 year follow-up)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	143/191 (74.9%)	137/191 (71.7%)	RR 1.04 (0.92 to 1.18)	29 more per 1000 (from 57 fewer to 129 more)	⊕OOO VERY LOW	CRITICAL

1 Dolan 2003/2005 - appropriate randomisation and allocation concealment; unclear blinding and available case analysis
2 Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or two boundaries of the defined minimally important difference(MID) for the outcome (imprecision) respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.
3 Shearer 2006 - unclear randomisation and allocation concealment; unclear blinding; available case analysis
4 Evidence was downgraded by one level due to serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of 50%-74.99%) and by two levels due to very serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of >75%).
5 Rich 2015 - appropriate randomisation and allocation concealment; unclear blinding; ITT analysis

Alpha-adrenergic agonists versus opioid maintenance

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Alpha-adrenergics	Opioid maintenance	Relative (95% CI)	Absolute		
Total withdrawal symptoms (follow-up mean 10 days; Better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	29	34	-	MD 24 higher (73.86 lower to 121.86 higher)	⊕⊕OO LOW	CRITICAL

[1] optimal information size criterion not met; confidence interval of effect includes both appreciable benefit and harm

Opioid substitution versus active intervention or placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Opioid substitution therapy versus active intervention	Control	Relative (95% CI)	Absolute		
Drop-out												
2	randomised trials	very serious[1]	no serious inconsistency	serious[2]	serious[3]	none	21/102 (20.6%)	29/104 (27.9%)	RR 0.75 (0.46 to 1.22)	70 fewer per 1000 (from 151 fewer to 61 more)	⊕OOO VERY LOW	CRITICAL
Abstinence - Post-treatment												

1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious[5]	none	74/100 (74%)	79/113 (69.9%)	RR 1.06 (0.9 to 1.25)	42 more per 1000 (from 70 fewer to 175 more)	⊕⊕OO LOW	CRITICAL
Abstinence - 1 month follow-up												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[6]	none	45/72 (62.5%)	64/87 (73.6%)	RR 0.85 (0.68 to 1.06)	110 fewer per 1000 (from 235 fewer to 44 more)	⊕OOO VERY LOW	CRITICAL
Abstinence - 3 month follow-up												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[7]	none	31/46 (67.4%)	27/48 (56.3%)	RR 1.2 (0.87 to 1.65)	113 more per 1000 (from 73 fewer to 366 more)	⊕OOO VERY LOW	CRITICAL
Abstinence - 6 month follow-up												
2	randomised trials	very serious[4,8]	no serious inconsistency	no serious indirectness	very serious[9]	none	26/75 (34.7%)	21/75 (28%)	RR 1.08 (0.74 to 1.59)	22 more per 1000 (from 73 fewer to 165 more)	⊕OOO VERY LOW	CRITICAL
Opioid abuse (3 month follow-up)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	very serious[10]	none	32/60 (53.3%)	37/56 (66.1%)	RR 0.81 (0.6 to 1.09)	126 fewer per 1000 (from 264 fewer to 59 more)	⊕OOO VERY LOW	CRITICAL
Self-reported injection drug use - Post-treatment												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[11]	none	8/24 (33.3%)	7/12 (58.3%)	RR 0.57 (0.27 to 1.2)	251 fewer per 1000 (from 426 fewer to 117 more)	⊕⊕OO LOW	CRITICAL
Self-reported injection drug use - 3 month follow-up												

1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[12]	none	7/24 (29.2%)	6/12 (50%)	RR 0.58 (0.25 to 1.35)	210 fewer per 1000 (from 375 fewer to 175 more)	⊕⊕OO LOW	CRITICAL
Number of times rearrested (3 month follow-up) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	very serious[10]	none	60	56	-	SMD 0.02 lower (0.39 lower to 0.34 higher)	⊕OOO VERY LOW	CRITICAL
Rearrest for drug crimes (3 month follow-up)												
1	randomised trials	serious[12]	no serious inconsistency	serious[12]	very serious[13]	none	8/60 (13.3%)	13/56 (23.2%)	RR 0.57 (0.26 to 1.28)	100 fewer per 1000 (from 172 fewer to 65 more)	⊕OOO VERY LOW	CRITICAL
Reincarceration (post-treatment)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	very serious[10]	none	24/60 (40%)	28/56 (50%)	RR 0.8 (0.53 to 1.2)	100 fewer per 1000 (from 235 fewer to 100 more)	⊕OOO VERY LOW	CRITICAL

1 high risk performance of bias
2 serious indirectness Maguara 2009 due to population)
3 Optimal information size criterion not met (combined n = 206)
4 high risk performance of bias
5 Optimal information size criterion not met (n = 213)
6 Optimal information size criterion not met (n = 159)
7 Optimal information size criterion not met (n = 94)
8 ROB - Sheared = high performance bias + unclear detection bias + 2 unclear bias.
9 Optimal information size criterion not met (Combined n = 150)
10 Optimal information size criterion not met (n = 116)
11 Optimal information size criterion not met (n = 36)
12 Optimal information size criterion not met (events<100) and CI of effect includes appreciable benefit and harm

N.2.3 Combined pharmacological and psychological interventions

Antidepressants plus psychological therapy versus psychological therapy alone

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antidepressants + psychological therapy	Psychological therapy only	Relative (95% CI)	Absolute		
No. participants who failed to complete treatment (follow-up mean 12 weeks)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	13/31 (41.9%)	9/29 (31%)	RR 1.35 (0.68 to 2.67)	109 more per 1000 (from 99 fewer to 518 more)	⊕⊕OO LOW	CRITICAL
Spielberger state anxiety inventory score (follow-up mean 12 weeks; Scale from 20 to 80; lower better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	31	29	-	MD 0.30 lower (6.44 lower to 5.84 higher)	⊕⊕OO LOW	CRITICAL
Hamilton depression rating scale score (follow-up mean 12 weeks; Scale from 0 to 52; lower better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	31	29	-	MD 3.10 lower (6.18 to 0.02 lower)	⊕⊕OO LOW	CRITICAL

[1] unclear selection, detection and attrition bias
[2] optimal information size criterion not met

N.2.4 Support and educational interventions

Psychoeducation versus control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Psychoeducation versus control/TAU	Control	Relative (95% CI)	Absolute		
Number of days with uncontrolled drinking (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	18	16	-	MD 4.85 days fewer (11.46 fewer to 1.76 more)	⊕OOO VERY LOW	CRITICAL

[1] high risk for performance, detection and selective reporting
[2] Optimal information size criterion not met (N<400); 95% CI of effect includes both appreciable benefit and harm

Employment workshop versus control or treatment as usual

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Employment workshops	Control/TAU	Relative (95% CI)	Absolute		
No. of participants employed (follow-up 12-52 weeks)												
2	randomised trials	very serious[1]	very serious[2]	serious[3]	serious[4]	none	220/272 (80.9%)	189/257 (73.5%)	RR 1.24 (0.84 to 1.81)	176 more per 1000 (from 118 fewer to 596 more)	⊕OOO VERY LOW	CRITICAL
Days in paid employment (follow-up mean 52 weeks; Better indicated by higher values)												
1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	no serious imprecision	none	244	233	-	MD 10.20 higher (11.8 lower to 32.2 higher)	⊕⊕OO LOW	CRITICAL

1 high risk of performance, detection and reporting bias, unclear bias on 3 other dimensions
2 I2=73%; random effects model used; no reasons for this heterogeneity were identified; study effect estimates were RR=1.58 [1.06, 2.36] for Hall (1961) and RR = 1.06 [0.97, 1.17] for Webster (2014)
3 Hall 1981-unclear whether the population have a current drug or other mental health problem
4 Hall 1981, small sample size
5 high risk of detection and performance bias, unclear risk on 3 other domains

N.2.5 Physical interventions

Acupuncture versus active intervention

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Acupuncture	active intervention	Relative (95% CI)	Absolute		
Drop-out												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	50/82 (61%)	32/76 (42.1%)	RR 1.45 (1.06 to 1.99)	189 more per 1000 (from 25 more to 417 more)	⊕⊕OO LOW	CRITICAL
Urine test positive for drug use post-treatment												

2	randomised trials	very serious[3]	very serious[4]	serious[5]	very serious[6]	none	15/46 (32.6%)	8/62 (12.9%)	RR 3.65 (0.33 to 41)	342 more per 1000 (from 86 fewer to 1000 more)	⊕OOO VERY LOW	CRITICAL

1 allocation concealment, attrition and selective reporting all high risk of bias
2 Optimal information size criterion not met (N<300 events)
3 Both studies had allocation concealment, attrition and selective reporting all high risk of bias
4 I^2 66% - random effects model used; large variation in effect sizes: Berman 16.39, Konefal 1.59, but no explanation for the heterogeneity was identified
5 For one study (Konefal 1995) - only 51% of participants were in contact with CJS
6 Optimal information size criterion not met (N <300 events) and CI of effect includes both appreciable benefit and harm

N.3 Interventions for 'other' mental health disorders

N.3.1 Depression

Psychotherapy vs PSYCHOED

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Psychotherapy**	**PSYCHOED**	**Relative (95% CI)**	**Absolute**		
Depression by HRSD scales (Scale from 0 to 52; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	19	19	-	MD 6.5 lower (12.52 to 0.48 lower)	⊕OOO VERY LOW	CRITICAL
Depression by HRSD scales (13 weeks Follow-up) (Scale from 0 to 52; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	19	19	-	MD 3.8 higher (3.83 lower to 11.43 higher)	⊕OOO VERY LOW	CRITICAL

[1] Johnson 2012 - Unclear risk of bias, unclear blinding of participants and care administrators, blinding of outcome assessors, low attrition bias, unclear selective outcome bias, low other risk of bias
[2] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.

Group therapy vs Individual therapy for depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Group therapy	Individual therapy	Relative (95% CI)	Absolute		
Depression by BDI (Scale from 0 to 20; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	5	5	-	MD 3.2 lower (13.56 lower to 7.16 higher)	⊕OOO VERY LOW	CRITICAL
Depression by Hopeless scale (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	5	5	-	MD 2.6 higher (4.98 lower to 10.18 higher)	⊕OOO VERY LOW	CRITICAL
Depression by MMPI D scale (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	Serious[2]	none	5	5	-	MD 12.6 higher (3.38 lower to 28.58 higher)	⊕OOO VERY LOW	CRITICAL
Depression by MMPI D scale (39 weeks Follow-up) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	5	5	-	MD 4.8 higher (9.68 lower to 19.28 higher)	⊕OOO VERY LOW	CRITICAL
Depression by Multiple affect adjective Check list D scale (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	5	5	-	MD 0.6 higher (4.93 lower to 6.13 higher)	⊕OOO VERY LOW	CRITICAL

[1] Wilson 1990 - Unclear selection bias, No blinding, low attrition rate, low selective outcome reporting, low other risk of bias
[2] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.

Arts-based therapy vs TAU for depression

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Arts-based therapy	TAU	Relative (95% CI)	Absolute		
Change in Adult Nowicki-Strickland Locus of Control Scale (ANS) (Better indicated by lower values)												
1	randomised	very	no serious	no serious	serious[2]	none	72	50	-	MD 3.88 lower (8.15 lower	⊕OOO	CRITICAL

	trials	serious[1]	inconsistency	indirectness						to 0.39 higher)	VERY LOW	
Change in Adult Nowicki-Strickland Locus of Control Scale (ANS) - Male (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	Serious[2]	none	37	25	-	MD 2.26 lower (4.18 to 0.34 lower)	⊕OOO VERY LOW	CRITICAL
Change in Adult Nowicki-Strickland Locus of Control Scale (ANS) - Female (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	35	25	-	MD 6.81 lower (11.97 to 1.65 lower)	⊕OOO VERY LOW	CRITICAL
Change in Beck Depression Inventory (BDI): Total (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	Serious[2]	none	111	45	-	MD 6.5 lower (9.33 to 3.67 lower)	⊕OOO VERY LOW	CRITICAL
Change in Beck Depression Inventory (BDI): Total - Male (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	35	25	-	MD 6.81 lower (11.97 to 1.65 lower)	⊕OOO VERY LOW	CRITICAL
Change in Formal Elements of Arts Therapy Scale rating guide(FEATS): Prominence of color (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	65	19	-	MD 0.81 lower (1.51 to 0.11 lower)	⊕OOO VERY LOW	LIMITEDIMPORTANT
Change in Formal Elements of Arts Therapy Scale rating guide (FEATS): color fit (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	65	19	-	MD 0.45 lower (0.84 to 0.06 lower)	⊕OOO VERY LOW	LIMITED IMPORTANT

[1] *Gussak 2009 - Unclear randomization and allocation, No blinding of patients and care administrators, Blinding of outcome assessors*Unclear randomization and allocation, No blinding of patients and care administrators, Blinding of outcome assessors, (Gussak 2009)

[2] *The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries*

N.3.2 Vulnerable inmates with suicidal risks

Social problem solving group vs No treatment control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Social problem solving group for vulnerable**	**No treatment**	**Relative (95%**	**Absolute**		

							inmates		CI)			
Depression by HADS scale (Scale from 0 to 21; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	23	23	-	MD 3.6 lower (5.76 to 1.44 lower)	⊕OOO VERY LOW	CRITICAL
Anxiety by HADS scales (Scale from 0 to 21; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	Serious[2]	none	23	23	-	MD 2.9 lower (4.67 to 1.13 lower)	⊕OOO VERY LOW	CRITICAL
Depression by Beck Hopeless scales (Scale from 0 to 20; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	23	23	-	MD 2.5 lower (4.89 to 0.11 lower)	⊕OOO VERY LOW	CRITICAL
Decision making ability by SPSI:R scales (Scale from 0 to 21; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	23	23	-	MD 5.3 higher (2.66 to 7.94 higher)	⊕⊕OO LOW	CRITICAL
Depression by HADS scale (13 weeks Follow-up) (Scale from 0 to 20; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	23	23	-	MD 3.3 lower (5.19 to 1.41 lower)	⊕OOO VERY LOW	CRITICAL
Anxiety by HADS scales (13 weeks Follow-up) (Scale from 0 to 21; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	23	23	-	MD 2.7 lower (4.61 to 0.79 lower)	⊕OOO VERY LOW	CRITICAL
Depression by Beck Hopeless scales (13 weeks Follow-up) (Scale from 0 to 20; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	23	23	-	MD 2.8 lower (5.13 to 0.47 lower)	⊕OOO VERY LOW	CRITICAL

[1]Biggam 2002 - Unclear risk of selection bias, No blinding, low attrition bias, unclear selective outcome reporting, low other risk of bias
[2] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries .

N.3.3 Anxiety disorders

Self-help therapy vs Wait-list control

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Self-help materials**	**Control**	**Relative (95% CI)**	**Absolute**		
Anxiety by HADS scales (Scale from 0 to 21; lower better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	15	18	-	MD 1.06 lower (3.63 lower to 1.51 higher)	⊕⊕OO LOW	CRITICAL
Anxiety by HADS scales (4 weeks follow-up) (Scale from 0 to 21; lower better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	15	18	-	MD 2.98 lower (5.82 to 0.14 lower)	⊕⊕OO LOW	CRITICAL

[1] *Maunder 2009 - low selection risk of bias, No blinding of participants but blinding of care administrators (+), unclear outcome assessor, unclear attrition risk of bias, unclear other risk of bias (blocked randomization with single blinded trial)*
[2] *The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries .*

N.3.4 PTSD

Psychotherapy vs Wait-list/No-contact control

Quality assessment							**No of patients**		**Effect**		**Quality**	**Importance**
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Psychotherapy**	**Control**	**Relative (95% CI)**	**Absolute**		
Trauma by TSI - Group Therapy (Wait-list/No-contact Control) (Scale from 0 to 300; lower better)												
2	randomised trials	very serious[1,2]	very serious[3]	no serious indirectness	serious[4]	none	17	23	-	MD 11.67 lower (30.36 lower to 7.02 higher)	⊕OOO VERY LOW	CRITICAL

[1] *Cole 2007 - high risks of selection bias, No blinding, Unclear attrition bias, low selective outcome bias and low other risk of bias*
[2] *Bradley 2003 - unclear risks of selection bias, No blinding, Unclear attrition, High selective outcomes bias and low other risks of bias* I2=83%; studies combined by randomised model because similar population, intervention and the outcome measured by same measure.
[3] *Evidence was downgraded by one level due to serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of 50%-74.99%) and by two levels due to very serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of >75%).*
[4] *The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.*

TIR vs Wait-list control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	TIR	Wait-list control	Relative (95% CI)	Absolute		
Depression by BDI - Traumatic Incident Reduction (Scale from 0 to 63; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	56	67	-	MD 3.8 lower (7.52 to 0.08 lower)	⊕OOO VERY LOW	CRITICAL
Depression by BDI total (13 weeks Follow-up) (Scale from 0 to 63; lower better))												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[2]	none	56	67	-	MD 7.8 lower (12.64 to 2.96 lower)	⊕OOO VERY LOW	CRITICAL
PTSD by PSS scales at post-treatment (Scale from 0 to 51; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	Serious[2]	none	56	67	-	MD 4.1 lower (7.96 to 0.24 lower)	⊕OOO VERY LOW	CRITICAL
PTSD by PSS scales (13 weeks follow-up) (Scale from 0 to 51; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	56	67	-	MD 7.3 lower (11.49 to 3.11 lower)	⊕OOO VERY LOW	CRITICAL
Generalized Expectancy for Success Scale at post-treatment (Scale from 30 to 150; higher better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	56	67	-	MD 15.9 higher (5.7 to 26.1 higher)	⊕OOO VERY LOW	CRITICAL
Generalized Expectancy for Success Scale (13 weeks follow-up) (Scale from 30 to 150; higher better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	56	67	-	MD 3.6 higher (2.69 lower to 9.89 higher)	⊕OOO VERY LOW	CRITICAL
MH outcomes: Clinical Anxiety scale at post-treatment (Scale from 0 to 100; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	Serious[2]	none	56	67	-	MD 3.3 lower (8.55 lower to 1.95 higher)	⊕OOO VERY LOW	CRITICAL
MH outcomes: Clinical Anxiety scale (13 weeks follow-up) (Scale from 0 to 100; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	56	67	-	MD 7.8 lower (12.64 to 2.96 lower)	⊕OOO VERY LOW	CRITICAL

[1] *Valentine 2001 - high risk of selection bias, No blinding, unclear attrition bias, low selective outcome bias, low other risk of bias*

[2] *The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference*

(MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries .

