# Functional Family Therapy (FFT) for Adolescents with Antisocial Behaviours: Impacts on Mental Health Issues, Family Functioning and Recidivism

Debra Delaney (B Soc Sc (Hons))

This thesis is presented in partial fulfilment of the requirements for the degree of Master of Clinical Psychology, School of Psychological Sciences, University of Newcastle, Australia.

October, 2023

## Declarations

### Statement of Originality

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to the final version of my thesis being made available worldwide when deposited in the University library, subject to the provisions of the Copyright Act 1968.

### Acknowledgement of Collaboration

I hereby certify that the work embodied in this thesis contains a scholarly work of which I am a joint author, along with Jessica Birchall (Librarian), Dr Renate Thienel (Masters of Clinical Psychology research student) and Dr Tanya Hanstock (Supervisor). I contributed to the development of the research question, identifying relevant publications, assessing study quality, summarising the evidence, interpreting the findings, and writing the thesis. I have compiled this thesis as part requirement of a Master of Clinical Psychology and as such have taken the lead in writing of the initial and final versions of the thesis.

### Acknowledgement of Authorship

I hereby certify that the work embodied in this thesis contains scholarly work of which I am a joint author. I have included as part of the thesis a written statement, endorsed by my supervisor, attesting to my contribution to the scholarly work. I wrote the thesis manuscript, and co-author Dr Tanya Hanstock provided feedback and editing.

Debra Delaney (Research Student) Date: 19.10.2023

Dr Tanya Hanstock (Primary Supervisor) Date: 19.10.2023

**Acknowledgements**

I would like to express my gratitude to Dr Tanya Hanstock for guidance, patience and for helping me to develop my research skills. I have greatly appreciated your encouragement and support. I would also like to thank Jessica Birchall (Librarian) for helping with the search of databases for the Systematic review. I would like to thank Dr Renate Thienel for assistance with study screening, study quality assessment and proofreading. Thank you also to my family for their endless patience and understanding.

## Table of Contents

## List of Figures



## List of Tables



## List of Appendices

## Formatting Style Used for this Thesis

This thesis is formatted according to the American Psychological Association (APA) Seventh Edition. This is in accordance with the instructions to authors for the Journal of Family Psychology (Appendix A).

# Functional Family Therapy (FFT) for Adolescents with Antisocial Behaviours: Impacts on Mental Health Issues, Family Functioning and Recidivism

Debra Delaney, Renate Thienel & Tanya Hanstock

School of Psychological Sciences, The University of Newcastle, Callaghan, NSW, 2803, Australia.

Correspondence Author

Debra Delaney, School of Psychological Sciences, The University of Newcastle, University Drive, Callaghan, NSW Australia, 2308. Email: Debra.Delaney@uon.edu.au.

The authors declared no conflicts of interest with respect to the authorship and publication of this article.

Word Count: 8570

## Abstract

**Background**: FFT is a short-term family-based intervention designed to improve communication patterns in families with adolescents engaging in antisocial behaviour. Currently, FFT is disseminated to families with adolescents experiencing alcohol or other drug disorders, low IQ, serious offending behaviours and/or mental health (MH) issues or diagnoses.

**Aim**: The primary aim was to examine the effectiveness of FFT on the MH of adolescents who engage in antisocial behaviour and have comorbid MH issues or diagnoses. The secondary aim to was to explore the impact of FFT on adolescents' recidivism and family functioning in domains such as communication styles and conflict level.

**Method**: Eleven electronic databases were reviewed. Peer reviewed studies that used FFT to treat families with adolescents who engage in antisocial behaviour were included regardless of study design or if the participants had MH issues.

**Results**: The search yielded 220 studies, with 28 meeting inclusion criteria. FFT was effective in reducing a variety of MH issues for most adolescents. FFT was also effective in reducing recidivism and improving family functioning in most families. However, FFT did not prove to be superior to other interventions.

**Discussion**: FFT, as well other interventions, are potential efficacious treatments for MH issues in adolescents who engage in antisocial behaviour and have comorbid MH issues. However, these findings should be interpreted with caution, given the methodological problems and high risk of bias in the studies. More methodologically sound studies are required, especially focusing on adolescents with comorbid MH issues.


**Keywords**: *FFT, adolescents, antisocial, mental health*

# Functional Family Therapy (FFT) for Adolescents with Antisocial Behaviours: Impacts on Mental Health Issues, Recidivism, and Family Functioning

Functional Family Therapy (FFT) is a short-term family-based intervention designed for adolescents who engage in antisocial behaviour and their families (Alexander et al., 2013). FFT is based on a combination of behaviourism and systems theory and is purported to be effective in reducing adolescent antisocial behaviour through changes in the family system (Alexander et al., 2013). Adolescent antisocial behaviour is a significant public health issue and involves behaviours characterised by the violation of the rights of others or societal rules, as well as oppositional/defiant behaviours (American Psychiatric Association [APA], 2022). FFT has been widely implemented in the USA and other high-income countries (Littell et al., 2023). Although FFT was originally designed to modify maladaptive communication patterns in families with adolescents who engaged in antisocial behaviour, it has been disseminated to adolescents with a wide variety of characteristics, including many with mental health issues or diagnoses.

## Functional Family Therapy

FFT is a short-term, structured, strengths based intensive family intervention for families with an adolescent (aged 11 to 18 years) with an average of 12 sessions conducted over a 3- to 4-month period, usually in the home or clinic settings (Alexander et al., 2013). There are five distinct phases of FFT (Alexander et al., 2013). First, engagement to maximise family members participation. Second, motivation to decrease blame and negativity as well as increase hope and motivation for change. Third, relational assessment to identify how family relational dynamics and risk and protective factors and are associated with problematic behaviour. Fourth, behaviour change to improve skills in communication, conflict management and problem solving. Finally, generalisation, to maintain and extend skill use in

domains outside the family, prevent relapse and refer the family to additional support and services (Alexander et al., 2013).

FFT is based on a combination of behaviourism and systems theory (Alexander & Parsons, 1982). Family systems theory purports the family unit is a complex social system in which family members interact to influence each other's behaviour. Thus, change in one individual within a family will influence the entire system, and therefore change other members (Alexender & Parsons, 1973). Behaviourism is a theory of learning that posits all behaviours are learned through interaction with the environment through conditioning (Alexander & Parsons, 1982). According to behaviourism, all behaviour is a response to environmental stimuli (Alexander & Parsons, 1982). FFT posits the functional outcome of a behaviour is the reason the behaviour is carried out (Alexander & Parsons, 1982). Within this model, all behaviour, both adaptive and maladaptive, is maintained because it functions to meet individual and relationship needs. Thus, the FFT treatment approach was developed to address the unique construct of relational functions (Alexander et al., 2013).

## Adolescent Antisocial Behaviour

Early FFT interventions aimed to modify maladaptive communication patterns in families with adolescents who engaged in antisocial behaviour (Alexander & Parsons, 1973; Parsons & Alexander, 1973). It was hypothesised that altering family communication patterns to those more characteristic of healthy families, would allow families to adjust to the stressors of having adolescents who engaged in antisocial behaviour. Antisocial behaviours are those characterised by the types of behaviours found in the diagnostic criteria within the Diagnostic and Statistical Manual of Mental Disorders (APA, 2022) for Conduct Disorder (CD) and Oppositional Defiant Disorder (ODD). The over-arching common theme of CD type behaviours is the violation of the basic rights of others or age-appropriate societal norms, rules, or laws. This can include aggression to people or animals, destruction of property, and

deceitfulness or theft. ODD type behaviours include argumentative/defiant behaviour, or vindictiveness that can be present in multiple settings and impair social functioning. Antisocial behaviours can also involve running away from home, school truancy, and alcohol or other drug (AOD) use (APA, 2022). For some adolescents, the number or type of antisocial behaviours do not meet the diagnostic threshold for any of the specific disorders, even though the symptoms may be associated with clinically significant impairment (APA, 2022).

## Impact of Adolescent Antisocial Behaviour

The impairment associated with engaging in adolescent antisocial behaviour can have lifelong negative consequences for individuals, including poorer overall health (Paradis et al., 2016) and reduced educational and occupational opportunities (Sawyer et al., 2015). Antisocial behaviour also has negative effects on the adolescent's peer relationships, as well as being an extremely challenging time for families (Alexander et al., 2013). These types of antisocial behaviours are also a significant public health issue that causes serious concern for communities (Alexander et al., 2013). In Australia, adolescent antisocial behaviour is a costly issue, and includes costs for youth justice services, law enforcement and victim assistance, as well as costs related to mental and physical health services (Heerde et al., 2018). The estimated rate of adolescent antisocial behaviour in Australia is estimated to be 5% to 20% (Heerde et al., 2018). The financial cost of crime in Australia has been estimated to be $36 billion per year, with adolescence contributing strongly to the costs associated with crime (Hemphill, 2015).

## The Family and Adolescent Antisocial Behaviour

There are a range of factors within the community, school, peer group and individual domains that can contribute to an adolescent's risk of engaging in antisocial behaviours (Cox et al., 2016). The family has been shown to play a crucial role in its start, escalation, and

maintenance (Alexander & Parsons, 1982; Uink et al., 2017). Some authors assert that the family is the primary source of adolescent behaviour both in and outside the home, and that conduct problems arise when relational dynamics (e.g., level and quality of communication) in the family become dysfunctional (Alexander et al., 2000; Sexton & Turner, 2010). Thus, family-based interventions have increasingly become popular in attempts to reduce adolescent antisocial behaviour. Family focussed interventions, primarily focus on family level risk factors as the reduction of these have been linked to diminished behaviour problems (Carr, 2014). One such approach is FFT, which, since its inception in the early 1970's, has expanded its reach to include over 1,600 practitioners from over 330 organisations providing FFT to 50,000 families each year in 45 North American states and 10 countries around the world (FFT Limited Liability Company, 2022).

## Participant Characteristics

This expansion has resulted in FFT being delivered to adolescents with a much wider range of characteristics than those for whom it was originally intended. Participants in the early iterations of FFT were mostly white, middle class, first-time offenders in the early years of adolescents, with minor offences (Alexander & Parsons, 1973; Parsons & Alexander, 1973). Participants were also recruited from a region in the United States where most of the people are known for their dedication to Morman religious practices and who place great value on family life (Gordon et al., 1988). However, adolescents who engage in antisocial behaviour is a population with significant heterogeneity in individual differences (Dopp et al., 2017; Weisz et al., 2019). More recently, FFT has been delivered to adolescents who have a very wide variety of characteristics including those who have multiple offences (Gordon et al., 1988), serious offences (Sexton & Turner, 2010), below average IQ (Humayun et al., 2017), low SES (Hartnett, et al., 2016), substance use disorder (Waldron et al., 2001), alcohol use disorder (Slesnick & Prestopnik, 2009), engaged with a forensic-psychiatric outpatient

clinic (Van der Put et al., 2012), diagnosed with depression (Rohde et al., 2014), or were in treatment for mental health issues (Celinska et al., 2019).

## Mental Health Issues

Antisocial behaviours are associated with a significantly increased risk of a range of mental health issues (National Institute for Health and Care Excellence [NICE], 2017). For adolescents involved in the juvenile justice system, some estimate that between 65% and 75% have at least one mental health disorder (NICE, 2017). Further, 46% percent of boys and 36% of girls with CD or ODD are known to have at least one comorbid mental health issue (NICE, 2017). Common comorbid disorders in adolescents with conduct problems include depressive and anxiety disorders, substance use problems, trauma related issues, language impairment, and learning difficulties (Fairchild et al., 2019). Between 11% and 50% of juvenile justice-involved adolescents have been shown to meet diagnostic criteria for Posttraumatic Stress Disorder (PTSD; Abram et al., 2013) and around 40% of young people who meet criteria for CD also have a diagnosis of Attention Deficit/Hyperactivity Disorder (ADHD; NICE, 2017).