TARGET vs SGT

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	TARGET	SGT	Relative (95% CI)	Absolute		
PTSD symptoms by CAPS scales (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	38	34	-	MD 0.5 lower (11.01 lower to 10.01 higher)	⊕⊕⊕O MODERATE	CRITICAL
Heartland forgiveness scale (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	23	9	-	MD 4.6 higher (6.73 lower to 15.93 higher)	⊕⊕OO LOW	CRITICAL

[1] *Ford 2013- low risk of selection bias, blinding of care administrators and outcome assessors but no blinding of participants, low attrition rate, low selective outcome bias, low other risk of bias* Ford - low risk of selection bias, blinding of care administrators and outcome assessors but no blinding of participants, low attrition rate, low selective outcome bias, low other risk of bias

[2] *The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries..*

Focused group therapy vs Wait-list control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Focused group therapy	Wait-list control	Relative (95% CI)	Absolute		
Symptom checklist-90-R: Global Severity Index (Scale from 0 to 90; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	4	5	-	MD 16.3 lower (26.23 to 6.37 lower)	⊕⊕OO LOW	CRITICAL
Symptom Checklist-90R: Positive Symptom Distress Index (Scale from 0 to 90; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	4	5	-	MD 13.9 lower (24.8 to 3 lower)	⊕OOO VERY LOW	CRITICAL
Symptom Checklist-90R: Positive Symptom Total (Scale from 0 to 90; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	4	5	-	MD 16.1 lower (26.67 to 5.53 lower)	⊕⊕OO LOW	CRITICAL

[1] Cole 2007 - high risks of selection bias, No blinding, Unclear attrition bias, low selective outcome bias and low other risk of bias
[2] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.

Group Therapy vs No contact control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Group Therapy	No contact control	Relative (95% CI)	Absolute		
IIP-32 scales (Scale from 0 to 128; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness		none	13	18	-	MD 10.1 lower (24.43 lower to 4.23 higher)		CRITICAL

[1]Bradley 2003 - unclear risks of selection bias, No blinding, Unclear attrition, High selective outcomes bias and low other risks of bias

N.3.5 ADHD

N.3.6 Methylphenidate vs Placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Methylphenidate (MPH)	Placebo	Relative (95% CI)	Absolute		
Conner Adult ADHD rating scale - Observer: Screening Version (CAARS-OSV) – post-treatment (52 weeks) (Scale from 0 to 90; lower better)												
2	randomised trials	serious[1,2]	very serious[3]	no serious indirectness	Serious[4]	none	42	42	-	MD 12.85 lower (22.5 to 3.20 lower)	⊕OOO VERY LOW	CRITICAL
Conner Adult ADHD rating scale - Observer: Screening Version (CAARS-OSV) - Follow-up (3 years) (Copy) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	15	5	-	MD 16.9 lower (24.5 to 9.3 lower)	⊕⊕OO LOW	CRITICAL
Number of participants with drug negative urine												
1	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	serious[5]	none	6/27 (22.2%)	4/27 (14.8%)	RR 1.5 (0.48 to 4.72)	74 more per 1000 (from 77 fewer to 551 more)	⊕OOO VERY LOW	CRITICAL

[1]Ginsberg 2012 - high risk of selection bias, No blinding, low risk of attrition, unclear selective outcome reporting and low risk of other bias
[2]Konstenius 2013- low risk of selection bias, Blinding of participants, care administrators and outcome detectors, unclear attrition bias and unclear selective outcome reporting, low risk of other bias

[3] Evidence was downgraded by one level due to serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of 50%-74.99%) and by two levels due to very serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of >75%)
[4] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries. If SMD was used, +0.5 and -0.5 on the SMD scale as MID boundaries.'
[5] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or two boundaries of the defined minimally important difference for the outcome (imprecision) respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.

N.3.7 Antisocial personality disorders

Tiagabine vs Placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Tiagabine	Placebo	Relative (95% CI)	Absolute		
Aggressive Response (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	6	6	-	MD 1.86 lower (2.7 to 1.02 lower)	⊕OOO VERY LOW	CRITICAL
Number of subjects with adverse effects												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	6/157* (3.8%)	6/65* (9.2%)	RR 0.41 (0.14 to 1.24)	54 fewer per 1000 (from 79 fewer to 22 more)	⊕OOO VERY LOW	CRITICAL

[1] Gowin 2012- Unclear risk of selection bias, blinding to participants and care person involved (+), low risk of attrition, unclear selective outcome reporting, low risk of other bias.
[2] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.
[3] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or two boundaries of the defined minimally important difference (MID) for the outcome (imprecision) respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.
**Denominator - total number of 'Yes' reports to the side-effects at least once*

N.3.8 Severe mental illness

IM Paliperidone vs Oral Antipsychotics for schizophrenia

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	IM Paliperidone	Oral antipsychotics	Relative (95% CI)	Absolute		
First-time treatment failure*												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	90/226 (39.8%)	117/218 (53.7%)	RR 0.74 (0.61 to 0.91)	140 fewer per 1000 (from 48 fewer to 209 fewer)	⊕OOO VERY LOW	CRITICAL
Incidence of prolactin-related side-effects												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	53/226 (23.5%)	9/219 (4.1%)	RR 5.71 (2.89 to 11.28)	194 more per 1000 (from 78 more to 422 more)	⊕⊕OO LOW	LIMITED IMPORTANCE

[1] *Alphs 2015a- Unclear risk of selection bias, no blinding, low risk of attrition bias, low risk of selective outcome bias, low risk of other bias*

[2] *The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or two boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.*

* First-time treatment failure - arrest/incarceration, psychiatric hospitalization, suicide, discontinuation of antipsychotic treatment due to inadequate efficacy, treatment supplementation with another antipsychotic due to inadequate efficacy, discontinuation of antipsychotic treatment due to safety or tolerability concerns, or an increase in the level of psychiatric services to prevent imminent psychiatric hospitalization

The Citizenship project for severe mental illness

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	The Citizenship Project	TAU	Relative (95% CI)	Absolute		
Overall quality of life (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	73	41	-	MD 0.68 higher (0 to 1.36 higher)	⊕OOO VERY LOW	CRITICAL
Number of all convictions (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	73	41	-	MD 0.05 higher (0.79 lower to 0.89 higher)	⊕⊕OO LOW	CRITICAL
Alcohol composite ratio (Scale from 0 to 9; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	Serious[2]	none	40	29	-	MD 0.29 lower (0.57 to 0.01 lower)	⊕OOO VERY LOW	IMPORTANT
Brief Psychiatric Rating Scale: Withdrawal symptoms (Scale from 1 to 7; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	Serious[2]	none	73	41	-	MD 0.28 higher (0.01 to 0.55 higher)	⊕OOO VERY LOW	IMPORTANT
Addition severity index: Drug index (Scale from 0 to 9; lower better)												
1	randomised	very	no serious	no serious	serious[2]	none	73	41	-	MD 0.04 lower (0.08 lower	⊕OOO	IMPORTANT

	trials	serious[1]	inconsistency	indirectness						to 0 higher)	VERY LOW	

[1] *Clayton 2013 - Unclear selection bias, No blinding, Unclear attrition, low risk of selective outcome reporting, low risk of other bias*
[2]*The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries .*

Individual Placement and Support vs Peer support for severe mental illness

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Individual Placement and Support (IPS)	Peer support	Relative (95% CI)	Absolute		
Competitive job placement												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	13/42 (31%)	3/43 (7%)	RR 4.44 (1.36 to 14.46)	240 more per 1000 (from 25 more to 939 more)	⊕⊕OO LOW	IMPORTANT
Number of hospitalizations (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	41	43	-	MD 0.5 higher (0.07 lower to 1.07 higher)	⊕⊕OO LOW	CRITICAL
Number of days being hospitalized (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	41	43	-	MD 5.51 higher (1.91 lower to 12.93 higher)	⊕OOO VERY LOW	CRITICAL

[1] *Bond 2015 - Appropriate randomization with concealed allocation, no blinding of participants and care administrators, ITT analysis, appropriate outcome report*
[2] *The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or two boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.*
[3] *The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries..*

N.3.9 Uncategorised mental health disorders

Parenting from inside vs wait-list control

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Parenting from the Inside (PFI)	Wait-list control	Relative (95% CI)	Absolute		
Parenting Stress Index (Scale from 27 to 135; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	60	76	-	MD 0.04 higher (0.17 lower to 0.25 higher)	⊕⊕OO LOW	CRITICAL
Brief Symptom Inventory (BSI): Total (Scale from 0 to 212; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	60	76	-	MD 0.2 higher (0.12 lower to 0.52 higher)	⊕OOO VERY LOW	CRITICAL
Parenting Alliance (Scale from 20 to 100; higher better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	60	76	-	MD 0.31 lower (6.23 lower to 5.61 higher)	⊕⊕OO LOW	IMPORTANT

[1] *Loper 2011 - Unclear selection bias; No blinding; Unclear attrition bias, low risk of selective outcomes, low risk of other bias*
[2] *The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.*

Music therapy vs standard care for anxiety and depression disorders

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Music therapy	Standard care	Relative (95% CI)	Absolute		
State and Trait Anxiety Inventory - State (Scale from 20 to 80; lower better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	93	91	-	MD 8.05 lower (10.74 to 5.36 lower)	⊕⊕⊕O MODERATE	CRITICAL
State and Trait Anxiety Inventory - Trait (Scale from 20 to 80; lower better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	93	91	-	MD 8.51 lower (10.91 to 6.11 lower)	⊕⊕⊕O MODERATE	CRITICAL
Brief Symptom Inventory (BSI): Total (Scale from 0 to 212; lower better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	93	91	-	MD 8.81 lower (11.82 to 5.8 lower)	⊕⊕⊕O MODERATE	CRITICAL
Rosenberg self-esteem inventory (Scale from 0 to 30; higher better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	93	91	-	MD 2.26 higher (0.98 to 3.54 higher)	⊕⊕⊕O MODERATE	CRITICAL
Texas social behaviour inventory (Scale from 0 to 128; higher better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	93	91	-	MD 7.54 higher (3.24 to 11.84 higher)	⊕⊕⊕O MODERATE	CRITICAL

[1] *Chen 2015 - Appropriate randomization with proper concealment; blinding of care administrators, but not participants; ITT analysis; appropriate outcome report*

Music therapy vs wait-list control for antisocial personality disorders

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Music therpy	Wait-list control	Relative (95% CI)	Absolute		
ASP-1: Self-management of psychiatric symptoms (Scale from 0 to 4; higher better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	8	5	-	MD 0.44 higher (0.03 lower to 0.91 higher)	⊕OOO VERY LOW	CRITICAL
ASP-4: self-management of assaultive symptoms (Scale from 0 to 4; higher better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	8	5	-	MD 0.11 lower (0.67 lower to 0.45 higher)	⊕OOO VERY LOW	CRITICAL
ASP-9: Interpersonal skills (Scale from 0 to 4; higher better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	8	5	-	MD 0.02 higher (0.06 lower to 0.1 higher)	⊕⊕OO LOW	CRITICAL
Social dysfunction and aggression scale (Scale from 0 to 44; lower better)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	8	5	-	MD 0.8 lower (2.73 lower to 1.13 higher)	⊕OOO VERY LOW	CRITICAL
forensic psychiatric profiles 40 (FP40): positive coping skills (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	8	5	-	MD 0.43 higher (0.12 to 0.74 higher)	⊕OOO VERY LOW	CRITICAL

[1] *Hakvoort 2013 - unclear randomisation and concealment; No blinding; available case analysis; appropriate outcome report*

[2] *The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.*

N.4 Interventions for acquired cognitive impairment

A narrative overview of indirectly relevant systematic reviews was performed for this question. The evidence was not from criminal justice system populations and was not subject to critical appraisal of quality.

N.5 Interventions for paraphilic disorders

N.5.1 Medroxyprogesterone + psychological intervention compared to psychological intervention only for paraphilic disorders

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Medroxyprogesterone + psych intervention	psych intervention only	Relative (95% CI)	Absolute (95% CI)		
Repetition of anomalous behaviour (follow up: range 15-52 weeks to; assessed with: self-report questionnaire and case notes)												
2	randomised trials	serious	serious [1]	not serious [2]	very serious[2]	none	2/25 (8.0%)	6/27 (22.2%)	**RR 0.58** (0.04 to 8.30)	**93 fewer per 1,000** (from 213 fewer to 1,000 more)	⊕◯◯◯ VERY LOW	CRITICAL
Reduced anomalous desires (follow up: 52; assessed with: self-report questionnaire)												
1	randomised trials	serious [3]	not serious	not serious	very serious [2]	none	5/10 (50.0%)	6/10 (60.0%)	**RR 0.83** (0.12 to 1.55)	**102 fewer per 1,000** (from 330 more to 528 fewer)	⊕◯◯◯ VERY LOW	CRITICAL
Dropout (follow up: 15; assessed with: number of participants who did not complete treatment)												
1	randomised trials	serious [3]	not serious	not serious	serious [2]	none	10/15 (66.7%)	5/17 (29.4%)	**RR 2.27** (1.00 to 5.14)	**374 more per 1,000** (from 0 fewer to 1,000 more)	⊕⊕◯◯ LOW	IMPORTANT

CI: Confidence interval; **RR:** Risk ratio

1. Downgraded for inconsistency
2. Confidence interval of the effect estimate includes appreciable benefit, harm and no effect
3. High risk of selection and performance bias

N.5.2 Medroxyprogesterone compared to imaginal desensitisation for paraphilic disorders

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Medroxyprogesterone	imaginal desensitisation	Relative (95% CI)	Absolute (95% CI)		
Reduced anomalous desire (follow up: 52; assessed with: self-report questionnaire)												
1	randomised trials	serious [1]	not serious	not serious	serious [2,3]	none	3/10 (30.0%)	6/10 (60.0%)	**RR 0.50** (0.17 to 1.46)	**300 fewer per 1,000** (from 276 more to 498 fewer)	⨁⨁◯◯ LOW	CRITICAL
Reduced anomalous behaviour (follow up: 52; assessed with: self-report questionnaire)												
1	randomised trials	serious [1]	not serious	not serious	serious [2,3]	none	1/10 (10.0%)	3/10 (30.0%)	**RR 0.33** (0.04 to 2.69)	**201 fewer per 1,000** (from 288 fewer to 507 more)	⨁⨁◯◯ LOW	CRITICAL

CI: Confidence interval; **RR:** Risk ratio

1. High risk of performance and attrition bias.
2. Optimal information size criterion not met (event rate less than 300)
3. Confidence interval for the effect estimate spans both MID thresholds (0.80 to 1.25).

N.5.3 Psychoeducational interventions, principally CBT-informed psychoeducation (including SOTP) versus treatment as usual, no treatment or waitlist control for paraphilic disorders.