Given the high prevalence of mental health issues among adolescents with conduct problems and those involved in the juvenile justice system, it is crucial to examine the gaps in knowledge about whether adolescent mental health issues are associated with positive treatment outcomes (Kang et al., 2018). Mental health issues could impair an adolescent's ability to understand and participate in FFT, so participants' mental health is a critical characteristic that warrants particular attention among adolescents with antisocial behaviour (Kang et al., 2018). Some research studies into the effectiveness of FFT have included measures of mental health and explored how mental health issues moderate FFT outcomes (e.g., Slesnick & Prestopnik, 2009).

## Research Outcome Measures

The literature on FFT to date has mostly focus on the impact of FFT on two main outcome measures: recidivism and family functioning (e.g., Alexander & Parsons, 1973; Parsons & Alexander, 1973). Recidivism refers to an individual's relapse into criminal behaviour after previous offending, while family functioning within the literature refers to a wide range of domains, such as those in the Family Environment Scale (Moos & Moos, 1986). This includes the use of adaptive communication styles, the level of support family members provides each other, the level of encouragement to express feelings, and the amount of conflict and verbal aggression (Moos & Moos, 1986). Family Functioning can also include the extent to which family members personal growth is encouraged, as well as the organisation of family activities and responsibilities (Stratton et al., 2010). The program developers' early studies first reported significantly reduced participant recidivism rates of 26%, compared to families receiving two alternate family therapies: client centred (47%) and psychodynamic (73%), and no treatment (50%; Alexander & Parsons, 1973). A later analysis of the siblings from this intervention, also used recidivism as a measure of the program's success and reported the rate of sibling contact with the court about three years later was 40% for the FFT families, and 59-63% for the comparison groups (Klein et al., 1977). However, these early analyses had significant methodological problems. The use of the unique demographic of mostly families of the Morman religion raises concern about the generalisability of these findings as the culture and class of the Mormon community may have been a factor in the successful outcomes presented (Gordon et al., 1988). There was also no mention of the mental health of the adolescents or their family members, which may have moderated the results. Later studies continued to use recidivism and family functioning, but also began to include mental health related variables with participants with a wider variety of characteristics.

## FFT and Mental Health Outcomes

In the early 2000's, studies into the effects of FFT began to include outcome variables related to mental health (e.g., Waldron et al., 2001). Mental health domains were included in Slesnick and Prestopnik's (2009) analysis including mood, anxiety, social problems, CD, and ODD. In a comparison of FFT to another family therapy, participants in both programs improved over time on measures of psychological functioning. An analysis of the role of callous-unemotional traits (CU) often found in young people with CD and juvenile offenders, reported that FFT led to greater improvements in conduct problems for youths with CU traits compared to the adolescents without these traits (White et al., 2013). In contrast, Thøgersen et al., (2022) used a randomised control group to examine the relationship between CU traits and treatment outcomes with Norwegian adolescents and found no difference between FFT and the comparable Multi-Systemic Therapy (MST). Both groups showed large reductions in aggressive and rule breaking behaviour, and increased prosocial behaviour, and family cohesion. More recent studies have used standardised assessment tools to measure mental health related outcomes.

## Mental Health Outcome Assessment Tools

Recent researchers have examined improvements in a range of mental health domains in FFT participants using psychometrically sound measures. Graham et al., (2014) used the the Strength and Difficulties Questionnaire (SDQ; Goodman, 2001), to measure clinical improvement in emotional and behavioural functioning and found those who completed FFT showed a significant improvement in all but one of the five SDQ domains from baseline to discharge. Clinical recovery rates using the SDQ clinical cut-off criterion were approximately 40%. Although there was no control group so it could be argued the changes were due to developmental factors. In contrast, Hartnett et al. (2016) found clinically significant improvement on the SDQ compared with a control group. From Baseline to 20 weeks, 50%

of youth in the FFT condition moved from clinical or high-risk to non-clinical or low-risk range on the SDQ, compared with 18.2% in a waitlist control group and these effects were maintained at 3-month follow-up. Others have also reported significant pre-post improvements on most of the SDQ domains (Graham, et al., 2014; Marshall et al., 2018).

The Strengths and Needs Assessment (SNA; Lyons, 2009) has been used to assess behavioural/emotional needs, and risk behaviours (Celinska et al., 2013, 2019; Celinska & Cheng, 2017) as well as the Outcome Questionnaire (OQ; Gan et al., 2021) with mixed success. Some have used participants with clinical mental health diagnoses (ODD and CD; Humayun et al., 2017) and found significant overall reduction in CD and ODD symptoms at 18-month follow-up for both FFT and the justice system treatment as usual (TAU) group. Other examinations of participants with CD found no differences between FFT and Cognitive Behaviour Therapy (CBT) on recidivism two years post intervention (Van der Put et al., 2013). This growing body of research into FFT and mental health outcomes has, to date, produced mixed results.

## Previous Reviews into FFT outcomes

There have been a variety of reviews and meta-analyses into FFT's effectiveness in reducing adolescent antisocial behaviour (e.g., Hartnett et al., 2017). However, few have examined mental health issues in their analysis. Weisman and Montgomery (2019) reported on recidivism and substance abuse in their overview of 31 reviews and concluded there was insufficient quality evidence to make conclusions about FFT. In another review of 14 studies into FFT, Hartnett et al. (2017) reported that FFT was more effective than alternate treatments or no-treatment, but not more effective than TAU for recidivism, family functioning and behavioural problems. The most recent review and meta-analysis of 20 studies also evaluated recidivism, family functioning, behaviour problems, and substance use (Littell et al., 2023). These authors also evaluated peer relations, prosocial behaviour, self-

esteem, school attendance, and school performance. However, with regard to mental health issues, their evaluations were collapsed into one category: internalising behaviour problems. This refers to negative behaviours focussed inward such as social withdrawal and somatic complaints. To support such widespread use of FFT in a population known to have 50-80% of its members meet the diagnostic criteria for a mental health issue, it is especially important to review the current body of evidence to determine whether an intervention designed to treat antisocial behaviours could also be effective in reducing mental health issues.