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Psychoeducational intervention: principally CBT-informed psychoeducation (including SOTP)	Treatment as usual, no treatment or waitlist control	Relative (95% CI)	Absolute (95% CI)		
Cognitive distortions (Abel and Becker Cognition Scale, ABCS; Children and Sex Questionnaire) – RCT (Scale from 26 to 130; higher better)												
1	randomised trials	serious [1]	not serious	not serious	not serious	none	30	30	-	MD **13.43 lower** (20.05 lower to 6.81 lower)	⨁⨁⨁◯ MODERATE	IMPORTANT
Cognitive distortions (Abel and Becker Cognition Scale, ABCS; Children and Sex Questionnaire) - Controlled non-randomised studies												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Psychoeducational intervention: principally CBT-informed psychoeducation (including SOTP)	Treatment as usual, no treatment or waitlist control	Relative (95% CI)	Absolute (95% CI)		
1	observational studies	very serious [2]	not serious	not serious	serious [3]	none	38	19	-	MD **8.6 lower** (14.48 lower to 2.72 lower)	⨁◯◯◯ VERY LOW	IMPORTANT
Cognitive distortions (Abel and Becker Cognition Scale [ABCS]; number of participants who 'improved' [pre- to post-test score increased by at least 10 points]) - Controlled non-randomised studies												
1	observational studies	very serious [4]	not serious	not serious	very serious [5]	none	4/19 (21.1%)	2/5 (40.0%)	**RR 0.53** (0.13 to 2.10)	**188 fewer per 1,000** (from 348 fewer to 440 more)	⨁◯◯◯ VERY LOW	IMPORTANT
Cognitive distortions (Multiphasic Sex Inventory [MSI]: Cognitive distortions; number of participants who 'improved' [pre- to post-test score increased by at least 2 points]) - Controlled non-randomised studies												
1	observational studies	very serious [4]	not serious	not serious	very serious [5]	none	6/19 (31.6%)	1/5 (20.0%)	**RR 1.58** (0.24 to 10.28)	**116 more per 1,000** (from 152 fewer to 1,000 more)	⨁◯◯◯ VERY LOW	IMPORTANT
Emotional congruence with children (Children and Sex Questionnaire) - Controlled non-randomised studies												
1	observational studies	very serious [2]	not serious	not serious	serious [6]	none	38	19	-	MD **1.29 lower** (8.8 lower to 6.22 higher)	⨁◯◯◯ VERY LOW	IMPORTANT
Victim empathy distortions (Victim Empathy Distortions scale) - Controlled non-randomised studies												
1	observational studies	very serious [2]	not serious	not serious	serious [7]	none	38	19	-	MD **13 lower** (21.56 lower to 4.44 lower)	⨁◯◯◯ VERY LOW	IMPORTANT
Acceptance of accountability (Multiphasic Sex Inventory [MSI]: Justifications) - RCT												
1	randomised trials	serious [1]	not serious	not serious	serious [8]	none	30	30	-	MD **0.8 lower** (6.13 lower to 4.53 higher)	⨁⨁◯◯ LOW	IMPORTANT
Acceptance of accountability (Multiphasic Sex Inventory [MSI]: Justifications; number of participants who 'improved' [pre- to post-test score increased by at least 2 points]) - Controlled non-randomised studies												
1	observational studies	very serious [4]	not serious	not serious	very serious [5]	none	6/19 (31.6%)	2/5 (40.0%)	**RR 0.79** (0.22 to 2.79)	**84 fewer per 1,000** (from 312 fewer to 716 more)	⨁◯◯◯ VERY LOW	IMPORTANT
Levels of denial (Multiphasic Sex Inventory [MSI]: Lie; number of participants who 'improved' [pre- to post-test score increased by at least 2 points]) - Controlled non-randomised studies												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Psychoeducational intervention: principally CBT-informed psychoeducation (including SOTP)	Treatment as usual, no treatment or waitlist control	Relative (95% CI)	Absolute (95% CI)		
1	observational studies	very serious [4]	not serious	not serious	very serious [5]	none	3/19 (15.8%)	0/5 (0.0%)	**RR 2.10** (0.13 to 35.20)	**0 fewer per 1,000** (from 0 fewer to 0 fewer)	⨁◯◯◯ VERY LOW	IMPORTANT
Openness/honesty about sexual outlets (Multiphasic Sex Inventory [MSI]: Child molest; number of participants who 'improved' [pre- to post-test score increased by at least 2 points]) - Controlled non-randomised studies												
1	observational studies	very serious [4]	not serious	not serious	very serious [5]	none	7/19 (36.8%)	2/5 (40.0%)	**RR 0.92** (0.27 to 3.13)	**32 fewer per 1,000** (from 292 fewer to 852 more)	⨁◯◯◯ VERY LOW	IMPORTANT
Sexual anxiety (Multiphasic Sex Inventory [MSI]: Sexual inadequacies) - RCT												
1	randomised trials	serious [1]	not serious	not serious	serious [9]	none	30	30	-	MD **6.2 lower** (13.43 lower to 1.06 higher)	⨁⨁◯◯ LOW	IMPORTANT
Anxiety (Social Anxiety and Distress Scale, SADS) – RCT (Scale from 0 to 28; lower better)												
2	randomised trials	serious [1,10]	not serious	serious [11]	not serious	none	38	37	-	MD **2.19 lower** (7.31 lower to 2.92 higher)	⨁⨁◯◯ LOW	CRITICAL
Rearrest (CJS database; controlled non-randomised studies; longest follow-up available) - 2-year follow-up												
2	randomised trials	very serious [12,13]	not serious	not serious	very serious [5]	none	38/197 (19.3%)	72/367 (19.6%)	**RR 1.00** (0.63 to 1.59)	**0 fewer per 1,000** (from 73 fewer to 116 more)	⨁◯◯◯ VERY LOW	CRITICAL
Rearrest (CJS database; controlled non-randomised studies; longest follow-up available) - 3-year follow-up												
2	observational studies	very serious [14,15]	not serious	serious [16]	serious [5]	none	436/1317 (33.1%)	1000/2118 (47.2%)	**RR 0.78** (0.71 to 0.86)	**104 fewer per 1,000** (from 66 fewer to 137 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sex offence rearrest (CJS database; controlled non-randomised studies; longest follow-up available) - 2-year follow-up												
2	observational studies	very serious [12,13]	not serious	not serious	very serious [5]	none	17/197 (8.6%)	26/367 (7.1%)	**RR 1.03** (0.58 to 1.84)	**2 more per 1,000** (from 30 fewer to 60 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sex offence rearrest (CJS database; controlled non-randomised studies; longest follow-up available) - 3-year follow-up												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Psychoeducational intervention: principally CBT-informed psychoeducation (including SOTP)	Treatment as usual, no treatment or waitlist control	Relative (95% CI)	Absolute (95% CI)		
2	observational studies	very serious [14,15]	not serious	serious	serious [5]	none	105/1317 (8.0%)	199/2118 (9.4%)	**RR 0.80** (0.57 to 1.12)	**19 fewer per 1,000** (from 11 more to 40 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Violent rearrest (CJS database; controlled non-randomised studies; longest follow-up available) - 2-year follow-up												
1	observational studies	very serious [12]	not serious	not serious	very serious [5]	none	1/119 (0.8%)	5/159 (3.1%)	**RR 0.27** (0.03 to 2.26)	**23 fewer per 1,000** (from 31 fewer to 40 more)	⨁◯◯◯ VERY LOW	CRITICAL
Violent rearrest (CJS database; controlled non-randomised studies; longest follow-up available) - 3-year follow-up												
2	observational studies	very serious [14,15]	not serious	serious [16]	serious [5]	none	197/1317 (15.0%)	485/2118 (22.9%)	**RR 0.71** (0.60 to 0.83)	**66 fewer per 1,000** (from 39 fewer to 92 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Criminal charges (CJS database; controlled non-randomised studies; longest follow-up available) - 2-year follow-up												
1	observational studies	very serious [17]	not serious	serious [16]	very serious [5]	none	2/54 (3.7%)	1/14 (7.1%)	**RR 0.52** (0.05 to 5.32)	**34 fewer per 1,000** (from 68 fewer to 309 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sex offence charges (CJS database; controlled non-randomised studies; longest follow-up available) - 2-year follow-up												
1	observational studies	very serious [17]	not serious	serious [16]	very serious [5]	none	0/54 (0.0%)	1/14 (7.1%)	**RR 0.09** (0.00 to 2.12)	**65 fewer per 1,000** (from -- to 80 more)	⨁◯◯◯ VERY LOW	CRITICAL
Reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 2-year follow-up												
3	observational studies	very serious [13,18,19]	very serious [20]	not serious	very serious [5]	none	37/243 (15.2%)	247/493 (50.1%)	**RR 0.54** (0.16 to 1.82)	**230 fewer per 1,000** (from 411 more to 421 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 3-year follow-up												
1	observational studies	very serious [21]	not serious	not serious	serious [5]	none	4/94 (4.3%)	11/86 (12.8%)	**RR 0.33** (0.11 to 1.01)	**86 fewer per 1,000** (from 1 more to 114 fewer)	⨁◯◯◯ VERY LOW	CRITICAL

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Psychoeducational intervention: principally CBT-informed psychoeducation (including SOTP)	Treatment as usual, no treatment or waitlist control	Relative (95% CI)	Absolute (95% CI)		
Reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 4-year follow-up												
1	observational studies	very serious [22]	not serious	serious [16]	not serious	none	3/49 (6.1%)	23/74 (31.1%)	**RR 0.20** (0.06 to 0.62)	**249 fewer per 1,000** (from 118 fewer to 292 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 5-year follow-up												
3	observational studies	very serious [23,24,25]	serious [26]	serious [16]	not serious	none	81/549 (14.8%)	116/484 (24.0%)	**RR 0.53** (0.30 to 0.92)	**113 fewer per 1,000** (from 19 fewer to 168 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 7-year follow-up												
1	observational studies	very serious [27]	not serious	serious [16]	not serious	none	199/403 (49.4%)	160/321 (49.8%)	**RR 0.99** (0.85 to 1.15)	**5 fewer per 1,000** (from 75 fewer to 75 more)	⨁◯◯◯ VERY LOW	CRITICAL
Reconviction at 2-year follow-up (risk of reconviction sub-analyses) - Low risk												
1	randomised trials	very serious [18]	not serious	serious [16]	not serious	none	15/263 (5.7%)	65/969 (6.7%)	**RR 0.85** (0.49 to 1.47)	**10 fewer per 1,000** (from 32 more to 34 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Reconviction at 2-year follow-up (risk of reconviction sub-analyses) - Medium-low risk												
1	randomised trials	very serious [18]	not serious	serious [16]	not serious	none	30/225 (13.3%)	166/655 (25.3%)	**RR 0.53** (0.37 to 0.75)	**119 fewer per 1,000** (from 63 fewer to 160 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Reconviction at 2-year follow-up (risk of reconviction sub-analyses) - Medium-high risk												
1	observational studies	very serious [18]	not serious	serious [16]	not serious	none	23/109 (21.1%)	229/229 (100.0%)	**RR 0.21** (0.15 to 0.31)	**790 fewer per 1,000** (from 690 fewer to 850 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Reconviction at 2-year follow-up (risk of reconviction sub-analyses) - High risk												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Psychoeducational intervention: principally CBT-informed psychoeducation (including SOTP)	Treatment as usual, no treatment or waitlist control	Relative (95% CI)	Absolute (95% CI)		
1	observational studies	very serious [18]	not serious	serious [16]	very serious [5]	none	18/50 (36.0%)	22/57 (38.6%)	**RR 0.93** (0.57 to 1.53)	**27 fewer per 1,000** (from 166 fewer to 205 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sexual reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 2-year follow-up												
2	observational studies	very serious [18,19]	not serious	not serious	very serious [5]	none	20/703 (2.8%)	55/1966 (2.8%)	**RR 0.99** (0.59 to 1.68)	**0 fewer per 1,000** (from 11 fewer to 19 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sexual reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 3-year follow-up												
1	observational studies	very serious [21]	not serious	not serious	very serious [5]	none	1/94 (1.1%)	4/86 (4.7%)	**RR 0.23** (0.03 to 2.01)	**36 fewer per 1,000** (from 45 fewer to 47 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sexual reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 4-year follow-up												
2	observational studies	very serious [22]	not serious	not serious	very serious [5]	none	5/93 (5.4%)	17/118 (14.4%)	**RR 0.42** (0.13 to 1.34)	**84 fewer per 1,000** (from 49 more to 125 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sexual reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 5-year follow-up												
3	observational studies	very serious [24,25,28]	not serious	serious [16]	serious [5]	none	23/276 (8.3%)	48/241 (19.9%)	**RR 0.37** (0.16 to 0.83)	**125 fewer per 1,000** (from 34 fewer to 167 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sexual reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 7-year follow-up												
1	observational studies	very serious [27]	not serious	serious [16]	very serious [5]	none	61/403 (15.1%)	46/321 (14.3%)	**RR 1.06** (0.74 to 1.50)	**9 more per 1,000** (from 37 fewer to 72 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sexual reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 11-year follow-up												
1	observational studies	serious [29]	not serious	serious [16]	serious [5]	none	66/616 (10.7%)	21/104 (20.2%)	**RR 0.53** (0.34 to 0.83)	**95 fewer per 1,000** (from 34 fewer to 133 fewer)	⨁◯◯◯ VERY LOW	CRITICAL

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Psychoeducational intervention: principally CBT-informed psychoeducation (including SOTP)	Treatment as usual, no treatment or waitlist control	Relative (95% CI)	Absolute (95% CI)		
Sexual reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - Length of follow-up not reported												
1	observational studies	very serious [30]	not serious	not serious	very serious [5]	none	12/95 (12.6%)	17/145 (11.7%)	**RR 1.08** (0.54 to 2.15)	**9 more per 1,000** (from 54 fewer to 135 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sexual reconviction at 2 or 11-year follow-up (risk of reconviction sub-analyses) - Low risk												
2	observational studies	very serious [18,29]	not serious	serious [16]	very serious [5]	none	12/511 (2.3%)	14/994 (1.4%)	**RR 0.68** (0.26 to 1.78)	**5 fewer per 1,000** (from 10 fewer to 11 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sexual reconviction at 2 or 11-year follow-up (risk of reconviction sub-analyses) - Medium-low risk												
2	observational studies	very serious [18,29]	not serious	serious [16]	very serious [5]	none	25/393 (6.4%)	25/683 (3.7%)	**RR 0.71** (0.23 to 2.16)	**11 fewer per 1,000** (from 28 fewer to 42 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sexual reconviction at 2 or 11-year follow-up (risk of reconviction sub-analyses) - Medium-high risk												
2	observational studies	very serious [18,29]	not serious	serious [16]	very serious [5]	none	27/253 (10.7%)	19/260 (7.3%)	**RR 0.67** (0.36 to 1.28)	**24 fewer per 1,000** (from 20 more to 47 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sexual reconviction at 2 or 11-year follow-up (risk of reconviction sub-analyses) - High risk												
2	observational studies	very serious [18,29]	serious [26]	serious [16]	very serious [5]	none	19/106 (17.9%)	17/77 (22.1%)	**RR 0.68** (0.26 to 1.76)	**71 fewer per 1,000** (from 163 fewer to 168 more)	⨁◯◯◯ VERY LOW	CRITICAL
Violent reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 3-year follow-up												
1	observational studies	very serious [21]	not serious	not serious	serious [5]	none	1/94 (1.1%)	7/86 (8.1%)	**RR 0.13** (0.02 to 1.04)	**71 fewer per 1,000** (from 3 more to 80 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Violent reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 5-year follow-up												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Psychoeducational intervention: principally CBT-informed psychoeducation (including SOTP)**	**Treatment as usual, no treatment or waitlist control**	**Relative (95% CI)**	**Absolute (95% CI)**		
2	observational studies	very serious [24,25]	not serious	serious [16]	not serious	none	16/176 (9.1%)	32/141 (22.7%)	**RR 0.43** (0.25 to 0.74)	**129 fewer per 1,000** (from 59 fewer to 170 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Violent reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 7-year follow-up												
1	observational studies	very serious [27]	not serious	serious [16]	serious [5]	none	136/403 (33.7%)	99/321 (30.8%)	**RR 1.09** (0.88 to 1.35)	**28 more per 1,000** (from 37 fewer to 108 more)	⨁◯◯◯ VERY LOW	CRITICAL
Violent reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - Length of follow-up not reported												
1	observational studies	very serious [30]	not serious	not serious	very serious [5]	none	11/95 (11.6%)	24/145 (16.6%)	**RR 0.70** (0.36 to 1.36)	**50 fewer per 1,000** (from 60 more to 106 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Violent reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 11-year follow-up												
1	observational studies	very serious [29]	not serious	serious [16]	not serious	none	163/616 (26.5%)	46/104 (44.2%)	**RR 0.60** (0.46 to 0.77)	**177 fewer per 1,000** (from 102 fewer to 239 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Violent reconviction at 11-year follow-up (risk of reconviction sub-analyses) - Low risk												
1	observational studies	very serious [29]	not serious	serious [16]	serious [5]	none	28/248 (11.3%)	6/25 (24.0%)	**RR 0.47** (0.22 to 1.03)	**127 fewer per 1,000** (from 7 more to 187 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Violent reconviction at 11-year follow-up (risk of reconviction sub-analyses) - Medium-low risk												
1	observational studies	very serious [29]	not serious	serious [16]	very serious [5]	none	56/168 (33.3%)	11/28 (39.3%)	**RR 0.85** (0.51 to 1.41)	**59 fewer per 1,000** (from 161 more to 193 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Violent reconviction at 11-year follow-up (risk of reconviction sub-analyses) - Medium-high risk												
1	observational studies	very serious [29]	not serious	serious [16]	serious [5]	none	53/144 (36.8%)	16/31 (51.6%)	**RR 0.71** (0.48 to 1.07)	**150 fewer per 1,000** (from 36 more to 268 fewer)	⨁◯◯◯ VERY LOW	CRITICAL

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Psychoeducational intervention: principally CBT-informed psychoeducation (including SOTP)	Treatment as usual, no treatment or waitlist control	Relative (95% CI)	Absolute (95% CI)		
Violent reconviction at 11-year follow-up (risk of reconviction sub-analyses) - High risk												
1	observational studies	very serious [15]	not serious	serious [16]	not serious	none	26/56 (46.4%)	13/20 (65.0%)	**RR 0.71** (0.47 to 1.10)	**189 fewer per 1,000** (from 65 more to 345 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Incarceration (CJS database; controlled non-randomised studies; longest follow-up available) - 3-year follow-up												
1	observational studies	very serious [15]	not serious	serious [16]	not serious	none	35/297 (11.8%)	228/1098 (20.8%)	**RR 0.57** (0.41 to 0.79)	**89 fewer per 1,000** (from 44 fewer to 123 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Incarceration for sexual offence (CJS database; controlled non-randomised studies; longest follow-up available) - 3-year follow-up												
1	observational studies	very serious [15]	not serious	serious [16]	very serious [5]	none	9/297 (3.0%)	42/1098 (3.8%)	**RR 0.79** (0.39 to 1.61)	**8 fewer per 1,000** (from 23 fewer to 23 more)	⨁◯◯◯ VERY LOW	CRITICAL
Incarceration for violent offence (CJS database; controlled non-randomised studies; longest follow-up available) - 3-year follow-up												
1	observational studies	very serious [15]	not serious	serious [16]	serious [5]	none	9/297 (3.0%)	74/1098 (6.7%)	**RR 0.45** (0.23 to 0.89)	**37 fewer per 1,000** (from 7 fewer to 52 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Revocation, breaches of the Sex Offender Register or probation violation (CJS database; controlled non-randomised studies; longest follow-up available) - 2-year follow-up												
2	observational studies	very serious [13,17]	very serious [20]	not serious	very serious [5]	none	31/132 (23.5%)	31/222 (14.0%)	**RR 0.88** (0.12 to 6.74)	**17 fewer per 1,000** (from 123 fewer to 802 more)	⨁◯◯◯ VERY LOW	CRITICAL
Revocation, breaches of the Sex Offender Register or probation violation (CJS database; controlled non-randomised studies; longest follow-up available) - 5-year follow-up												
2	observational studies	very serious [15,24]	not serious	serious [16]	not serious	none	66/231 (28.6%)	643/1361 (47.2%)	**RR 0.64** (0.51 to 0.80)	**170 fewer per 1,000** (from 94 fewer to 231 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Revocation, breaches of the Sex Offender Register or probation violation (CJS database; controlled non-randomised studies; longest follow-up available) - Length of follow-up not reported												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Psychoeducational intervention: principally CBT-informed psychoeducation (including SOTP)	Treatment as usual, no treatment or waitlist control	Relative (95% CI)	Absolute (95% CI)		
1	observational studies	very serious [30]	not serious	not serious	not serious	none	7/95 (7.4%)	35/145 (24.1%)	**RR 0.31** (0.14 to 0.66)	**167 fewer per 1,000** (from 82 fewer to 208 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Global treatment response: Any change (positively rated for improvements in denial, positive changes on scales and attendance at therapy, and negatively rated for reconvictions and breach of probation order or parole licence) - Controlled non-randomised studies												
1	observational studies	very serious [4]	not serious	not serious	very serious [5]	none	15/20 (75.0%)	3/5 (60.0%)	**RR 1.25** (0.59 to 2.67)	**150 more per 1,000** (from 246 fewer to 1,000 more)	⨁◯◯◯ VERY LOW	IMPORTANT
Global treatment response: Major change (positively rated for improvements in denial, positive changes on scales and attendance at therapy, and negatively rated for reconvictions and breach of probation order or parole licence) - Controlled non-randomised studies												
1	observational studies	very serious [4]	not serious	not serious	very serious [5]	none	4/20 (20.0%)	0/5 (0.0%)	**RR 2.57** (0.16 to 41.34)	**0 fewer per 1,000** (from 0 fewer to 0 fewer)	⨁◯◯◯ VERY LOW	IMPORTANT
Sub-analysis by country: Reconviction(Any)												
9	observational studies	very serious [13,18,19,21,22,23,24,25,27]	serious [26]	serious [16]	very serious [5]	none	324/1338 (24.2%)	557/1458 (38.2%)	**RR 0.49** (0.30 to 0.82)	**195 fewer per 1,000** (from 69 fewer to 267 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Reconviction(Any) - UK												
1	observational studies	very serious [18]	not serious	serious [16]	not serious	none	23/109 (21.1%)	229/229 (100.0%)	**RR 0.21** (0.15 to 0.31)	**790 fewer per 1,000** (from 690 fewer to 850 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Reconviction(Any) - Netherlands												
1	observational studies	very serious [19]	not serious	not serious	very serious [5]	none	12/56 (21.4%)	14/56 (25.0%)	**RR 0.86** (0.44 to 1.69)	**35 fewer per 1,000** (from 140 fewer to 173 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Reconviction(Any) - Spain												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Psychoeducational intervention: principally CBT-informed psychoeducation (including SOTP)	Treatment as usual, no treatment or waitlist control	Relative (95% CI)	Absolute (95% CI)		
1	observational studies	very serious [22]	not serious	serious [16]	not serious	none	3/49 (6.1%)	23/74 (31.1%)	**RR 0.20** (0.06 to 0.62)	**249 fewer per 1,000** (from 118 fewer to 292 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Reconviction(Any) - US												
4	observational studies	very serious [13,23,24,25]	serious [26]	serious [16]	serious [5]	none	83/627 (13.2%)	120/692 (17.3%)	**RR 0.57** (0.34 to 0.96)	**75 fewer per 1,000** (from 7 fewer to 114 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Reconviction(Any) - Canada												
2	observational studies	very serious [21,27]	serious [26]	serious [16]	very serious [5]	none	203/497 (40.8%)	171/407 (42.0%)	**RR 0.66** (0.23 to 1.88)	**143 fewer per 1,000** (from 324 fewer to 370 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Sexual reconviction												
11	observational studies	very serious [18,19,21,22,24,25,27,28,29,30,31]	serious [26]	serious [16]	serious [5]	none	188/2280 (8.2%)	208/2981 (7.0%)	**RR 0.66** (0.47 to 0.93)	**24 fewer per 1,000** (from 5 fewer to 37 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Sexual reconviction - UK												
3	observational studies	very serious [18,30,31]	not serious	not serious	very serious [5]	none	32/786 (4.1%)	75/2099 (3.6%)	**RR 0.96** (0.64 to 1.44)	**1 fewer per 1,000** (from 13 fewer to 16 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Sexual reconviction - US												
3	observational studies	very serious [24,25,28]	not serious	serious [16]	serious [5]	none	23/276 (8.3%)	48/241 (19.9%)	**RR 0.37** (0.16 to 0.83)	**125 fewer per 1,000** (from 34 fewer to 167 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Sexual reconviction - Netherlands												
1	observational studies	very serious [19]	not serious	not serious	very serious [5]	none	3/56 (5.4%)	1/56 (1.8%)	**RR 3.00** (0.32 to 27.97)	**36 more per 1,000** (from 12 fewer to 482 more)	⨁◯◯◯ VERY LOW	CRITICAL