### Aims and Research Questions

This systematic literature review primarily examined the effectiveness of FFT on the mental health of adolescents who engage in antisocial behaviour and have comorbid mental health issues or diagnoses. This review identified, selected, critically evaluated research, and synthesised findings into FFT to determine if FFT results in improvements in the mental health in adolescents who engage in antisocial behaviours and have comorbid mental health issues or diagnoses. The secondary aim was to assess the effectiveness of FFT in reducing recidivism and improving family functioning. The review compared studies regarding risk of bias and provided a comprehensive and up-to-date account of the effectiveness of FFT as a therapeutic utility for improving the mental health of adolescents who engage in antisocial behaviour.

The following research questions were asked:

1. Is FFT effective in improving mental health issues in adolescents with antisocial behaviour and comorbid mental health issues or diagnoses.
2. Is FFT effective in reducing recidivism and improving family functioning?

## Method

### Design

The protocol for this review was registered with the International Prospective Register for Systematic Reviews (PROSPERO, **http://www.crd.york.ac.uk/PROSPERO**, registration number: CRD42023451290, see Appendix B). This review was conducted using a systematic search strategy to identify outcomes for adolescents and their families who have participated in FFT. This review had a specific focus on FFT treatment outcomes related to mental health symptom reduction. The review also aimed to ascertain whether FFT is effective in reducing adolescent antisocial behaviour, specifically recidivism and family functioning. To identify literature related to FFT for families with an adolescent who engages in antisocial behaviour, a systematic approach was undertaken, guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2009), see Figure 1.

## Search Strategy

The following databases were searched during July 2023: Ovid hosted databases (APA PsycArticles, APA PsycINFO, Medline, and Embase); EBSCO hosted databases (Academic Search Ultimate, Humanities Source Ultimate, Psychology and Behavioral Sciences Collection); Proquest hosted databases (Social sciences, Psychology, and Criminal justice); as well as Scopus. The following key terms were used to search within the title and abstract fields: FUNCTIONAL FAMILY THERAPY and TEENAGER OR TEEN OR YOUNG PERSON OR YOUTH OR ADOLESCEN*. Mental health issues were not included as search terms, in order to capture all of the nuanced references to a wide range of mental health related variables in the literature. Only peer reviewed documents were included in the search. The references from selected articles were also assessed for eligibility in the review process. All the potential articles were evaluated by two independent reviewers through systematic approaches involving the inclusion and exclusion criteria.

## Inclusion/Exclusion Criteria

Studies were included if they:

1) Described an FFT intervention applied to the broad range of antisocial behaviours.
2) Were published in English in a peer reviewed academic journal.
3) Distinctly used an FFT therapeutic technique.
4) Were delivered to adolescents with any type of characteristics, including those with mental health issues or diagnoses.

Studies were excluded if they:

1) Described a hypothetical or non FFT intervention.
2) Described an intervention that is a variation of FFT such as FFT-CW (child welfare).
3) Were delivered to a population younger than 11 years of age or older than 18 years of age.
4) Were conference abstracts or unpublished work such as doctoral theses.

Data Extraction

Data was extracted from the included articles regarding study location, design, sample size, age range, characteristics, ethnicity, sex, comparison group, outcome measures, and results by the first author. The data was organised using a literature table (see Table 1). Studies were organised by date and findings regarding adolescent mental health outcomes.

Quality Appraisal

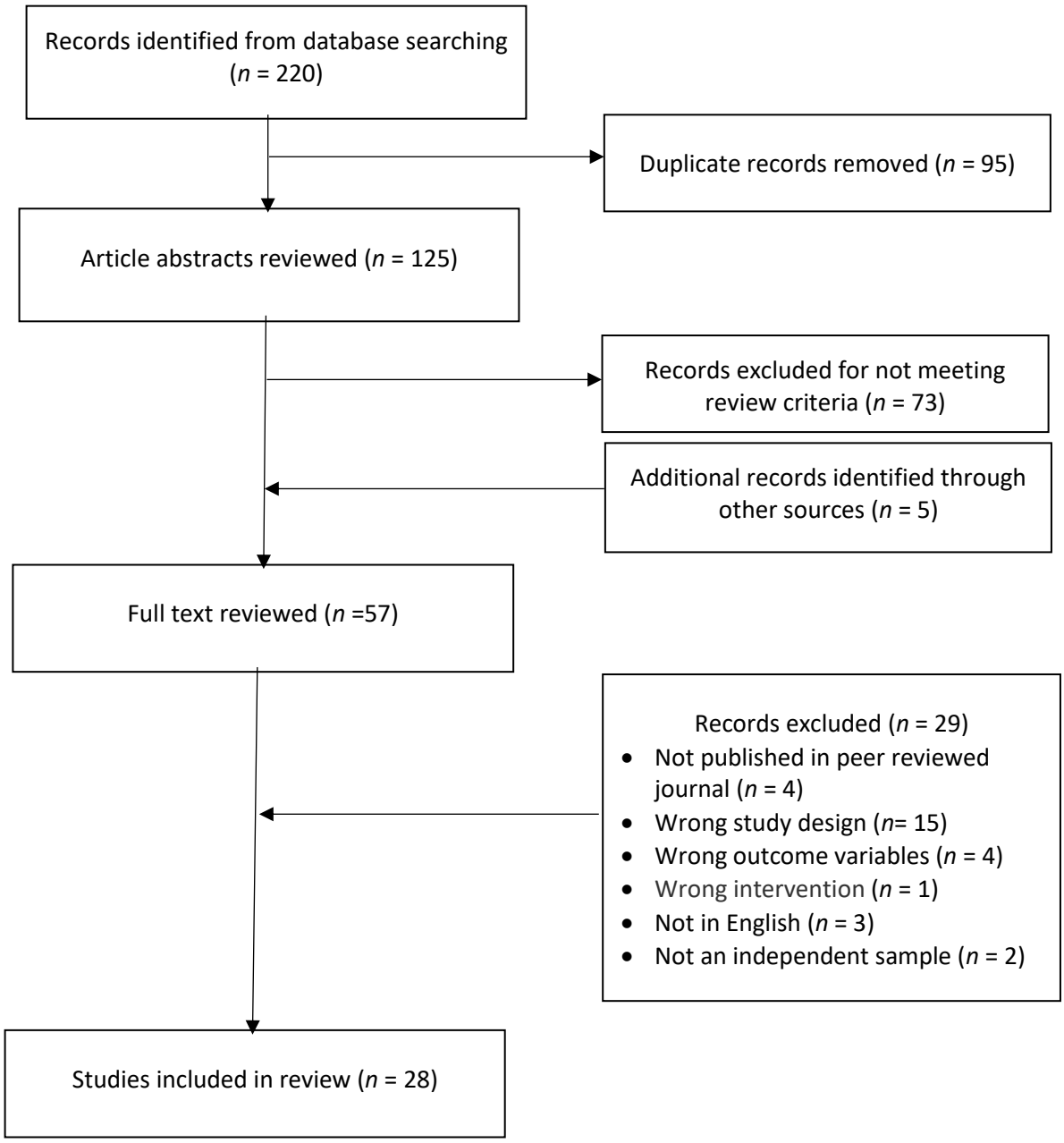
The first author assessed risk of bias for randomised controlled trials (RCT) using the Revised Cochrane Risk-of-Bias Tool for Randomised Trials (RoB 2; Higgins et al., 2011; Appendix C). Risk of bias for non-randomised studies was assessed by the first and second authors with the Risk of Bias in Non-Randomised Studies – of Interventions (ROBINS-I) Assessment Tool (Sterne et al., 2016; Appendix D). Any discrepancies between the two authors were discussed and agreement was reached.