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Psychoeducational intervention: principally CBT-informed psychoeducation (including SOTP)	Treatment as usual, no treatment or waitlist control	Relative (95% CI)	Absolute (95% CI)		
Sub-analysis by country: Sexual reconviction - Spain												
1	observational studies	very serious [22]	not serious	serious [16]	serious [5]	none	2/49 (4.1%)	13/74 (17.6%)	**RR 0.23** (0.05 to 0.98)	**135 fewer per 1,000** (from 4 fewer to 167 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Sexual reconviction - Canada												
3	observational studies	very serious [21,27,29]	serious [26]	not serious	serious [16]	none	128/1113 (11.5%)	71/511 (13.9%)	**RR 0.69** (0.36 to 1.33)	**43 fewer per 1,000** (from 46 more to 89 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Violent reconviction												
6	observational studies	very serious [30]	not serious	not serious	serious [5]	none	327/1384 (23.6%)	208/797 (26.1%)	**RR 0.62** (0.40 to 0.96)	**99 fewer per 1,000** (from 10 fewer to 157 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Violent reconviction - UK												
1	observational studies	very serious [24,25]	not serious	not serious	very serious [5]	none	11/95 (11.6%)	24/145 (16.6%)	**RR 0.70** (0.36 to 1.36)	**50 fewer per 1,000** (from 60 more to 106 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Violent reconviction - US												
2	observational studies	very serious [24,25]	not serious	serious [16]	not serious	none	16/176 (9.1%)	32/141 (22.7%)	**RR 0.43** (0.25 to 0.74)	**129 fewer per 1,000** (from 59 fewer to 170 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Violent reconviction - Canada												
3	observational studies	very serious [21,27,29]	very serious [20]	serious [16]	very serious [5]	none	300/1113 (27.0%)	152/511 (29.7%)	**RR 0.71** (0.39 to 1.31)	**86 fewer per 1,000** (from 92 more to 181 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Revocation												
5	observational studies	very serious [13,15,17,24,30]	serious [20]	serious [16]	serious [5]	none	104/458 (22.7%)	709/1728 (41.0%)	**RR 0.66** (0.35 to 1.23)	**140 fewer per 1,000** (from 94 more to 267 fewer)	⨁◯◯◯ VERY LOW	CRITICAL

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Psychoeducational intervention: principally CBT-informed psychoeducation (including SOTP)	Treatment as usual, no treatment or waitlist control	Relative (95% CI)	Absolute (95% CI)		
Sub-analysis by country: Revocation - UK												
1	observational studies	very serious [30]	not serious	not serious	not serious	none	7/95 (7.4%)	35/145 (24.1%)	**RR 0.31** (0.14 to 0.66)	**167 fewer per 1,000** (from 82 fewer to 208 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Revocation - US												
4	observational studies	very serious [13,15,17,24]	very serious [20]	serious [16]	very serious [5]	none	97/363 (26.7%)	674/1583 (42.6%)	**RR 0.77** (0.39 to 1.55)	**98 fewer per 1,000** (from 234 more to 260 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by setting: Any reconviction												
9	observational studies	very serious [13,18,19,21,22,24,25,27,32]	very serious [20]	serious [16]	serious [5]	none	349/1054 (33.1%)	582/1204 (48.3%)	**RR 0.52** (0.33 to 0.80)	**232 fewer per 1,000** (from 97 fewer to 324 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by setting: Any reconviction - Inpatient												
1	observational studies	very serious [32]	very serious [20]	serious [16]	serious [5]	none	55/89 (61.8%)	66/89 (74.2%)	**RR 0.83** (0.68 to 1.02)	**126 fewer per 1,000** (from 15 more to 237 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by setting: Any reconviction - Prison												
4	observational studies	very serious [18,21,22,25]	serious [26]	serious [16]	serious [5]	none	74/357 (20.7%)	315/479 (65.8%)	**RR 0.33** (0.13 to 0.81)	**441 fewer per 1,000** (from 125 fewer to 572 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by setting: Any reconviction - Community												
4	observational studies	very serious [13,19,24,27]	serious [26]	serious [16]	very serious [5]	none	220/608 (36.2%)	201/636 (31.6%)	**RR 0.67** (0.32 to 1.40)	**104 fewer per 1,000** (from 126 more to 215 fewer)	⨁◯◯◯ VERY LOW	CRITICAL

CI: Confidence interval; **MD:** Mean difference; **RR:** Risk ratio; SOTP, sex offender treatment programme

1. Anderson-Varney 1991 - unclear risk of selection bias; no blinding; low risk of attrition bias; low risk of selective outcome bias; low risk of other bias
2. O'Reilly 2010 - Controlled Non-RCT; high risk of selection bias (significant difference in age between groups); No blinding; low risk of attrition bias; low risk of selective outcome bias; low risk of other bias
3. The MID calculated from SD of control was +/-6.26.
4. Craissati 1997 - Controlled Non-RCT; at baseline, men in the group condition was more likely to have abused children within the family; Increased loss of data in individual treatment programme (68%) than group treatment (38%); no selective outcome bias, no other risk of bias
5. The 95% CI considered for imprecision was 0.8 to 1.25.
6. The MID calculated from SD of control was +/-6.39.
7. The MID calculated from SD of control was +/-9.11.
8. The MID calculated from SD of control was +/-5.41.
9. The MID calculated from SD of control was +/-6.01.
10. Hopkins 1991 - Unclear selection bias; No blinding; low risk of attrition bias; low risk of selective outcome bias; low risk of other bias.
11. Hopkins 1991 - Participants involved roughly equal numbers of incarcerated paedophile and rapists.
12. Song 1995 - Controlled Non-RCT; significant group differences at baseline in race, prior sex offences and type of offence; no blinding; unclear risk of attrition bias; low risk of selective outcome bias; low risk of other bias;
13. Stalans 2001 - Controlled Non-RCT; significant group differences at baseline in current offence and on prior criminal history; no blinding; unclear risk of attrition bias; low risk of selective outcome bias; low risk of other bias
14. Duwe 2009 - Controlled Non-RCT; high risk of selection bias; no blinding; low risk of attrition bias; high risk of selective outcome bias; low risk of other bias;
15. Lowden 2003 - Controlled Non-RCT; significant group differences at baseline on age, marital status and criminal history; high risk of selection bias; no blinding; low risk of attrition bias; high risk of selective outcome bias; low risk of other bias
16. Unclear proportion of paraphilia participants
17. McGuire 2000 - Controlled Non-RCT; no blinding; unclear risk of attrition bias; low risk of selective outcome bias; low risk of other bias
18. Friendship 2003 - Controlled Non-RCT; confounders controlled in analysis; no blinding; unclear risk of attrition bias; high risk of selective outcome bias; low risk of other bias
19. Ruddijs 2000 - Controlled Non-RCT; no blinding; unclear risk of attrition bias; low risk of selective outcome bias; low risk of other bias
20. I2>80%
21. Marshall 2008 - Controlled Non-RCT; no blinding; unclear risk of attrition bias; low risk of selective outcome bias; low risk of other bias
22. Illescas 2008 - Controlled Non-RCT; no blinding; unclear risk of attrition bias; low risk of selective outcome bias; low risk of other bias
23. Aytes 2001 - Controlled Non-RCT; significant group differences at baseline in prior incarceration and prior felony conviction; no blinding; unclear risk of attrition bias; low risk of selective outcome bias; low risk of other bias
24. McGrath 1998 - Controlled Non-RCT; significant group differences at baseline in prior convictions; average time incarcerated and type of sexual offence committed; no blinding; low risk of attrition bias; low risk of selective outcome bias; low risk of other bias
25. McGrath 2003 - Controlled Non-RCT; significant group differences at baseline on prior convictions and time at risk in the community; no blinding; low risk of attrition bias; low risk of selective outcome bias; low risk of other bias
26. 50%<I2<80%
27. Hanson 2004 - Controlled Non-RCT; higher proportion of prior sexual offences in intervention group compared with control group; no blinding; unclear risk of attrition bias; low risk of selective outcome bias; low risk of other bias
28. Turner 2000 - McGrath 1998 - Controlled Non-RCT; no blinding; unclear risk of attrition bias; low risk of selective outcome bias; low risk of other bias
29. Olver 2013a - Controlled Non-RCT; low risk of selection bias (profounders properly controlled); no blinding; unclear risk of attrition bias; low risk of selective outcome bias; low risk of other bias
30. Craissati 2009 - Controlled Non-RCT; high risk of selection bias; no blinding; unclear risk of attrition bias; low risk of selective outcome bias; low risk of other bias
31. Procter 1996 - Controlled Non-RCT; no blinding; low risk of attrition bias; low risk of selective outcome bias; low risk of other bias
32. Looman 2000 - Controlled Non-RCT; no blinding; unclear risk of attrition bias; low risk of selective outcome bias; unclear risk of other bias

N.5.4 Good Lives Model (GLM) versus Relapse Prevention (RP) for paraphilic disorders

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Good Lives Model (GLM)	Relapse Prevention (RP)	Relative (95% CI)	Absolute (95% CI)		
Cognitive distortions (Children and Sex Questionnaire) (Scale from 0 to 75; lower better) - Controlled non-randomised studies												
1	observational studies	very serious [1]	not serious	not serious	serious [2]	none	207	294	-	MD **7.15 lower** (9.06 lower to 5.25 lower)	⊕◯◯◯ VERY LOW	IMPORTANT
Emotional congruence with children (Children and Sex Questionnaire) (Scale from 0 to 75; lower better) - Controlled non-randomised studies												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Good Lives Model (GLM)	Relapse Prevention (RP)	Relative (95% CI)	Absolute (95% CI)		
1	observational studies	very serious [1]	not serious	not serious	serious [3]	none	207	294	-	MD **7.72 lower** (10.13 lower to 5.3 lower)	⨁◯◯◯ VERY LOW	IMPORTANT
Victim empathy distortions (Victim Empathy Distortions scale) (Scale from 0 to 120; lower better) - Controlled non-randomised studies												
1	observational studies	very serious [1]	not serious	not serious	not serious [4]	none	207	294	-	MD **0.44 higher** (2.56 lower to 3.44 higher)	⨁◯◯◯ VERY LOW	IMPORTANT
Treatment response for pro-offending attitudes (using algorithm based on pre-post change and comparison with scores of a non-offender on Beliefs about Children Scale [cognitive distortions and emotional congruence with children subscales] and Victim Empathy Scale) - Controlled non-randomised studies												
1	randomised trials	very serious [5]	not serious	not serious	not serious	none	46/67 (68.7%)	366/520 (70.4%)	**RR 0.98** (0.82 to 1.16)	**14 fewer per 1,000** (from 113 more to 127 fewer)	⨁⨁◯◯ LOW	CRITICAL
Drop-out (any cause) - Controlled non-randomised studies												
1	observational studies	very serious [5]	not serious	not serious	very serious [6]	none	2/87 (2.3%)	2/182 (1.1%)	**RR 2.09** (0.30 to 14.60)	**12 more per 1,000** (from 8 fewer to 149 more)	⨁◯◯◯ VERY LOW	IMPORTANT

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

1. Barnett 2014 - Controlled Non-RCT; no blinding; data on drop-out was not available for some outcomes; low risk of other bias.
2. The MID calculated from SD of control was +/-6.79.
3. The MID calculated from SD of control was +/-7.95.
4. The MID calculated from SD of control was +/-8.48.
5. Harkins 2012 - Controlled Non-RCT; No blinding; data for individual scales were not reported; low other risk of bias.
6. The 95% CI considered for imprecision was 0.8 to 1.25.

N.5.5 Reintegration programmes versus treatment as usual for paraphilic disorders

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Reintegration programme (COSA)	Treatment as usual	Relative (95% CI)	Absolute (95% CI)		

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Reintegration programme (COSA)	Treatment as usual	Relative (95% CI)	Absolute (95% CI)		
Rearrest at 2-year follow-up (CJS database) - RCT												
1	randomised trials	very serious [1]	not serious	serious [2]	serious [3]	none	12/31 (38.7%)	20/31 (64.5%)	**RR 0.60** (0.36 to 1.00)	**258 fewer per 1,000** (from 0 fewer to 413 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sex offence rearrest at 2-year follow-up (CJS database) - RCT												
1	randomised trials	very serious [1]	not serious	serious [2]	very serious [3]	none	0/31 (0.0%)	1/31 (3.2%)	**RR 0.33** (0.01 to 7.88)	**22 fewer per 1,000** (from 32 fewer to 222 more)	⨁◯◯◯ VERY LOW	CRITICAL
Reconviction at 2- to 4-year follow-up (CJS database) - RCT (2-year follow-up)												
1	randomised trials	very serious [1]	not serious	serious [2]	serious [3]	none	8/31 (25.8%)	14/31 (45.2%)	**RR 0.57** (0.28 to 1.16)	**194 fewer per 1,000** (from 72 more to 325 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Reconviction at 2- to 4-year follow-up (CJS database) - Controlled non-randomised studies (3- or 4-year follow-up)												
3	observational studies	very serious [4]	not serious	serious [5]	serious [3]	none	29/175 (16.6%)	57/175 (32.6%)	**RR 0.52** (0.33 to 0.81)	**156 fewer per 1,000** (from 62 fewer to 218 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sex offence reconviction at 3- or 4-year follow-up (CJS database) - Controlled non-randomised studies												
3	observational studies	very serious [4]	not serious	serious [5]	serious [3]	none	8/175 (4.6%)	21/175 (12.0%)	**RR 0.41** (0.18 to 0.94)	**71 fewer per 1,000** (from 7 fewer to 98 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sex offence reconviction (contact) at 4-year follow-up (CJS database) - Controlled non-randomised studies												
1	observational studies	very serious [6]	not serious	not serious	very serious [3]	none	0/71 (0.0%)	3/71 (4.2%)	**RR 0.14** (0.01 to 2.72)	**36 fewer per 1,000** (from 42 fewer to 73 more)	⨁◯◯◯ VERY LOW	CRITICAL
Violent reconviction at 3- or 4-year follow-up (CJS database) - Controlled non-randomised studies												
3	observational studies	very serious [4]	not serious	serious [5]	not serious	none	13/175 (7.4%)	43/175 (24.6%)	**RR 0.34** (0.19 to 0.61)	**162 fewer per 1,000** (from 96 fewer to 199 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Resentence at 2-year follow-up (CJS database) - RCT												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Reintegration programme (COSA)	Treatment as usual	Relative (95% CI)	Absolute (95% CI)		
1	randomised trials	very serious [1]	not serious	serious [2]	very serious [3]	none	3/31 (9.7%)	8/31 (25.8%)	**RR 0.38** (0.11 to 1.28)	**160 fewer per 1,000** (from 72 more to 230 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Any reincarceration at 2-year follow-up (CJS database) - RCT												
1	randomised trials	very serious [1]	not serious	serious [2]	very serious [3]	none	15/31 (48.4%)	19/31 (61.3%)	**RR 0.79** (0.50 to 1.25)	**129 fewer per 1,000** (from 153 more to 306 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Reincarceration for a technical violation revocation or failure to comply with Sex Offender's Register (SOR) requirements at 2- or 4-year follow-up (CJS database) - RCT (reincarceration for revocation; 2-year follow-up)												
1	randomised trials	very serious [1]	not serious	serious [2]	serious [3]	none	13/27 (48.1%)	17/25 (68.0%)	**RR 0.71** (0.44 to 1.14)	**197 fewer per 1,000** (from 95 more to 381 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Reincarceration for a technical violation revocation or failure to comply with Sex Offender's Register (SOR) requirements at 2- or 4-year follow-up (CJS database) - Controlled non-randomised studies (failure to comly with SOR requirements; 4-year follow-up)												
1	observational studies	very serious [6]	not serious	not serious	very serious [3]	none	4/71 (5.6%)	6/71 (8.5%)	**RR 0.67** (0.20 to 2.26)	**28 fewer per 1,000** (from 68 fewer to 106 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Reconviction (Any) - UK (controlled non-randomised)												
1	observational studies	very serious [6]	not serious	not serious	serious [3]	none	7/71 (9.9%)	14/71 (19.7%)	**RR 0.50** (0.21 to 1.16)	**99 fewer per 1,000** (from 32 more to 156 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Reconviction (Any) - US (RCT)												
1	randomised trials	very serious [1]	not serious	serious [2]	serious [3]	none	8/31 (25.8%)	14/31 (45.2%)	**RR 0.57** (0.28 to 1.16)	**194 fewer per 1,000** (from 72 more to 325 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Reconviction (Any) - Canada (controlled non-randomised)												
2	observational studies	very serious [7]	serious [8]	serious [5]	serious [3]	none	22/104 (21.2%)	43/104 (41.3%)	**RR 0.48** (0.22 to 1.04)	**215 fewer per 1,000** (from 17 more to 323 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Reconviction (sexual)												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Reintegration programme (COSA)	Treatment as usual	Relative (95% CI)	Absolute (95% CI)		
3	observational studies	very serious [4]	not serious	serious [5]	serious [3]	none	8/175 (4.6%)	21/175 (12.0%)	**RR 0.41** (0.18 to 0.94)	**71 fewer per 1,000** (from 7 fewer to 98 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Reconviction (sexual) - UK												
1	observational studies	very serious [6]	not serious	not serious	very serious [3]	none	4/71 (5.6%)	5/71 (7.0%)	**RR 0.80** (0.22 to 2.86)	**14 fewer per 1,000** (from 55 fewer to 131 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Reconviction (sexual) - Canada												
2	observational studies	very serious [7]	not serious	serious [5]	not serious	none	4/104 (3.8%)	16/104 (15.4%)	**RR 0.26** (0.09 to 0.75)	**114 fewer per 1,000** (from 38 fewer to 140 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Reconviction (violent)												
3	observational studies	very serious [4]	not serious	serious [5]	not serious	none	13/175 (7.4%)	43/175 (24.6%)	**RR 0.34** (0.19 to 0.61)	**162 fewer per 1,000** (from 96 fewer to 199 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Reconviction (violent) - UK												
1	observational studies	very serious [6]	not serious	not serious	serious [3]	none	0/71 (0.0%)	7/71 (9.9%)	**RR 0.07** (0.00 to 1.15)	**92 fewer per 1,000** (from -- to 15 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Reconviction (violent) - Canada												
2	observational studies	very serious [7]	not serious	serious [5]	not serious	none	13/104 (12.5%)	36/104 (34.6%)	**RR 0.37** (0.21 to 0.65)	**218 fewer per 1,000** (from 121 fewer to 273 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Revocation - UK (failure to comply with SOR requirements: 4-year follow-up; controlled non-randomised studies)												
1	observational studies	very serious [4]	not serious	not serious	very serious [3]	none	4/71 (5.6%)	6/71 (8.5%)	**RR 0.67** (0.20 to 2.26)	**28 fewer per 1,000** (from 68 fewer to 106 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sub-analysis by country: Revocation - US (reincarceration for revocation; 2-year follow-up; RCT)												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Reintegration programme (COSA)	Treatment as usual	Relative (95% CI)	Absolute (95% CI)		
1	randomised trials	very serious [1]	not serious	not serious	serious [3]	none	13/27 (48.1%)	17/25 (68.0%)	**RR 0.71** (0.44 to 1.14)	**197 fewer per 1,000** (from 95 more to 381 fewer)	⨁◯◯◯ VERY LOW	CRITICAL

CI: Confidence interval; **RR:** Risk ratio

1. Duwe 2013 - high risk of selection bias (Prior sex crime conviction was 32% in intervention group compared with 10% in control group); No blinding; low attrition risks; low selective outcome bias; low risk of other bias.
2. 'Sex offender' - unclear proportion of participants with a paraphilic disorder
3. The 95% CI considered for imprecision was 0.8 to 1.25.
4. Bates 2014 - Controlled Non-RCT; high risk of selection bias; no blinding; unclear attrition risk of bias; low risk of selective outcome bias; low risk of other bias; Wilson 2007, Wilson 2009 - Controlled Non-RCT; high risk of selection bias; significant differences in baseline risk factors between groups; no blinding; unclear attrition risk of bias; low risk of selective outcome bias; low risk of other bias
5. Proportion of participants with paraphilia was unclear (Wilson 2009); only over half (Wilson 2007); majority (86%) of sample (Bates 2014).
6. Bates 2014 - Controlled Non-RCT; high risk of selection bias; no blinding; unclear attrition risk of bias; low risk of selective outcome bias; low risk of other bias;
7. Wilson 2007, Wilson 2009 - Controlled Non-RCT; high risk of selection bias; significant differences in baseline risk factors between groups; no blinding; unclear attrition risk of bias; low risk of selective outcome bias; low risk of other bias
8. I2>50%.

N.5.6 Therapeutic communities versus no treatment for paraphilic disorders

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Therapeutic communities	No treatment	Relative (95% CI)	Absolute (95% CI)		
Rearrest (CJS database; controlled non-randomised studies; longest follow-up available)												
1	observational studies	serious [1]	not serious	not serious	not serious	none	41/119 (34.5%)	607/1098 (55.3%)	**RR 0.62** (0.48 to 0.80)	**210 fewer per 1,000** (from 111 fewer to 287 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Rearrest (CJS database; controlled non-randomised studies; longest follow-up available) - 3-year follow-up												
1	observational studies	serious [1]	not serious	not serious	not serious	none	41/119 (34.5%)	607/1098 (55.3%)	**RR 0.62** (0.48 to 0.80)	**210 fewer per 1,000** (from 111 fewer to 287 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sex offence rearrest (CJS database; controlled non-randomised studies; longest follow-up available)												
1	observational studies	serious [1]	not serious	not serious	serious [2]	none	8/119 (6.7%)	81/1098 (7.4%)	**RR 0.91** (0.45 to 1.84)	**7 fewer per 1,000** (from 41 fewer to 62 more)	⨁◯◯◯ VERY LOW	CRITICAL

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Therapeutic communities	No treatment	Relative (95% CI)	Absolute (95% CI)		
Sex offence rearrest (CJS database; controlled non-randomised studies; longest follow-up available) - 3-year follow-up												
1	observational studies	serious [1]	not serious	not serious	serious [2]	none	8/119 (6.7%)	81/1098 (7.4%)	**RR 0.91** (0.45 to 1.84)	**7 fewer per 1,000** (from 41 fewer to 62 more)	⨁◯◯◯ VERY LOW	CRITICAL
Violent rearrest (CJS database; controlled non-randomised studies; longest follow-up available)												
1	observational studies	serious [1]	not serious	not serious	serious [2]	none	26/119 (21.8%)	288/1098 (26.2%)	**RR 0.83** (0.58 to 1.19)	**45 fewer per 1,000** (from 50 more to 110 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Violent rearrest (CJS database; controlled non-randomised studies; longest follow-up available) - 3-year follow-up												
1	observational studies	serious [1]	not serious	not serious	serious [2]	none	26/119 (21.8%)	288/1098 (26.2%)	**RR 0.83** (0.58 to 1.19)	**45 fewer per 1,000** (from 50 more to 110 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Incarceration (CJS database; controlled non-randomised studies; longest follow-up available)												
1	observational studies	serious [1]	not serious	not serious	serious [2]	none	12/119 (10.1%)	228/1098 (20.8%)	**RR 0.49** (0.28 to 0.84)	**106 fewer per 1,000** (from 33 fewer to 150 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Incarceration (CJS database; controlled non-randomised studies; longest follow-up available) - 3-year follow-up												
1	observational studies	serious [1]	not serious	not serious	serious [2]	none	12/119 (10.1%)	228/1098 (20.8%)	**RR 0.49** (0.28 to 0.84)	**106 fewer per 1,000** (from 33 fewer to 150 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Incarceration for sexual offence (CJS database; controlled non-randomised studies; longest follow-up available)												
1	observational studies	serious [1]	not serious	not serious	serious [2]	none	6/119 (5.0%)	42/1098 (3.8%)	**RR 1.32** (0.57 to 3.04)	**12 more per 1,000** (from 16 fewer to 78 more)	⨁◯◯◯ VERY LOW	CRITICAL
Incarceration for sexual offence (CJS database; controlled non-randomised studies; longest follow-up available) - 3-year follow-up												
1	observational studies	serious [1]	not serious	not serious	serious [2]	none	6/119 (5.0%)	42/1098 (3.8%)	**RR 1.32** (0.57 to 3.04)	**12 more per 1,000** (from 16 fewer to 78 more)	⨁◯◯◯ VERY LOW	CRITICAL
Incarceration for violent offence (CJS database; controlled non-randomised studies; longest follow-up available)												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Therapeutic communities	No treatment	Relative (95% CI)	Absolute (95% CI)		
1	observational studies	serious [1]	not serious	not serious	serious [2]	none	3/119 (2.5%)	74/1098 (6.7%)	**RR 0.37** (0.12 to 1.17)	**42 fewer per 1,000** (from 11 more to 59 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Incarceration for violent offence (CJS database; controlled non-randomised studies; longest follow-up available) - 3-year follow-up												
1	observational studies	serious [1]	not serious	not serious	serious [2]	none	3/119 (2.5%)	74/1098 (6.7%)	**RR 0.37** (0.12 to 1.17)	**42 fewer per 1,000** (from 11 more to 59 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Revocation (CJS database; controlled non-randomised studies; longest follow-up available)												
1	observational studies	serious [1]	not serious	not serious	serious [2]	none	18/115 (15.7%)	625/1310 (47.7%)	**RR 0.33** (0.21 to 0.50)	**320 fewer per 1,000** (from 239 fewer to 377 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Revocation (CJS database; controlled non-randomised studies; longest follow-up available) - 5-year follow-up												
1	observational studies	serious [1]	not serious	not serious	serious [2]	none	18/115 (15.7%)	625/1310 (47.7%)	**RR 0.33** (0.21 to 0.50)	**320 fewer per 1,000** (from 239 fewer to 377 fewer)	⨁◯◯◯ VERY LOW	CRITICAL

CI: Confidence interval; **RR:** Risk ratio

1. Lowden 2003 - Controlled Non-RCT; significant group differences at baseline on age, marital status and criminal history; high risk of selection bias; no blinding; low risk of attrition bias; high risk of selective outcome bias; low risk of other bias
2. The 95% C.I. considered for imprecision was 0.80 to 1.25

N.5.7 Cognitive behavioural therapy (CBT) versus treatment as usual for paraphilic disorders

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	CBT	Treatment as usual	Relative (95% CI)	Absolute (95% CI)		
Sexual reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 4-year follow-up (exhibitionists)												
1	observational studies	very serious [1]	not serious	not serious	serious [2]	none	4/17 (23.5%)	12/21 (57.1%)	**RR 0.41** (0.16 to 1.05)	**337 fewer per 1,000** (from 29 more to 480 fewer)	⨁◯◯◯ VERY LOW	CRITICAL

CI: Confidence interval; **RR:** Risk ratio

1. Marshall 1988a/b/1991 - Controlled Non-RCT with 4 and 9-year follow-up; No baseline risk differences; No blinding; unclear attrition risk of bias; low risk of selective outcome bias; low risk of other bias
2. The 95% CI considered for imprecision was 0.8 to 1.25.

N.5.8 Behavioural therapies versus treatment as usual for paraphilic disorders

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Behavioural therapies	Treatment as usual	Relative (95% CI)	Absolute (95% CI)		
Sexual reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 4-year follow-up (sex offenders against children)												
1	observational studies	very serious [1]	not serious	not serious	serious [2]	none	6/24 (25.0%)	12/20 (60.0%)	**RR 0.42** (0.19 to 0.91)	**348 fewer per 1,000** (from 54 fewer to 486 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sexual reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 9-year follow-up (exhibitionists)												
1	observational studies	very serious [1]	not serious	not serious	very serious [2]	none	9/23 (39.1%)	12/21 (57.1%)	**RR 0.68** (0.36 to 1.29)	**183 fewer per 1,000** (from 166 more to 366 fewer)	⨁◯◯◯ VERY LOW	CRITICAL

CI: Confidence interval; **RR:** Risk ratio

1. Marshall 1988a/b/1991 - Controlled Non-RCT with 4 and 9-year follow-up; No baseline risk differences; No blinding; unclear attrition risk of bias; low risk of selective outcome bias; low risk of other bias
2. The 95% CI considered for imprecision was 0.8 to 1.25.

N.5.9 Imaginal desensitization plus MPA versus MPA for paraphilic disorders

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Imaginal desensitization + medroxyprogesterone	Medroxyprogesterone only	Relative (95% CI)	Absolute (95% CI)		
Number of people who had a reduction in anomalous behaviours (26 weeks follow-up)												
1	randomised trials	very serious [1]	not serious	serious [2]	very serious [3]	none	9/10 (90.0%)	8/10 (80.0%)	**RR 1.12** (0.78 to 1.63)	**96 more per 1,000** (from 176 fewer to 504 more)	⨁◯◯◯ VERY LOW	CRITICAL
Number of people who had a reduction in anomalous desires (26 weeks follow-up)												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Imaginal desensitization + medroxyprogesterone	Medroxyprogesterone only	Relative (95% CI)	Absolute (95% CI)		
1	randomised trials	very serious [1]	not serious	serious [2]	very serious [3]	none	5/10 (50.0%)	3/10 (30.0%)	**RR 1.67** (0.54 to 5.17)	**201 more per 1,000** (from 138 fewer to 1,000 more)	⨁◯◯◯ VERY LOW	CRITICAL

CI: Confidence interval; **RR:** Risk ratio

1. McConaghy 1988 - unclear risk of selection bias, no blinding, low risk of attrition bias, high risk of selective outcome bias, low risk of other bias.
2. Unclear what percentage are currently in contact with the criminal justice system
3. The 95% CI considered for imprecision was 0.8 to 1.25.

N.5.10 Imaginal desensitization versus covert sensitization for paraphilic disorders

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Imaginal desensitization versus Covert sensitization only (Inpatient)	Control	Relative (95% CI)	Absolute		
Number of people who had a reduction in anomalous behaviours												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	very serious[3]	none	7/10 (70%)	4/10 (40%)	RR 1.75 (0.74 to 4.14)	300 more per 1000 (from 104 fewer to 1000 more)	⨁◯◯◯ VERY LOW	CRITICAL
Number of people who had a reduction in anomalous desires												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	very serious[3]	none	3/10 (30%)	5/10 (50%)	RR 0.6 (0.19 to 1.86)	200 fewer per 1000 (from 405 fewer to 430 more)	⨁◯◯◯ VERY LOW	CRITICAL

[1] McConaghy 1985 - unclear selection bias, no blinding, high risk of attrition bias, high risk of selective outcome bias, low other risk of bias,
[2] 13/20 had previously received convictions but unclear what percentage of the sample were currently in contact with the criminal justice system. Also 5 individuals requested treatment due to being homosexual, which would no longer be considered a paraphilia.
[3] The 95% CI considered for imprecision was 0.8 to 1.25.

N.5.11 Aversive conditioning and milieu therapy versus treatment as usual for paraphilic disorders

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aversive conditioning training and milieu therapy	Treatment as usual	Relative (95% CI)	Absolute (95% CI)		
Sexual and/or violent reconvictions at 21-year follow-up (CJS database) - Controlled non-randomised studies												
1	observational studies	very serious [1]	not serious	not serious	serious [2]	none	47/106 (44.3%)	35/91 (38.5%)	**RR 1.15** (0.82 to 1.61)	**58 more per 1,000** (from 69 fewer to 235 more)	⨁◯◯◯ VERY LOW	CRITICAL

CI: Confidence interval; **RR:** Risk ratio

1. Hanson 1993 - Controlled Non-RCT; significant baseline risk differences (+); no blinding; unclear attrition risk of bias; low risk of selective outcome bias; low risk of other bias.
2. The 95% CI considered for imprecision was 0.8 to 1.25.

N.5.12 Psychotherapy versus no treatment or treatment as usual for paraphilic disorders

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Psychotherapy	No treatment or treatment as usual	Relative (95% CI)	Absolute (95% CI)		
Rearrest (source of data not reported; controlled non-randomised studies; longest follow-up available) - 2-year follow-up												
1	observational studies	very serious [1]	not serious	serious [2]	not serious	none	3/92 (3.3%)	20/75 (26.7%)	**RR 0.12** (0.04 to 0.40)	**235 fewer per 1,000** (from 160 fewer to 256 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sex offence rearrest (source of data not reported; controlled non-randomised studies; longest follow-up available) - 2-year follow-up												
1	observational studies	very serious [1]	not serious	serious [2]	serious [3]	none	1/92 (1.1%)	6/75 (8.0%)	**RR 0.14** (0.02 to 1.10)	**69 fewer per 1,000** (from 8 more to 78 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Sexual reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - Length of follow-up not reported												
1	observational studies	very serious [4]	not serious	not serious	very serious [3]	none	5/23 (21.7%)	17/145 (11.7%)	**RR 1.85** (0.76 to 4.54)	**100 more per 1,000** (from 28 fewer to 415 more)	⨁◯◯◯ VERY LOW	CRITICAL
Violent reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - Length of follow-up not reported												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Psychotherapy	No treatment or treatment as usual	Relative (95% CI)	Absolute (95% CI)		
1	observational studies	very serious [4]	not serious	not serious	very serious [3]	none	3/23 (13.0%)	24/145 (16.6%)	**RR 0.79** (0.26 to 2.41)	**35 fewer per 1,000** (from 122 fewer to 233 more)	⨁◯◯◯ VERY LOW	CRITICAL
Breaches of the Sex Offender Register (CJS database; controlled non-randomised studies; longest follow-up available) - Length of follow-up not reported												
1	observational studies	very serious [4]	not serious	not serious	very serious [3]	none	8/23 (34.8%)	35/145 (24.1%)	**RR 1.44** (0.77 to 2.70)	**106 more per 1,000** (from 56 fewer to 410 more)	⨁◯◯◯ VERY LOW	CRITICAL

CI: Confidence interval; **RR:** Risk ratio

1. Peters 1968 - Controlled Non-RCT; group differences at baseline; no blinding; unclear attrition risk of bias; low risk of selective outcome bias; low risk of other bias.
2. 'Sex offender' - unclear proportion of participants with a paraphilic disorder; also an unknown proportion of participants in the intervention group had treatment delivered in a psychiatric inpatient unit
3. The 95% CI considered for imprecision was 0.8 to 1.25.
4. Craissati 2009 - Controlled Non-RCT; there might have selection bias issues such as unequal baseline risks between 2 groups and the individual psychoeducation group was also offered to to those who had already attempted group work; No blinding; only participants with available follow-up data were included; low risk of selective outcome bias; low risk of other bias

N.5.13 Polygraph testing versus treatment as usual for paraphilic disorders

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Polygraph testing	Treatment as usual	Relative (95% CI)	Absolute (95% CI)		
Reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 5-year follow-up												
1	observational studies	very serious [1]	not serious	not serious	serious [2]	none	41/104 (39.4%)	36/104 (34.6%)	**RR 1.14** (0.80 to 1.63)	**48 more per 1,000** (from 69 fewer to 218 more)	⨁◯◯◯ VERY LOW	CRITICAL
Sexual reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 5-year follow-up												
1	observational studies	very serious [1]	not serious	not serious	serious [2]	none	6/104 (5.8%)	7/104 (6.7%)	**RR 0.86** (0.30 to 2.46)	**9 fewer per 1,000** (from 47 fewer to 98 more)	⨁◯◯◯ VERY LOW	CRITICAL
Violent reconviction (CJS database; controlled non-randomised studies; longest follow-up available) - 5-year follow-up												

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Polygraph testing	Treatment as usual	Relative (95% CI)	Absolute (95% CI)		
1	observational studies	very serious [1]	not serious	not serious	serious [2]	none	3/104 (2.9%)	12/104 (11.5%)	**RR 0.25** (0.07 to 0.86)	**87 fewer per 1,000** (from 16 fewer to 107 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Incarceration (CJS database; controlled non-randomised studies; longest follow-up available) - 5-year follow-up												
1	observational studies	very serious [1]	not serious	not serious	serious [2]	none	49/104 (47.1%)	40/104 (38.5%)	**RR 1.23** (0.89 to 1.68)	**88 more per 1,000** (from 42 fewer to 262 more)	⨁◯◯◯ VERY LOW	CRITICAL
Violation of supervision conditions (CJS database; controlled non-randomised studies; longest follow-up available) - 5-year follow-up												
1	observational studies	very serious [1]	not serious	not serious	serious [2]	none	54/104 (51.9%)	47/104 (45.2%)	**RR 1.15** (0.87 to 1.52)	**68 more per 1,000** (from 59 fewer to 235 more)	⨁◯◯◯ VERY LOW	CRITICAL

CI: Confidence interval; **RR:** Risk ratio

1. McGrath 2007 - Controlled Non-RCT; baseline characters were similar between the groups; no blinding; low risk of detection bias; low attrition bias; low selective outcome bias; low risk of other bias
2. The 95% CI considered for imprecision was 0.80 to 1.25.