## Results

A total of 42 potential studies were identified from 115 articles after the first round of screening. Full text review in the second round of screening resulted in 28 included articles for the final analysis. Figure 1 shows the study selection procedure in PRISMA format and Table 1 presents key data from the 28 included articles.

Figure 1

*PRISMA Flow Diagram of Process of Selected Studies*



## Sample Characteristics

### *Year and Country*

The 28 studies included in the review were published between 1973 and 2022. Most studies were conducted by researchers from the United States (*n* = 17, 61%), one study was published by Danish researchers (4%), and two (21%) each were undertaken in the Netherlands, Ireland, and Norway, with one (14%) each from New Zealand, England, Scotland and Singapore.

### *Study Design and Sample*

The most common study design type was RCTs (*n* = 12), followed by Quasi Experimental Design (QED) with a comparison group (*n* = 9), and without a comparison group (*n* = 7). The wide range of characteristics of participants are shown in Table 1. Adolescents were typically aged between 11 and 18 years old, and the majority were male in 20 of 28 studies. Sample sizes ranged from 21 to 2471, and there was significant variability in participant characteristics. Some were substance dependent (e.g., Slesnick & Prestopnik, 2009), criminal offenders ranging from mild to serious offences (e.g., Alexander & Parsons, 1973; White et al., 2013), and some had significant family problems such as aggressive or violent behaviour between family members (Hartnett, et al., 2016). One sample were residents in a psychiatric outpatient clinic (van der Put et al., 2013), and in an out of home placement (Darnell & Schuler, 2015), and in a runaway shelter for adolescents (Slesnick & Prestopnik, 2009), or with below average IQ (Humayun et al., 2017) or had anxiety and/or depression symptoms (Waldron et al., 2001). Nineteen studies reported on the ethnicity of participants, with the majority being Caucasian followed by Hispanic and African American.

Table 1

*Summary of Included Studies and Outcomes*

| Author/s, Year, Location | Design, Sample size | Program Length (PL), Follow up (FU) | Age (A), Sex (S) | Ethnicity | Sample Characteristics | Comparison Groups/s | Outcome Variable/s | Outcome Measure/s | Outcome/s | Risk of Bias |
|---|---|---|---|---|---|---|---|---|---|---|
| Alexander & Parsons (1973) USA | RCT *n*=86 | PL: 10-12 weeks FU: 6-18 months | A: 13-16 S: m=38 f=48 | Not stated | Juvenile Court: run away, shoplifting, ungovernable, truancy, possession of alcohol, soft drugs, tobacco | 1. Client-centred family group (CCFG) 2. Psychodynamic family treatment (PFT) 3. No Treatment | 1. Family communication interactions 2. Recidivism | 1. Observation of family interactions on 4 measures: silence, simultaneous speech frequency & duration, verbal reciprocity. 2. Juvenile Justice Records | 1. FFT: significant changes in interaction measures group. 2. FFT: significantly reduced recidivism rates at follow-up (26%) vs CCFG (47%), PFT (73%), No Treatment (50%). | High |
| Parsons & Alexander (1973) USA | RCT *n*=40 | PL: Not stated FU: Program end | A: *M*=15 S: m=18 f=22 | Not stated | Juvenile Court: run away, shoplifting, ungovernable, truancy, possession of alcohol, soft drugs, tobacco | 2 x treatment groups: a) pretest, FFT, post-test, b) 2. FFT, post-test. 2 x control: a) pretest post-test, b) psychologist led group discussion | Family communication interactions | Observation of family interactions on 4 measures: silence, simultaneous speech frequency & duration, verbal reciprocity | Treatment groups significantly different from untreated control group and discussion group. Significant changes in family interaction patterns: less silent, talked more equally, increase in frequency and duration of simultaneous speech. | High |
| Alexander, Barton, Schiaro, & Parsons (1976) USA | QED (pre-post) *n*=21 | PL: Not stated FU: 12-15 months | A: 13-16 S: m=10 f=11 | Not stated | Juvenile Court: run away, shoplifting, ungovernable, truancy | nil | 1. Rates of supportive and defensive communication 2. Recidivism | 1. Observations of family interactions 2. Juvenile Justice records | 1. Supportive/defensive communication ratios were significantly different pre-post 2. Recidivism rate: (23.8%). | Critical |
| Klein, Alexander, Parsons (1977) USA | RCT *n*=86* | PL: Not stated FU: 2.5 to 3.5 years | A: 13-16 S: m=38 f=48 | Not stated | *Same sample as Alexander & Parsons (1973) | *Same comparison group as Alexander & Parsons (1973) | Rate of sibling contact with court | Juvenile Justice records | Significant reduction in sibling court involvement: FFT: 40%, No treatment: 50%, Client-centered: 59%, Eclectic-dynamic: 63% | High |
| Gordon, Arbuthnot, Gustafson, & McGreen (1988) USA | QED *n*=54 | PL: M sessions= 16 FU: 28 months | A: *M*=15 S: m=38 f=16 | All Caucasian | 1/3 Out of home care, 2 offences/varying severity, probation visit, court referred, family conflict, 80% low income, 60% single parent | Probation visit, lower risk delinquents not court mandated, juvenile court, less serious offences, not placed outside home, no counselling. | Recidivism: Juvenile court contact | Juvenile Justice records | Recidivism: Intervention group 11% (follow up average 28 month), Comparison 67% (follow up average 32 months, 25%) | Critical |