N.6 Service delivery models

N.6.1 Street Triage (Before and After)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Street triage	Control	Relative (95% CI)	Absolute		
Total s136 detentions per 100,000												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	89/100000* (0.09%)	107/100000* (0.11%)	RR 0.83 (0.63 to 1.1)	18.2 fewer per 100,000 (from 39.6 fewer to 10.7 more)	⨁OOO VERY LOW	CRITICAL
Number of s136 detentions in custody												
2	observational studies	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	strong association[4]	6085/24687 (24.6%)	9100/25227 (36.1%)	RR 0.68 (0.67 to 0.7)	115 fewer per 1000 (from 108 fewer to 119 fewer)	⨁⨁OO LOW	CRITICAL

Number of s136 detentions in hospital												
3	observational studies	serious[5]	no serious inconsistency	no serious indirectness	no serious imprecision	none	18613/24703 (75.3%)	16139/25250 (63.9%)	RR 1.18 (1.16 to 1.19)	115 more per 1000 (from 102 more to 121 more)	⊕OOO VERY LOW	CRITICAL

[1] *Reveruzzi 2016 – before and after study; low risk of selection bias as the groups were formed by before and after implementation of street triage; high risk of performance bias as there was no blinding involved; high rate of missing data and complete case analysis*

[2] *The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or two boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.*

[3] *Hywel Dda 2015 – before and after study; low risk of selection bias as the groups were formed by before and after implementation of street triage; high risk of performance bias as there was no blinding involved; high rate of missing data and complete case analysis.*

[4] *Evidence was upgraded if the effect estimate was considered to be large(I.e. 95% CI of RR <0.75 or RR>1.25).*

[5] *Powys 2015 – before and after study; low risk of selection bias as the groups were formed by before and after implementation of street triage; high risk of performance bias as there was no blinding involved; high rate of missing data and complete case analysis*

*The total population being looked at was not provided and the data was calculated per 100,000.

N.6.2 Diversion services

Before and After Diversion services

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Court Diversion**	**Control**	**Relative (95% CI)**	**Absolute**		
Duration between remand and assessment (days) (Better indicated by lower values)												
2	observational studies	Serious[1,2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	294	317	-	MD 31.76 lower (69.55 lower to 6.03 higher)	⊕OOO VERY LOW	CRITICAL
Days of total time on remand (Better indicated by lower values)												
1	observational studies	Serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	280	285	-	MD 17.6 lower (28.64 to 6.56 lower)	⊕OOO VERY LOW	IMPORTANT
Proportions of prisoners on bail												
1	observational studies	Serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	none	31/122 (25.4%)	20/98 (20.4%)	RR 1.25 (0.76 to 2.04)	51 more per 1000 (from 49 fewer to 212 more)	⊕OOO VERY LOW	CRITICAL
Attendance at alcohol and drug treatment programmes												
1	observational studies	Serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	none	13/41 (31.7%)	9/29 (31%)	RR 1.02 (0.51 to 2.07)	6 more per 1000 (from 152 fewer to 332 more)	⊕OOO VERY LOW	CRITICAL

OPD attendance rate for those release on bail												
1	observational studies	Serious[3]	no serious inconsistency	no serious indirectness	serious[4]	none	11/23 (47.8%)	7/13 (53.8%)	RR 0.89 (0.46 to 1.72)	59 fewer per 1000 (from 291 fewer to 388 more)	⊕OOO VERY LOW	CRITICAL
Registration of care programmes (CPA) and supervision registration (SR)												
1	observational studies	Serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	none	10/122 (8.2%)	4/98 (4.1%)	RR 2.01 (0.65 to 6.21)	41 more per 1000 (from 14 fewer to 213 more)	⊕OOO VERY LOW	CRITICAL

[1] Exworthy 1997- before and after study with no confounder being controlled; no blinding; unclear drop out and available case analysis
[2] Weaver 1997 – before and after study with no confounder being controlled; no blinding; unclear dropout with available case analysis
[3] Chambers 1999 – controlled cohort study with no confounder being controlled; no blinding; unclear drop out and available case analysis
[4] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or two boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.

Court diversion vs Community diversion services

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Court**	**Community Diversion**	**Relative (95% CI)**	**Absolute**		
Rate of re-incarceration in two years after index discharge												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	strong association[2]	60/214 (28%)	11/214 (5.1%)	RR 5.45 (2.95 to 10.08)	229 more per 1000 (from 100 more to 467 more)	⊕⊕OO LOW	IMPORTANT
100% attendance rate of appointments												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	47/214 (22%)	79/214 (36.9%)	RR 0.59 (0.44 to 0.81)	151 fewer per 1000 (from 70 fewer to 207 fewer)	⊕OOO VERY LOW	CRITICAL
Number of days in hospital (Better indicated by lower values)												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	214	214	-	MD 17 lower (64.44 lower to 30.44 higher)	⊕OOO VERY LOW	CRITICAL
Number of diverted participants with no mental health disorders												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	6/214 (2.8%)	0/214 (0%)	RR 13 (0.74 to 229.33)	-	⊕OOO VERY LOW	CRITICAL

[1] James 2002 - controlled cohort study; No blinding; Few missing cases and available case data analysis
[2] The effect size is considered large if 95% of RR<0.8 or RR>1.25.
[3] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or two boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.

N.6.3 Patient Navigation Intervention (PNI): Motivational feedback vs Control for substance misuse disorders (26 weeks follow-up)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Patient navigation intervention (at 26 weeks follow-up)	Control	Relative (95% CI)	Absolute		
Number of participants who used drugs												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	1/8 (12.5%)	2/10 (20%)	RR 0.62 (0.07 to 5.72)	76 fewer per 1000 (from 186 fewer to 944 more)	⊕OOO VERY LOW	CRITICAL
Number of participants who used alcohol to intoxication												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	1/8 (12.5%)	3/10 (30%)	OR 0.33 (0.03 to 4.04)	176 fewer per 1000 (from 287 fewer to 334 more)	⊕OOO VERY LOW	CRITICAL
Average days when mental health was not good in the last 30 days (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	8	10	-	MD 1.1 lower (9.74 lower to 7.54 higher)	⊕OOO VERY LOW	CRITICAL

[1] Binswanger 2015 - unclear randomization with appropriate allocation concealment, no blinding and appropriate attrition rate
[2] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or two boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.
[3] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.

N.6.4 Neighbourhood outreach (Before and After)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Neighbourhood outreach	Control	Relative (95% CI)	Absolute		
Proportion of crime contacts with policing team escalated to court												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	65/198 (32.8%)	149/308 (48.4%)	RR 0.68 (0.54 to 0.85)	155 fewer per 1000 (from 73 fewer to 223 fewer)	⊕OOO VERY LOW	CRITICAL

[1] Earl 2015 – before and after study; available case analysis; high risk of selective outcome report
[2] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or two boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25..

N.6.5 Drug Rehabilitation Requirement (DRR) (formerly known as Drug Testing Treatment Order (DTTO) vs TAU for substance misuse disorders

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	DRR	Control	Relative (95% CI)	Absolute		
MAP total scores (Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	25	27	-	MD 20.2 lower (52 lower to 11.6 higher)	⊕OOO VERY LOW	CRITICAL
Overall satisfaction (Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	25	27	-	MD 2.1 higher (1.16 to 3.04 higher)	⊕OOO VERY LOW	CRITICAL
HoNOS total score (Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	25	27	-	MD 0.2 lower (2.44 lower to 2.04 higher)	⊕OOO VERY LOW	CRITICAL

[1] Naeem 2007 – controlled cohort study; missing data imputed by regression

[2] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries. .

N.6.6 Case Management

Case Management vs TAU for substance misuse disorders

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Case management	TAU	Relative (95% CI)	Absolute		
Rearrest - Post-treatment												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	serious[3]	none	137/369 (37.1%)	56/135 (41.5%)	RR 0.9 (0.7 to 1.14)	41 fewer per 1000 (from 124 fewer to 58 more)	⊕OOO VERY LOW	IMPORTANT
Rearrest - 3 month follow-up												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness[2]	serious[3]	none	56/224 (25%)	48/238 (20.2%)	RR 1.24 (0.88 to 1.74)	48 more per 1000 (from 24 fewer to 149 more)	⊕⊕OO LOW	IMPORTANT

Reconviction –Post-treatment												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	serious[3]	none	58/369 (15.7%)	28/135 (20.7%)	RR 0.76 (0.51 to 1.14)	50 fewer per 1000 (from 102 fewer to 29 more)	⊕OOO VERY LOW	IMPORTANT
Reincarceration - Post-treatment												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	serious[3]	none	99/369 (26.8%)	44/135 (32.6%)	RR 0.82 (0.61 to 1.11)	59 fewer per 1000 (from 127 fewer to 36 more)	⊕OOO VERY LOW	IMPORTANT
Reincarceration - 3 month follow-up												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[3]	none	54/224 (24.1%)	55/238 (23.1%)	RR 1.04 (0.75 to 1.45)	9 more per 1000 (from 58 fewer to 104 more)	⊕OOO VERY LOW	IMPORTANT
Reincarceration - 12 month follow-up: female sample												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	very serious	none	16/77 (20.8%)	22/77 (28.6%)	RR 0.73 (0.41 to 1.27)	77 fewer per 1000 (from 169 fewer to 77 more)	⊕OOO VERY LOW	IMPORTANT
Reincarceration - 12 month follow-up: male sample												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[3]	none	120/354 (33.9%)	127/354 (35.9%)	RR 0.94 (0.77 to 1.16)	22 fewer per 1000 (from 83 fewer to 57 more)	⊕⊕OO LOW	IMPORTANT
Number of days jailed in past 6 months (12 month follow-up) (Better indicated by lower values)												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	no serious imprecision	none	207	204	-	MD 0.47 higher (6.65 lower to 7.59 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Drug related crimes in past 6 months (12 month follow-up) (Better indicated by lower values)												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[3]	none	207	204	-	MD 25.6 lower (235.88 lower to 184.68 higher)	⊕⊕OO LOW	IMPORTANT
Drug related criminal activity during treatment (12 months follow-up)												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[3]	none	32/147 (21.8%)	33/137 (24.1%)	RR 0.9 (0.59 to 1.39)	24 fewer per 1000 (from 99 fewer to 94 more)	⊕OOO VERY LOW	IMPORTANT
Self-reported alcohol use - During treatment												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	serious[3]	none	85/151 (56.3%)	93/137 (67.9%)	RR 0.83 (0.69 to 0.99)	115 fewer per 1000 (from 7 fewer to 210 fewer)	⊕⊕OO LOW	CRITICAL
Self-reported alcohol use - Post-treatment												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[3]	none	102/346 (29.5%)	90/334 (26.9%)	RR 1.09 (0.86 to 1.39)	24 more per 1000 (from 38 fewer to 105 more)	⊕⊕OO LOW	CRITICAL
Self-reported alcohol use - 12 month follow-up: female sample at post-treatment												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	no serious imprecision	none	4/77 (5.2%)	22/77 (28.6%)	RR 0.18 (0.07 to 0.5)	234 fewer per 1000 (from 143 fewer to 266 fewer)	⊕⊕⊕O MODERATE	CRITICAL
Self-reported alcohol use - 12 month follow-up: male sample at post-treatment												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[3]	none	138/354 (39%)	166/354 (46.9%)	RR 0.83 (0.7 to 0.99)	80 fewer per 1000 (from 5 fewer to 141 fewer)	⊕⊕OO LOW	CRITICAL
Self-reported drug use - During treatment (marijuana)												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	serious[3]	none	44/151 (29.1%)	49/137 (35.8%)	RR 0.81 (0.58 to 1.14)	68 fewer per 1000 (from 150 fewer to 50 more)	⊕⊕OO LOW	CRITICAL
Self-reported drug use - During treatment (hard drugs)												

1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[3]	none	76/151 (50.3%)	69/137 (50.4%)	RR 1 (0.79 to 1.26)	0 fewer per 1000 (from 106 fewer to 131 more)	⊕OOO VERY LOW	CRITICAL
Self-reported drug use - Post-treatment												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[3]	none	100/346 (28.9%)	90/334 (26.9%)	RR 1.07 (0.84 to 1.37)	19 more per 1000 (from 43 fewer to 100 more)	⊕⊕OO LOW	CRITICAL
Self-reported drug use - 12 month follow-up: female sample at post-treatment												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	very serious[3]	none	8/77 (10.4%)	13/77 (16.9%)	RR 0.62 (0.27 to 1.4)	64 fewer per 1000 (from 123 fewer to 68 more)	⊕OOO VERY LOW	CRITICAL
Self-reported drug use - 12 month follow-up: male sample at post-treatment												
1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	serious[3]	none	74/354 (20.9%)	95/354 (26.8%)	RR 0.78 (0.6 to 1.02)	59 fewer per 1000 (from 107 fewer to 5 more)	⊕⊕OO LOW	CRITICAL
Injection drug use (post-treatment)												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[3]	none	9/224 (4%)	12/238 (5%)	RR 0.8 (0.34 to 1.85)	10 fewer per 1000 (from 33 fewer to 43 more)	⊕OOO VERY LOW	CRITICAL
Abstinent - During treatment (at 12 months)												
1	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	serious[3]	none	42/147 (28.6%)	30/136 (22.1%)	RR 1.3 (0.86 to 1.94)	66 more per 1000 (from 31 fewer to 207 more)	⊕⊕OO LOW	CRITICAL
Abstinent - Post-treatment												
1	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[3]	none	54/224 (24.1%)	55/238 (23.1%)	RR 1.04 (0.75 to 1.45)	9 more per 1000 (from 58 fewer to 104 more)	⊕OOO VERY LOW	CRITICAL

[1] Hanlon 1999 - Unclear randomisation; No blinding; Unclear attrition
[2] Evidence was downgraded by one level because study population of one study (Hanlon 1999) differed from the review question in that the study included unclear proportion of ex-herion/cocaine users.
[3] Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or more boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.
[4] Scott 2012 - appropriate randomisation with concealment; No blinding; Unclear attrition bias; No selective outcomes report
[5] Johnson 2011/Friedmann 2012 - Unclear randomisation with unclear allocation concealment; No blinding; ITT analysis; Appropriate outcome report
[6] Rossman 1999 - Appropriate randomisation with allocation concealment; No blinding; Unclear drop-out; Appropriate selective outcome report
[7] Prendergast 2011 - Unclear randomisation with unclear allocation concealment; No blinding; Unclear attrition risk; high risk of selective outcome report

Case management vs active intervention for substance misuse disorders

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Case management**	**Active intervention**	**Relative (95% CI)**	**Absolute**		
Remained in treatment for 6 months												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	no serious imprecision	none	162/270 (60%)	34/99 (34.3%)	RR 1.75 (1.31 to 2.33)	258 more per 1000 (from 106 more to 457 more)	⊕OOO VERY LOW	CRITICAL
Rearrest - Post-treatment												
1	randomised	very	no serious	serious[2]	serious[3]	none	93/270	44/99	RR 0.78	98 fewer per 1000	⊕OOO	IMPORTANT

	trials	serious[1]	inconsistency				(34.4%)	(44.4%)	(0.59 to 1.02)	(from 182 fewer to 9 more)	VERY LOW	
Rearrest - 3 month follow-up												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	serious[3]	none	96/247 (38.9%)	93/264 (35.2%)	RR 1.1 (0.88 to 1.38)	35 more per 1000 (from 42 fewer to 134 more)	⊕OOO VERY LOW	IMPORTANT
Rearrest - 12 month follow-up												
1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	serious[3]	none	41/69 (59.4%)	33/64 (51.6%)	RR 1.15 (0.85 to 1.57)	77 more per 1000 (from 77 fewer to 294 more)	⊕OOO VERY LOW	IMPORTANT
Rearrest for drug crime (3 month follow-up)												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	very serious[6]	none	48/247 (19.4%)	49/264 (18.6%)	RR 1.05 (0.73 to 1.5)	9 more per 1000 (from 50 fewer to 93 more)	⊕OOO VERY LOW	IMPORTANT
Reconviction - Post-treatment												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	serious[3]	none	37/270 (13.7%)	21/99 (21.2%)	RR 0.65 (0.4 to 1.05)	74 fewer per 1000 (from 127 fewer to 11 more)	⊕OOO VERY LOW	IMPORTANT
Reconviction - 3 month follow-up												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	serious[3]	none	67/247 (27.1%)	54/264 (20.5%)	RR 1.33 (0.97 to 1.81)	68 more per 1000 (from 6 fewer to 166 more)	⊕OOO VERY LOW	IMPORTANT
Reincarceration - Post-treatment												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	very serious[3]	none	71/270 (26.3%)	28/99 (28.3%)	RR 0.93 (0.64 to 1.35)	20 fewer per 1000 (from 102 fewer to 99 more)	⊕OOO VERY LOW	IMPORTANT
Reincarceration - 3 month follow-up												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	serious[3]	none	88/247 (35.6%)	86/264 (32.6%)	RR 1.09 (0.86 to 1.39)	29 more per 1000 (from 46 fewer to 127 more)	⊕OOO VERY LOW	IMPORTANT
Any self-reported drug use (3 month follow-up)												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	serious[3]	none	100/247 (40.5%)	100/264 (37.9%)	RR 1.07 (0.86 to 1.33)	27 more per 1000 (from 53 fewer to 125 more)	⊕OOO VERY LOW	CRITICAL
Positive hair test (3 month follow-up) - Crack/Cocaine												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	serious[3]	none	97/247 (39.3%)	99/264 (37.5%)	RR 1.05 (0.84 to 1.3)	19 more per 1000 (from 60 fewer to 112 more)	⊕OOO VERY LOW	CRITICAL
Positive hair test (3 month follow-up) - Marijuana												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	serious[3]	none	50/247 (20.2%)	71/264 (26.9%)	RR 0.75 (0.55 to 1.03)	67 fewer per 1000 (from 121 fewer to 8 more)	⊕OOO VERY LOW	CRITICAL

[1] Hanlon 1999 - Unclear randomisation; No blinding; Unclear attrition
[2] Evidence was downgraded by one level because study population of one study (Hanlon 1999) differed from the review question in that the study included unclear proportion of ex-herion/cocaine users.
[3] Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or more boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.
[4] Needels 2005 - Unclear randomisation and allocation concealment; No blinding; Available case analysis with unclear drop-out; appropriate outcome report
[5] Kinlock 2007/Kinlock 2009/ Gordon 2008 - Permuted block randomisation with unclear allocation concealment; No blinding; ITT analysis with uncomparable drop-out rates
[6] Evidence was downgraded by one level due to serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of 50%-74.99%) and by two levels due to very serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of >75%). Random Effect Model was used if I-squared inconsistency statistic was more than or equal to 50%.

Assertive community treatment vs TAU for substance misuse disorders

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Assertive Community Treatment	TAU	Relative (95% CI)	Absolute		
Urine test positive for drug use during treatment												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	14/45 (31.1%)	6/45 (13.3%)	RR 2.33 (0.98 to 5.53)	177 more per 1000 (from 3 fewer to 604 more)	⊕OOO VERY LOW	CRITICAL
Injection drug use during treatment (self-report)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	10/56 (17.9%)	14/63 (22.2%)	RR 0.8 (0.39 to 1.66)	44 fewer per 1000 (from 136 fewer to 147 more)	⊕OOO VERY LOW	CRITICAL
Drug use during treatment (self-report)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	40/56 (71.4%)	40/63 (63.5%)	RR 1.13 (0.88 to 1.44)	83 more per 1000 (from 76 fewer to 279 more)	⊕OOO VERY LOW	CRITICAL
Reincarcerated during treatment												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	26/56 (46.4%)	32/63 (50.8%)	RR 0.91 (0.63 to 1.33)	46 fewer per 1000 (from 188 fewer to 168 more)	⊕OOO VERY LOW	IMPORTANT

[1] Martin 1993 - Unclear randomisation and allocation concealment; no blinding; Available case analysis with unclear drop-out; appropriate outcome report
[2] Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or more boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.

Case management vs TAU for mental health disorders other than substance misuse

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Case management	TAU	Relative (95% CI)	Absolute		
Service utilization												
2	randomised trials	very serious[1,2]	serious[3]	no serious indirectness	serious[4]	none	48/113 (42.5%)	52/110 (47.3%)	RR 0.98 (0.56 to 1.72)	9 fewer per 1000 (from 208 fewer to 340 more)	⊕OOO VERY LOW	CRITICAL
Rate of re-offending												
3	randomised trials	very serious[2,5,6]	no serious inconsistency	no serious indirectness	serious[4]	none	122/236 (51.7%)	99/196 (50.5%)	RR 1.03 (0.87 to 1.26)	15 more per 1000 (from 81 fewer to 136 more)	⊕OOO VERY LOW	CRITICAL
No of days in jail(up to 24 mths follow-up) (Better indicated by lower values)												
2	randomised trials	very serious[5,6]	no serious inconsistency	no serious indirectness	serious[4]	none	209	160	-	MD 12.24 higher (21.87 to 2.61 lower)	⊕OOO VERY LOW	CRITICAL
Quality of life (Better indicated by lower values)												
1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	serious[4]	none	53	39	-	MD 0.09 higher (0.51 lower to 0.69 higher)	⊕OOO VERY LOW	CRITICAL

[1] Jarrett 2012 – Unclear randomisation and allocation concealment; No blinding; Available case analysis
[2] Wang 2012 – Appropriate randomisation and allocation concealment; Unclear blinding; ITT analysis
[3]Evidence was downgraded by one level due to serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of 50%-74.99%) and by two levels due to very serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of >75%). Random Effect Model was used if I-squared inconsistency statistic was more than or equal to 50%.
[4] Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or more boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.
[5] Cosden 2003 – Unclear randomisation and allocation concealment; Unclear blinding; Available case analysis
[6] Solomon 1994 – Unclear randomisation and allocation concealment; No blinding; Unclear risk of attrition bias
[7] Cusack 2010 – Unclear randomisation and allocation concealment; ITT .

N.6.7 Drug court

Drug court vs TAU for substance misuse disorders

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Drug court	TAU	Relative (95% CI)	Absolute		
Days of substance use (12 month follow-up) - Alcohol (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	86	71	-	MD 43.10 lower (46.8 to 39.4 lower)	⊕⊕⊕O MODERATE	CRITICAL
Days of substance use (12 month follow-up) - Cocaine (Better indicated by lower values)												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	86	71	-	MD 43.70 lower (48.16 to 39.24 lower)	⊕⊕⊕O MODERATE	CRITICAL
Days of substance use (12 month follow-up) - Heroine (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	86	71	-	MD 54.50 lower (59.42 to 49.58 lower)	⊕⊕⊕O MODERATE	CRITICAL
Rearrest (12 month follow-up)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37/86 (43%)	46/71 (64.8%)	RR 0.66 (0.49 to 0.89)	220 fewer per 1000 (from 71 fewer to 330 fewer)	⊕⊕OO LOW	IMPORTANT
Maximum Crime Seriousness Scale (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	86	71	-	MD 1.12 lower (1.18 to 1.06 lower)	⊕⊕⊕O MODERATE	IMPORTANT

[1] Gottfredson 2005 - Unclear randomisation and allocation concealment; No blinding; Unclear analysis; Insufficient outcome report
[2] Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or more boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25

Drug court vs active intervention for substance misuse disorders

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Drug court versus active intervention**	**Control**	**Relative (95% CI)**	**Absolute**		
Removed from treatment due to unsatisfactory progress												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	11/85 (12.9%)	10/65 (15.4%)	RR 0.84 (0.38 to 1.86)	25 fewer per 1000 (from 95 fewer to 132 more)	⊕OOO VERY LOW	CRITICAL
Addiction Severity Index (ASI): alcohol composite score (Scale from 0 to 9; lower better)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	none	31	31	-	MD 0.02 lower (0.04 lower to 0 higher)	⊕OOO VERY LOW	CRITICAL
Addiction Severity Index (ASI): drug composite score (Scale from 0 to 9; lower better)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious	none	31	31	-	MD 0.01 lower (0.04 lower to 0.02 higher)	⊕OOO VERY LOW	CRITICAL
Number of sanctions during treatment (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	85	65	-	MD 0.90 lower (1.99 lower to 0.19 higher)	⊕⊕OO LOW	CRITICAL
Number of sanctions during treatment resulting in jail detention (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	67	54	-	MD 0.5 lower (0.99 to 0.01 lower)	⊕⊕OO LOW	IMPORTANT
Reincarceration during treatment												

1	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	very serious[6]	none	18/63 (28.6%)	25/68 (36.8%)	RR 0.78 (0.47 to 1.28)	81 fewer per 1000 (from 195 fewer to 103 more)	⊕OOO VERY LOW	IMPORTANT
Urine test positive for drugs (post-treatment)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[6]	none	2/31 (6.5%)	5/31 (16.1%)	RR 0.4 (0.08 to 1.91)	97 fewer per 1000 (from 148 fewer to 147 more)	⊕OOO VERY LOW	CRITICAL

[1] Messina 2012 - Inappropriate randomisation with adequate allocation concealment; No blinding; low risk of attrition bias; appropriate selective outcomes
[2] Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or more boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.
[3] Dakof 2010 - Unclear randomisation and allocation concealment; No blinding; ITT analysis; insufficient outcome report
[4] Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.
[5] Jones 2013 - Permuted block randomisation with unclear allocation concealment; No blinding; low risk of attrition bias; insufficient outcome report
[6] Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or more boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.

N.6.8 Opioid substitution therapy

Opioid substitution therapy + case management vs active intervention

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Opioid substitution therapy + case management**	**Active intervention**	**Relative (95% CI)**	**Absolute**		
Completed jail treatment												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	64/104 (61.5%)	68/107 (63.6%)	RR 0.96 (0.81 to 1.14)	25 fewer per 1000 (from 121 fewer to 89 more)	⊕⊕OO LOW	CRITICAL
Completed jail treatment - Female sample												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	27/32 (84.4%)	27/31 (87.1%)	RR 0.97 (0.79 to 1.18)	26 fewer per 1000 (from 183 fewer to 157 more)	⊕OOO VERY LOW	CRITICAL
Completed jail treatment - Male sample												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	37/72 (51.4%)	41/76 (53.9%)	RR 0.95 (0.7 to 1.29)	27 fewer per 1000 (from 162 fewer to 156 more)	⊕OOO VERY LOW	CRITICAL
Urine test positive for cocaine - 1 month follow-up												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[2]	none	31/70 (44.3%)	73/130 (56.2%)	RR 0.79 (0.58 to 1.07)	118 fewer per 1000 (from 236 fewer to 39 more)	⊕OOO VERY	CRITICAL

											LOW	
Urine test positive for cocaine - 6 month follow-up												
1	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	very serious[2]	none	15/25 (60%)	34/51 (66.7%)	RR 0.9 (0.62 to 1.31)	67 fewer per 1000 (from 253 fewer to 207 more)	⊕OOO VERY LOW	CRITICAL
Urine test positive for cocaine - 12 month follow-up												
1	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	serious[2]	none	19/44 (43.2%)	49/71 (69%)	RR 0.63 (0.43 to 0.91)	255 fewer per 1000 (from 62 fewer to 393 fewer)	⊕OOO VERY LOW	CRITICAL
Urine test positive for opioids - 1 month follow-up												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[2]	none	19/70 (27.1%)	67/130 (51.5%)	RR 0.53 (0.35 to 0.8)	242 fewer per 1000 (from 103 fewer to 335 fewer)	⊕OOO VERY LOW	CRITICAL
Urine test positive for opioids - 6 month follow-up												
1	randomised trials	very serious	no serious inconsistency	no serious indirectness	serious[2]	none	3/12 (25%)	26/45 (57.8%)	RR 0.43 (0.16 to 1.19)	329 fewer per 1000 (from 485 fewer to 110 more)	⊕OOO VERY LOW	CRITICAL
Urine test positive for opioids - 12 month follow-up												
1	randomised trials	very serious	no serious inconsistency	no serious indirectness	no serious imprecision	none	11/44 (25%)	40/71 (56.3%)	RR 0.44 (0.26 to 0.77)	315 fewer per 1000 (from 130 fewer to 417 fewer)	⊕⊕OO LOW	CRITICAL
Days of substance use (12 month follow-up) - Cocaine (Better indicated by lower values)												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[4]	none	71	133	-	MD 27.40 lower (47.25 to 7.55 lower)	⊕OOO VERY LOW	CRITICAL
Days of substance use (12 month follow-up) - Heroin (Better indicated by lower values)												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[4]	none	71	133	-	MD 36.80 lower (74.3 lower to 0.7 higher)	⊕OOO VERY LOW	CRITICAL
Self-reported days with drug use in past 30 days (6 month follow-up) - Crack/Cocaine												
1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	serious[2]	none	4/21 (19%)	19/41 (46.3%)	RR 0.41 (0.16 to 1.05)	273 fewer per 1000 (from 389 fewer to 23 more)	⊕OOO VERY LOW	CRITICAL
Self-reported days with drug use in past 30 days (6 month follow-up) - Heroin												
1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	no serious imprecision	none	3/21 (14.3%)	22/41 (53.7%)	RR 0.27 (0.09 to 0.79)	392 fewer per 1000 (from 113 fewer to 488 fewer)	⊕⊕OO LOW	CRITICAL
Self-reported days with drug use in past 30 days (6 month follow-up) - Marijuana												
1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	very serious[2]	none	2/21 (9.5%)	9/41 (22%)	RR 0.43 (0.1 to 1.83)	125 fewer per 1000 (from 198 fewer to 182 more)	⊕OOO VERY LOW	CRITICAL
Self-reported days with drug use in past 30 days (6 month follow-up) - Injection drug use												

1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	very serious[2]	none	2/21 (9.5%)	15/41 (36.6%)	RR 0.26 (0.07 to 1.03)	271 fewer per 1000 (from 340 fewer to 11 more)	⊕OOO VERY LOW	CRITICAL
Drug overdose - 6 month follow-up												
1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	very serious[2]	none	3/21 (14.3%)	7/41 (17.1%)	RR 0.84 (0.24 to 2.91)	27 fewer per 1000 (from 130 fewer to 326 more)	⊕OOO VERY LOW	CRITICAL
Drug overdose - 12 month follow-up												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	none	0/71 (0%)	6/133 (4.5%)	RR 0.14 (0.01 to 2.51)	39 fewer per 1000 (from 45 fewer to 68 more)	⊕OOO VERY LOW	CRITICAL
Rearrest - 6 month follow-up												
1	randomised trials	very serious[5]	no serious inconsistency	no serious indirectness	very serious[2]	none	7/21 (33.3%)	11/41 (26.8%)	RR 1.24 (0.56 to 2.73)	64 more per 1000 (from 118 fewer to 464 more)	⊕OOO VERY LOW	IMPORTANT
Rearrest - 12 month follow-up												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	none	38/71 (53.5%)	74/133 (55.6%)	RR 0.96 (0.74 to 1.25)	22 fewer per 1000 (from 145 fewer to 139 more)	⊕OOO VERY LOW	IMPORTANT
Self-reported days of criminal activity (12 months follow-up) (Better indicated by lower values)												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	71	133	-	MD 3.37 lower (35.27 lower to 28.53 higher)	⊕⊕OO LOW	IMPORTANT

[1] Gordon 2014 - Permuted blocks with adequate allocation concealment, No blinding with potential of effect size bigger in intervention group; available case analysis; appropriate outcome report
[2] Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or more boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.
[3] Kinlock 2007/Kinlock 2009/ Gordon 2008 - Permuted block randomisation with unclear allocation concealment; No blinding; ITT analysis with incomparable drop-out rates
[4] Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.
[5] McKenzie 2012 - Unclear randomisation and allocation concealment; No blinding with potential increased effect size in intervention arm; per protocol analysis; appropriate outcome report

N.6.9 Automated telephony with feedback vs Automated telephony alone

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Automated telephony with feedback**	**Automated telephony alone**	**Relative (95% CI)**	**Absolute**		
Improvement in Arnetz and Hasson stress questionnaire (AHSS) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	52	56	-	MD 2.5 higher (1.13 lower to 6.13 higher)	⊕⊕⊕O MODERATE	CRITICAL
Improvement in symptom checklist-8D (SCL-8D) (Better indicated by lower values)												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	52	56	-	MD 4.5 higher (0.22 to 8.78 higher)	⊕⊕OO LOW	CRITICAL
Improvement in daily stressor assessment (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	52	56	-	MD 1.91 higher (1.11 to 2.71 higher)	⊕⊕OO LOW	CRITICAL
Alcohol Urge Questionnaires: reduction in alcohol urge (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	52	56	-	MD 0.20 higher (0.35 lower to 0.75 higher)	⊕⊕⊕O MODERATE	CRITICAL
Alcohol Urge Questionnaires: reduction in alcohol use (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	52	56	-	MD 0.8 higher (0.11 to 1.49 higher)	⊕⊕OO LOW	CRITICAL
Alcohol Urge Questionnaires: reduction in drug use (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	52	56	-	MD 1 higher (0.41 to 1.59 higher)	⊕⊕OO LOW	CRITICAL
Alcohol Urge Questionnaires: reduction in drug urge (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	52	56	-	MD 0.3 higher (0.25 lower to 0.85 higher)	⊕⊕OO LOW	CRITICAL

[1] Andersson 2014 - Unclear randomisation with unclear allocation concealment; No blinding; Low drop-out rate with available rate analysis
[2] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.

N.6.10 IDDT vs TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**IDDT**	**Service as usual**	**Relative (95% CI)**	**Absolute**		
Rate of outpatient medication services												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	83/103 (80.6%)	51/79 (64.6%)	RR 1.25 (1.03 to 1.51)	161 more per 1000 (from 19 more to 329 more)	⊕⊕OO LOW	CRITICAL
No of days in hospital (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	103	79	-	MD 5.63 lower (9.59 to 1.67 lower)	⊕⊕OO LOW	CRITICAL
Rate of crisis visits (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	103	79	-	MD 2.26 lower (3.82 to 0.7 lower)	⊕⊕OO LOW	CRITICAL

[1] Chandler 2006 - Unclear randomization with unclear allocation concealment; Blinding was not reported; Analysis by imputation
[2] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or two boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25. *[3] The evidence was downgraded by one*

level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.

N.6.11 Housing first vs TAU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Housing First	TAU	Relative (95% CI)	Absolute		
Any offence												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	16/197 (8.1%)	19/100 (19%)	RR 0.43 (0.23 to 0.82)	108 fewer per 1000 (from 34 fewer to 146 fewer)	⊕⊕OO LOW	IMPORTANT
Any offence - Scattered HF+ACT												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	6/90 (6.7%)	11/50 (22%)	RR 0.3 (0.12 to 0.77)	154 fewer per 1000 (from 51 fewer to 194 fewer)	⊕⊕⊕O MODERATE	IMPORTANT
Any offence - Congregate HF												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	10/107 (9.3%)	8/50 (16%)	RR 0.58 (0.25 to 1.39)	67 fewer per 1000 (from 120 fewer to 62 more)	⊕OOO VERY LOW	IMPORTANT

[1] Somers 2013 - Unclear randomisation with unclear concealment; no blinding of participants and care administrators; ITT analysis

[2] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or two boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.