| Author/s, Year, Location | Design, Sample size | Program Length (PL), Follow up (FU) | Age (A), Sex (S) | Ethnicity | Sample Characteristics | Comparison Groups/s | Outcome Variable/s | Outcome Measure/s | Outcome/s | Risk of Bias |
|---|---|---|---|---|---|---|---|---|---|---|
| Waldron Slesnick, Brody, Turner, & Peterson (2001) USA | RCT *n*=120 | PL: FFT: 12hr CBT: 12hr Grp: 12hr Joint: 24hr FU: 4 & 7 months | A: *M*=15 S: m=96 f=24 | Hispanic 47%, Anglo American 38%, Native American 8%, mixed/ other 8% | 43% Juvenile justice mandated, substance abuse disorder (marijuana) 90% delinquent, 30% anxious/ depressed, 27% attention difficulties, 48% externalising & 45% internalising behaviour. | 1. CBT (*n*=31)<br>2. FFT+CBT (*n*=29)<br>3. Psychoeducation (*n*=30) | 1. Percentage of days marijuana used & minimal use.<br>2. Percentage of youths-minimal use. | 1a. Form 90D version of the Timeline follow-back interview<br>1b. POSIT<br>1c. Urine screens | 1. Significant reductions in percentage of days of substance use for the FFT and FFT+CBT.<br>2. Significant numbers of youths achieved minimal-use levels in the FFT, FFT+CBT and CBT. | Some concerns |
| Slesnick & Prestopnik (2009) USA | RCT *n*=119 | FFT *M* sessions: 6.51 EBFT *M* sessions: 10.31 FU: 3, 9 & 15 months | A: 12-17 M=15 S: m=54 f=65 | Hispanic 44% Anglo 29% Native American 13% Other 13% African American 5% | All alcohol problem from runaway shelters. 89% diagnosed alcohol abuse, 66% marijuana abuse, 22% other substance abuse | 1. Home based EBFT (*n*=37; Ecologically-Based Family Therapy).<br>2. SAU (*n*=42) Service as Usual through the shelter - case management and individual therapy. Mean stay 17 days | 1. Substance use<br>2. Family functioning<br>3a. Delinquency, aggression, attention, somatic, thought, and social problems.<br>3b. Mood, cognitive and somatic aspects of depression<br>3c. Theft, crimes against persons, drug sales, delinquency | 1a. Form 90 for Substance Use<br>1b. POSIT<br>1c. Urine screens<br><br>2a. FES<br>2b. PBI<br>2c. CTS<br><br>3a. YSR/CBCL<br>3b. BDI<br>3c. NYSDS<br>3d. CDISC | 1. 15 months post baseline: EBFT: 97% decline days alcohol use (FFT: 83%, SAU 59%). EBFT: 77% reduction number drinks consumed (FFT: 64%, SAU: no change). Drug Use: both family therapies: 72% reduction at 15 months. SAU returned to baseline use levels.<br>2. All improved in family functioning (verbal aggression, family cohesion and conflict), psychological functioning (psychiatric diagnoses, externalizing problems, delinquent behaviours, and days living at home). 3. Internalising problems and depression changes varied by age: younger adolescents, internalising problems and depression decreased significantly for EBFT and FFT. SAU did not. | High |
| Sexton & Turner (2010) USA | RCT *n*=917 | PL: *M*=12 sessions, 3-6 months FU: 12 months | A: 13-17 S: m=79 % f=21% | White 78% African American 10% Asian 5% Native | Juvenile offenders on probation: Drugs 85%, Alcohol 81%, mental health, or behavioural problems 27%, serious crimes 56%, less serious | Usual probation services (PAU): 85% weekly checking and supervision, 15% education and guidance | 1. Recidivism<br>2. Therapist model adherence | 1. Juvenile Justice records<br>2. Supervisor ratings | 1. FFT intervention no more effective for recidivism than supervised probations services<br>2. FFT effective for high model adherent therapists: significant reduction of (35%) in felony, a (30%) violent crime, and a | High |