N.6.12 TIMA vs service as usual

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	TIMA	Service as usual	Relative (95% CI)	Absolute		
Bipolar Disorder Symptom Scale (BDSS) (Scale from 7 to 70; lower better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	30	30	-	MD 0.27 lower (0.75 lower to 0.21 higher)	⊕⊕OO LOW	CRITICAL
Brief Psychiatric Rating Scale (BPRS) (Scale from 18 to 126; lower better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	30	30	-	MD 0.97 higher (1.78 lower to 3.72 higher)	⊕⊕OO LOW	CRITICAL

[1] Ehret 2013 - inappropriate randomization with unclear concealment; no blinding; available case analysis

[2] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.

N.6.13 Service Brokerage Intervention vs Control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Service brokerage intervention	Control	Relative (95% CI)	Absolute		
Number of participants in contact with MH service												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	55/665 (8.3%)	47/660 (7.1%)	RR 1.16 (0.8 to 1.69)	11 more per 1000 (from 14 fewer to 49 more)	⊕OOO VERY LOW	CRITICAL
Number of participants who have seen GP												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	21/665 (3.2%)	13/660 (2%)	RR 1.6 (0.81 to 3.17)	12 more per 1000 (from 4 fewer to 43 more)	⊕OOO VERY LOW	CRITICAL
Number of participants who attended alcohol or drug service												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	18/665 (2.7%)	17/660 (2.6%)	RR 1.05 (0.55 to 2.02)	1 more per 1000 (from 12 fewer to 26 more)	⊕OOO VERY LOW	CRITICAL

[1] Kinner 2013/2014a/2014b - RCT with unclear allocation concealment; Blinding of care administrators; ITT analysis
[2] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or two boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.

N.6.14 Therapeutic communities

Therapeutic community versus waitlist control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Therapeutic community	waitlist control	Relative (95% CI)	Absolute		
Days until reincarceration (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	199	142	-	MD 83.58 higher (32.69 to 134.47 higher)	⊕OOO VERY LOW	IMPORTANT

[1] Wexler 1999 - Unclear randomisation and allocation concealment; No blinding with potential of effect size bigger in intervention group; ITT analysis; appropriate outcome report
[2] Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.

Modified therapeutic communities versus CBT informed psychoeducation

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Modified therapeutic community	CBT informed psychoeducation	Relative (95% CI)	Absolute		
Substance use (12 month follow-up)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	23/75 (30.7%)	35/64 (54.7%)	RR 0.56 (0.37 to 0.84)	241 fewer per 1000 (from 88 fewer to 345 fewer)	⊕OOO VERY LOW	CRITICAL
Alcohol use (12 month follow-up)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	15/75 (20%)	24/64 (37.5%)	RR 0.53 (0.31 to 0.93)	176 fewer per 1000 (from 26 fewer to 259 fewer)	⊕OOO VERY LOW	CRITICAL
Drug use (12 month follow-up)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	18/75 (24%)	28/64 (43.8%)	RR 0.55 (0.34 to 0.89)	197 fewer per 1000 (from 48 fewer to 289 fewer)	⊕OOO VERY LOW	CRITICAL
Reincarceration (12 month follow-up)												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	7/75 (9.3%)	21/64 (32.8%)	RR 0.28 (0.13 to 0.63)	236 fewer per 1000 (from 121 fewer to 285 fewer)	⊕⊕OO LOW	IMPORTANT
Alcohol/drug offence (12 month follow-up)												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[2]	none	27/75 (36%)	37/64 (57.8%)	RR 0.62 (0.43 to 0.9)	220 fewer per 1000 (from 58 fewer to 330 fewer)	⊕OOO VERY LOW	IMPORTANT
Criminal activity (12 month follow-up)												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[2]	none	35/75 (46.7%)	45/64 (70.3%)	RR 0.66 (0.5 to 0.89)	239 fewer per 1000 (from 77 fewer to 352 fewer)	⊕OOO VERY LOW	IMPORTANT

[1] Sullivan 2007 - unclear randomisation and allocation concealment; No blinding; unclear analysis; self-reported data
[2] Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or more boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.
[3] Sacks 2004 - Unclear randomisation and allocation concealment; unclear blinding; Available case analysis; inadequate outcome report

Enhanced therapeutic community versus standard therapeutic community

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Enhanced therapeutic community**	**Standard therapeutic community**	**Relative (95% CI)**	**Absolute**		
Engagement with treatment (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	232	219	-	MD 0.03 higher (0.01 lower to 0.07 higher)	⊕⊕OO LOW	CRITICAL
Negative mood (as rated by counsellor) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	no serious imprecision	none	230	219	-	MD 1.79 lower (2.09 to 1.49 lower)	⊕OOO VERY LOW	CRITICAL

[1] Czuchry 2003 – unclear randomisation and allocation concealment; no blinding; unclear attrition
[2] Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.

Gender-responsive therapeutic community versus standard therapeutic community

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Gender-responsive therapeutic community**	**standard therapeutic community**	**Relative (95% CI)**	**Absolute**		
Addiction Severity Index (ASI): alcohol composite score (Better indicated by lower values)												
1	randomised trials	Very serious[1]	no serious inconsistency	serious[2]	very serious[3]	none	60	55	-	MD 0.04 lower (0.08 lower to 0 higher)	⊕OOO VERY LOW	CRITICAL
Addiction Severity Index (ASI): psychological composite score (Better indicated by lower values)												
1	randomised trials	Very serious[1]	no serious inconsistency	serious[2]	no serious imprecision	none	60	55	-	MD 0.01 lower (0.1 lower to 0.08 higher)	⊕OOO VERY LOW	CRITICAL
Addiction Severity Index (ASI): drug composite score (Better indicated by lower values)												
1	randomised trials	Very serious[1]	no serious inconsistency	serious[2]	very serious[3]	none	60	55	-	MD 0.02 higher (0.0 lower to 0.04 higher)	⊕OOO VERY LOW	CRITICAL
Addiction Severity Index (ASI): family composite score (Better indicated by lower values)												
1	randomised trials	Very serious[1]	no serious inconsistency	Serious[2]	very serious[3]	none	60	55	-	MD 0.04 lower (0.12 lower to 0.04 higher)	⊕OOO VERY LOW	CRITICAL
Participated in aftercare upon release												

1	randomised trials	Very serious[1]	no serious inconsistency	serious[2]	very serious[4]	none	28/60 (46.7%)	30/55 (54.5%)	RR 0.86 (0.6 to 1.23)	76 fewer per 1000 (from 218 fewer to 125 more)	⊕OOO VERY LOW	CRITICAL
Months spent in aftercare (Better indicated by lower values)												
1	randomised trials	Very serious[1]	no serious inconsistency	Serious[2]	very serious[3]	none	60	55	-	MD 1.50 higher (0.29 to 2.71 higher)	⊕OOO VERY LOW	CRITICAL
Disciplinary removal from first residential treatment post-release												
1	randomised trials	Very serious[1]	no serious inconsistency	Serious[2]	very serious[4]	none	8/60 (13.3%)	8/55 (14.5%)	RR 0.92 (0.37 to 2.28)	12 fewer per 1000 (from 92 fewer to 186 more)	⊕OOO VERY LOW	CRITICAL
Reincarceration (12 month follow-up)												
1	randomised trials	Very serious[1]	no serious inconsistency	Serious[2]	very serious[4]	none	18/60 (30%)	25/55 (45.5%)	RR 0.66 (0.41 to 1.07)	155 fewer per 1000 (from 268 fewer to 32 more)	⊕OOO VERY LOW	IMPORTANT
Voluntarily dropped-out from first residential treatment post-release												
1	randomised trials	Very serious[1]	no serious inconsistency	serious[2]	very serious[4]	none	10/60 (16.7%)	17/55 (30.9%)	RR 0.54 (0.27 to 1.08)	142 fewer per 1000 (from 226 fewer to 25 more)	⊕OOO VERY LOW	CRITICAL
Months until reincarceration (Better indicated by lower values)												
1	randomised trials	Very serious[1]	no serious inconsistency	Serious[2]	very serious[3]	none	60	55	-	MD 1.90 higher (0.5 to 3.3 higher)	⊕OOO VERY LOW	IMPORTANT

[1] *Messina 2010 - high risk of selection bias; No blinding; available case analysis; unclear selective outcome report*
[2] *Evidence was downgraded by one level because study population of one study (Messina 2010) differed from the review question in that not all the participants met the proxy measure criteria for substance misuse disorder.*
[3] *Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or both boundaries of the defined minimally important difference (MID) for the outcome, respectively. For continuous outcomes, +/-0.5 (mean for 2 studies and median for 3 or more studies) times SD of the control group (if MD was used) were considered as MID boundaries.*
[4] *Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or more boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.*

Gender-specific therapeutic community versus CBT informed psychoeducation

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Gender-specific therapeutic community**	**CBT informed psychoeducation**	**Relative (95% CI)**	**Absolute**		
Self-reported criminal activity (sexual)												
1	randomised	very	no serious	no serious	very serious[2]	none	3/163	8/151	RR 0.35	34 fewer per 1000	⊕OOO	IMPORTANT

	trials	serious[1]	inconsistency	indirectness			(1.8%)	(5.3%)	(0.09 to 1.29)	(from 48 fewer to 15 more)	VERY LOW	
Receiving mental health treatment at follow-up												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	65/163 (39.9%)	63/151 (41.7%)	RR 0.96 (0.73 to 1.25)	17 fewer per 1000 (from 113 fewer to 104 more)	⊕OOO VERY LOW	CRITICAL
Alcohol use (follow-up NR)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	41/163 (25.2%)	29/151 (19.2%)	RR 1.31 (0.86 to 2)	60 more per 1000 (from 27 fewer to 192 more)	⊕OOO VERY LOW	CRITICAL
Frequency of alcohol use (follow-up NR) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	87	75	-	MD 0.25 higher (0.42 lower to 0.92 higher)	⊕⊕OO LOW	CRITICAL
Frequency of drug use (follow-up NR) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	111	95	-	MD 0.42 lower (1.14 lower to 0.3 higher)	⊕⊕OO LOW	CRITICAL
Self-reported drug use - 6 month follow-up												
2	randomised trials	very serious[1,3]	no serious inconsistency	no serious indirectness	serious[2]	none	76/374 (20.3%)	87/328 (26.5%)	RR 0.77 (0.59 to 1.01)	61 fewer per 1000 (from 109 fewer to 3 more)	⊕OOO VERY LOW	CRITICAL
Self-reported drug use - 12 month follow-up												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[2]	none	50/207 (24.2%)	54/163 (33.1%)	RR 0.73 (0.53 to 1.01)	89 fewer per 1000 (from 156 fewer to 3 more)	⊕OOO VERY LOW	CRITICAL
Rearrest - 6 month follow-up												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[2]	none	19/211 (9%)	32/177 (18.1%)	RR 0.5 (0.29 to 0.85)	90 fewer per 1000 (from 27 fewer to 128 fewer)	⊕OOO VERY LOW	IMPORTANT
Rearrest - 12 month follow-up												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[2]	none	23/207 (11.1%)	11/163 (6.7%)	RR 1.65 (0.83 to 3.28)	44 more per 1000 (from 11 fewer to 154 more)	⊕OOO VERY LOW	IMPORTANT
Rearrest - Follow-up NR												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	42/163 (25.8%)	53/151 (35.1%)	RR 0.73 (0.52 to 1.03)	95 fewer per 1000 (from 168 fewer to 11 more)	⊕OOO VERY LOW	IMPORTANT
Reincarceration												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[2]	none	59/257 (23%)	59/211 (28%)	RR 0.82 (0.6 to 1.12)	50 fewer per 1000 (from 112 fewer to 34 more)	⊕OOO VERY LOW	IMPORTANT

Self-reported criminal activity (any) - 6 month follow-up												
2	randomised trials	very serious[1,3]	no serious inconsistency	no serious indirectness	serious[2]	none	130/374 (34.8%)	149/328 (45.4%)	RR 0.77 (0.64 to 0.92)	104 fewer per 1000 (from 36 fewer to 164 fewer)	⊕OOO VERY LOW	IMPORTANT
Self-reported criminal activity (any) - 12 month follow-up												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[2]	none	72/207 (34.8%)	67/163 (41.1%)	RR 0.85 (0.65 to 1.1)	62 fewer per 1000 (from 144 fewer to 41 more)	⊕OOO VERY LOW	IMPORTANT
Self-reported criminal activity (drugs) - 6 month follow-up												
2	randomised trials	very serious[1,3]	no serious inconsistency	no serious indirectness	serious[2]	none	106/374 (28.3%)	108/328 (32.9%)	RR 0.86 (0.69 to 1.08)	46 fewer per 1000 (from 102 fewer to 26 more)	⊕OOO VERY LOW	IMPORTANT
Self-reported criminal activity (drugs) - 12 month follow-up												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[2]	none	62/207 (30%)	60/163 (36.8%)	RR 0.81 (0.61 to 1.09)	70 fewer per 1000 (from 144 fewer to 33 more)	⊕OOO VERY LOW	IMPORTANT
Receiving substance abuse treatment at follow-up												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	109/163 (66.9%)	118/151 (78.1%)	RR 0.86 (0.75 to 0.98)	109 fewer per 1000 (from 16 fewer to 195 fewer)	⊕OOO VERY LOW	CRITICAL
Beck Depression Inventory (BDI) total score (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	163	151	-	MD 2.64 lower (5.26 to 0.02 lower)	⊕⊕OO LOW	CRITICAL
Brief Symptom Inventory (BSI) total score (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	163	151	-	MD 1.63 lower (4.45 lower to 1.19 higher)	⊕⊕OO LOW	CRITICAL
Post-traumatic Symptom Severity Scale (PSS) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	163	151	-	MD 2.90 lower (5.68 to 0.12 lower)	⊕⊕OO LOW	CRITICAL

[1] Sacks 2008 - unclear randomisation and allocation concealment; No blinding; analysis by regression technique; appropriate outcome report
[2] Evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or more boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.
[3] Sacks 2012a - unclear randomisation and allocation concealment; No blinding with potential of effect size bigger in intervention group; available case analysis

Re-entry modified therapeutic community versus treatment as usual

Quality assessment							No of patients		Effect		Quality	Importance
No of	Design	Risk of	Inconsistency	Indirectness	Imprecision	Other	Re-entry modified	treatment as	Relative	Absolute		

studies		bias				considerations	therapeutic community	usual	(95% CI)			
Reincarceration (12 month post prison release)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	14/71 (19.7%)	21/56 (37.5%)	RR 0.53 (0.29 to 0.94)	176 fewer per 1000 (from 23 fewer to 266 fewer)	⊕OOO VERY LOW	IMPORTANT
Criminal activity												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	25/63 (39.7%)	29/47 (61.7%)	RR 0.64 (0.44 to 0.94)	222 fewer per 1000 (from 37 fewer to 346 fewer)	⊕OOO VERY LOW	IMPORTANT
Alcohol/Drug offence												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	23/63 (36.5%)	27/47 (57.4%)	RR 0.64 (0.42 to 0.96)	207 fewer per 1000 (from 23 fewer to 333 fewer)	⊕OOO VERY LOW	IMPORTANT

[1] Sacks 2012b – inappropriate randomisation without allocation concealment; no blinding; ITT analysis; lack of outcome report on percentages of therapeutic community in prison
[2] The evidence was downgraded by one level and two levels if the confidence interval crossed or touched one or two boundaries of the defined minimally important difference (MID) for the outcome respectively. The MID boundaries for dichotomous outcomes (RR) were 0.8 to 1.25.

N.7 Staff Training

N.7.1 Organisational linkage intervention (OLI) plus medication-assisted training (MAT) vs Training alone for substance misuse disorders

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Organisational Linkage Intervention (OLI) plus training (RQ 5.1)	Training alone	Relative (95% CI)	Absolute		
MAT-Methadone: Familiarity with medication (Change from baseline to post intervention; range -4 to 4; higher is better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	383	464	Mean 0.26 (SD 1.01)	MD 0.14 higher (0.03 lower to 0.31 higher)	⊕⊕⊕O MODERATE	IMPORTANT
MAT-Methadone: Referral knowledge (Change from baseline to post intervention; range -4 to 4; higher is better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	383	464	Mean 0.24 (SD 1.23)	MD 0.04 higher (0.11 lower to 0.19 higher)	⊕⊕⊕O MODERATE	IMPORTANT
MAT-Methadone:Intent to refer clients to MAT (Change from baseline to post intervention; range -4 to 4; higher is better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	383	464	Mean 0.05 (SD 1.24)	MD 0.38 higher (0.19 to 0.57 higher)	⊕⊕⊕O MODERATE	IMPORTANT
MAT-Methadone: Overall perception and knowledge (Change from baseline to post intervention; range -4 to 4; higher is better)												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	383	464	Mean 0.01 (SD 0.04)	MD 0.2 higher (0.13 to 0.27 higher)	⊕⊕⊕O MODERATE	IMPORTANT
MAT-Buprenorphine: Familiarity with the Medication (Change from baseline to post intervention; range -4 to 4; higher is better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	383	464	Mean 0.39 (SD 1.52)	MD 0.01 higher (0.19 lower to 0.21 higher)	⊕⊕⊕O MODERATE	IMPORTANT
MAT-Buprenorphine: Referral Knowledge (Change from baseline to post intervention; range -4 to 4; higher is better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	383	464	Mean 0.34 (SD 1.33)	MD 0.07 higher (0.12 lower to 0.26 higher)	⊕⊕⊕O MODERATE	IMPORTANT
MAT-Buprenorphine: Intent to refer clients to MAT (Change from baseline to post intervention; range -4 to 4; higher is better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	383	464	Mean 0.15 (SD 1.35)	MD 0.15 higher (0.02 lower to 0.32 higher)	⊕⊕⊕O MODERATE	IMPORTANT
MAT- Buprenorphine: Overall perception and knowledge (Change from baseline to post intervention; range -4 to 4; higher is better)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	383	464	Mean 0.03 (SD 0.66)	MD 0.13 higher (0.05 to 0.21 higher)	⊕⊕⊕O MODERATE	IMPORTANT

1. Friedmann 2015 - unclear randomisation and concealment; comparable management of experimental and control group; appropriate outcome report