| Author/s, Year, Location | Design, Sample size | Program Length (PL), Follow up (FU) | Age (A), Sex (S) | Ethnicity | Sample Characteristics | Comparison Groups/s | Outcome Variable/s | Outcome Measure/s | Outcome/s | Risk of Bias |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | American 3% Not identified 4% | 42%, weapons 10%, gang 16%, out of home placements 11%, runaway 14%, school dropout 46%. | | | | marginally significant reduction (21%) in less serious recidivisms compared to the PAU. Low adherent therapists significantly higher than control group in recidivism rates. | |
| Celinska, Furrer, & Cheng (2013) USA | QED *n*=72 | FFT 3.4 months YCM 4.5 months | A: *M=15* S: m=47 f=25 | Treatment group: 36% African American 36% Latino. | Probation mandated 42%, family crisis 25%, family court 14%, youth services 8%, history aggression, property damage, chronic truancy. | Youth Case Managers individual therapy or mentoring. 44% African American, 33% Latino | 1. Risk and protective factors. 2. Life domains functioning | 1 & 2. SNA | 1. FFT: significant reduction risk behaviour, more domains, improvement-life domains: living situation, school behaviour, achievement, attendance, legal, vocational concerns, child behavioural/ emotional needs. 2. Both: significant improvement life domain, child strengths, and child risk behaviours scales. No change from pre- to post on acculturation, caregiver strengths, caregiver needs. | Critical |
| van der Put, Asscher, Stams, van der Laan, Breuk, Jongman, & Doreleijers (2013) Netherlands | QED *n*=241 | PL: Not stated FU: 2 years | A: 13-21 years (*M*=17 years) S: m=207 f=34 | Non-Dutch (ethnic minority) group) 48% | Forensic-psychiatric outpatient clinic. Average 4 previous offences. Conduct disorder & justice system. Multiple psychiatric disorders. | 1. CBT 2. CBT + parent training | Recidivism | Juvenile Justice records | No significant differences in recidivism between all groups 2 years post intervention. | Serious |
| White, Frick, Lawing, & Bauer (2013) USA | QED (pre-post) *n*=134 | PL: *M*=10. 1 to 19 sessions FU: Completion, 6 & 12 months | A: 11-17 *M*=15 S: m=96 f=38 | African American 59% European-American 35% Hispanic 4.5% No data 4% | Adolescent juvenile offender court diversion program: minor offences 48%, violent 22%, property 19%, drug 6%, no offending data 5% | Nil | 1. Emotional, behavioural, & social functioning. 2. Recidivism | 1a. ICU 1b. BASC 1c. COM/TOM 2. Juvenile Justice records | 1. No significant improvements on emotional or social adjustment (Emotional Symptoms Index, Relationship with Parents, and Interpersonal Relations). Significant improvements in parent reported BASC Aggression and Conduct Problems subscales. CU traits associated with greater improvements in | Critical |

| Author/s, Year, Location | Design, Sample size | Program Length (PL), Follow up (FU) | Age (A), Sex (S) | Ethnicity | Sample Characteristics | Comparison Groups/s | Outcome Variable/s | Outcome Measure/s | Outcome/s | Risk of Bias |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | adjustment- ended treatment with more severe adjustment problems.<br>2. Association between CU traits and risk for violent charges decreased after treatment at 6- and 12-month follow-ups. | |
| Graham, Carr, Rooney, Sexton, & Wilson Satterfield (2014) Ireland | QED (pre-post) *n*=118 | PL: M=17 sessions, 3–6 months FU: 9 & 46 months (*M*=23 months) | *A: M* =14 S: m=70 f=48 | Not stated | Court Diversion Behavioural problems: school & family relationship difficulties, aggressive behaviour, substance use, self-harm. Socially disadvantaged suburb | nil | 1. Conduct problems, hyperactivity, emotional symptoms, peer problems, prosocial behaviour.<br>2. Therapist adherence | 1. SDQ<br>2. Therapist Adherence Measure | 1. Significant improvement in group mean scores from intake to discharge on all but one SDQ domain (except peer problems). SDQ 40% clinical recovery rates.<br>2. High adherent therapists clinical recovery rates 60% for therapy completers. | Critical |
| Rohde, Waldron, Turner, Brody, & Jorgensen (2014) USA | RCT *n*=170 | PL: 12 sessions over 10 weeks FU: 6 & 12 months | A: 13-18 S: f=37 m=133 | 61% non-Hispanic White | Major depressive disorder 54%, dysthymia 18%, substance use disorders. Cannabis: abuse 21%, dependence 73%, alcohol: abuse 34%, dependence 31% | 1. FFT then Coping with Depression (FFT/CWD). 2. CWD then FFT (CWD/FFT). 3. Coordinated FFT and CWD (CT). | 1. Substance use<br>2. Depression | 1. TLFB<br>2a. CDRS-R<br>2b. K-SADS-PL | 1. FFT/CWD: better substance use outcomes than CT at posttreatment, and 6- and 12-month follow-ups. CWD/FFT lower substance use than either FFT/CWD or CT for MDD.<br>2. 3 groups: depressive symptoms decreased significantly. Depression remission during 40%, 1 year post 60% | High |
| Baglivio, Jackowski, Greenwald, & Wolff (2014) USA | QED *n*=2,203 | PL: MST Mean=119 days FFT Mean=95 days FU: 1 year | A: *M*=15 S: m=1593 f=610 | Non white 1007, white 1196 | FFT (*n*=1,574) 917 Alcohol use 1399 Drug Use | MST (*n*=629) | Recidivism | Juvenile Justice records | Both significant improvements except: FFT females lower recidivism rate, low-risk FFT fewer offences during program. Higher risk MST higher recidivism rates than higher risk FFT youth. | Serious |
| Darnell & Schuler (2015) USA | QED *n*=524 | PL: *M* FFT sessions | 11–18 m=403 | 60% Latino, 30% | Recently released from court-ordered out-of-home | 1. FFP (n = 216)<br>2. FFT + FFP (n = 539) | Subsequent out-of-home placement | Juvenile Justice records | 36 months: no significant differences in OHP between 4 groups. OHP: comparison | Critical |

| Author/s, Year, Location | Design, Sample size | Program Length (PL), Follow up (FU) | Age (A), Sex (S) | Ethnicity | Sample Characteristics | Comparison Groups/s | Outcome Variable/s | Outcome Measure/s | Outcome/s | Risk of Bias |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 9.1. *M* time to complete 4.2 months. FU: 36 months | f=121 | African American, less than 10% White. | placement (OHP) for family risk, maltreatment history, child behavioural health needs. | 3. Comparison (*n*=7,434): released from placement prior to interventions | (camp/JJ facility) | | =36%, FFT=34%, FFP=39%, FFT+FFP= 39%. OHP significantly lower for FFT & FFT+FFP compared to comparison first 2 months post release, but not at 3rd month (OHP similar to comparison youth) | |
| Hartnett, Carr, & Sexton (2016) Ireland | RCT *n*=97 | PL: 20 sessions 4–6 months FU: 3 months | A: *M*=14 S: m=60 f=37 | Most Irish | Half single parent, half parents unemployed, half low SES, behavioural problems, high risk mental health disorders. SDQ & SCORE clinical range | Waitlist (*n*=55) | 1. Adolescent behaviour problems and risk of mental health disorder 2. Family functioning | 1. SDQ 2. SCORE | 1. FFT Clinically significant improvement compared to controls. Baseline to 20 weeks, 50% in SDQ nonclinical or low-risk range (controls 18.2%). Parent SDQ (not self): significant reduction in risk of mental health problems. Post treatment effects maintained at a 3-month follow-up. 2. Significant improvements in family adjustment compared to controls. Significant reductions in severity and impact of main target problems. | Some concerns |
| Heywood & Fergusson (2016) New Zealand | QED (pre-post) *n*=59 | PL: 10.4 Sessions range: 1 to 31, *M*=10 FU: baseline, 6 & 12 months | A: 9-16, *M*=14 S: m=43 f=43 | Māori 55%, NZ Euro 33%, European 7%, Tongan 3%, Fijian 4% | Conduct problems, social/economic disadvantage, 78% parent high school only, 62% social welfare benefit, 69% sole parent, third of median household NZ income | Nil | 1. CD & ODD behaviours 2. Drug and Alcohol use 3. Delinquency behaviours | 1. CD and ODD. DSM criteria 2. Parent report AOD type & frequency previous two months, 3. SRD | 1. Significant reduction in Parent and teacher reported CD & OD behaviours. Significant reduction in two of five conduct problem measures at follow-up. Māori vs. non-Māori no difference between changes in conduct problems over time 2. No significant change parent reported AOD 3. Significant difference in parent reported delinquent behaviours from baseline to 12 months | Serious |
| Celinska & Cheng (2017) | QED (pre-post) | PL: 179 days, *M* days | A:11-17 | 35% (41) Black; 28% (33) | family court mandated, aggression, property damage, | Nil | 1. Life Domain Functioning 2. Recidivism | 1. SNA 2. Juvenile Justice records | 1. Both sexes improved significantly on life domain functioning, child behavioural | Serious |

| Author/s, Year, Location | Design, Sample size | Program Length (PL), Follow up (FU) | Age (A), Sex (S) | Ethnicity | Sample Characteristics | Comparison Groups/s | Outcome Variable/s | Outcome Measure/s | Outcome/s | Risk of Bias |
|---|---|---|---|---|---|---|---|---|---|---|
| USA | *n*=116 | attended: male 12, female 10 | S: m=72 f=44 | White; 25% (29) Latino | chronic truancy. Females less serious charges. Males mandated on probation | | | | emotional needs, child strengths, and child risk behaviour scales.<br>2. Number of convictions significantly decreased, and number of institutionalizations significantly increased for males. | |
| Humayun, Herlitz, Chesnokov, Doolan, Landau, & Scott (2017) England | RCT *n*=111 | PL: 12 sessions, 3 - 6 months. FU: Baseline, 6, & 18 months | 10-18 *M*=15 m=77 f=34 | White British (90%) | 67% youth offending services, 22% services for antisocial, 11% other crime agency, below average IQ (*M*=84). 55% single parent, 57% unemployed, 69% parents' education up to 16yrs | 1. FFT+MAU (*n*=65)<br>2. MAU (*n*=46).<br>MAU: case worker support, counselling, education, anger, employment, AOD sexual health, social skills, victim awareness programs | 1. Recidivism<br>2. ODD/CD symptoms<br>3. Parent-youth interactions/relationship | 1. Juvenile Justice records<br>2. APACS<br>3. APQ | 1. No significant differences between FFT+MAU and MAU at 6- or 18-month follow-up on any measure of antisocial behaviour: self-report delinquency, parent investigator-rated interview, directly observed child negativity or police records. MAU better directly observed positive interaction with parent at 6 months. Baseline severity of SRD did not moderate the effect of treatment.<br>2. No significant differences between groups at 6- or 18-months follow-up in either CD or ODD symptoms or diagnoses. Both groups: significant overall reduction in CD and ODD symptoms and diagnosis between baseline and 6- and 18-months follow-up.<br>3. No differences between groups on family functioning, no significant changes over time in parenting behaviour or the parent-child relationship | High |
| Eeren, Goossens, Scholte, Busschbach, & Van der | QED *n*=697 | PL: MST 149 days FFT 196 days | *M*=15 m=62 % m=27 5 | Not stated | *n*=275 54% male, 95.8% Netherlands born | MST (*n*=422) MST, 67% male, 83% Netherlands born<br><br>*With court order* | 1. Problem behaviour<br>2. Living at home, engaged in school or | 1. Parents CBCL<br>2. YSR<br>3. NOSI | 1. No differences in externalizing problem behaviour. | Critical |

| Author/s, Year, Location | Design, Sample size | Program Length (PL), Follow up (FU) | Age (A), Sex (S) | Ethnicity | Sample Characteristics | Comparison Groups/s | Outcome Variable/s | Outcome Measure/s | Outcome/s | Risk of Bias |
|---|---|---|---|---|---|---|---|---|---|---|
| Rijken (2018) Netherlands | | FU: not stated | f=134 | | | *(higher risk)-more often assigned to MST *Propensity score method could not balance groups* | work, and police contact during treatment 3. Parenting Stress | | 2. MST more often engaged in school or work post intervention. No court order group: MST more effective on externalising problems * | |
| Marshall, Hamilton, & Cairns (2018) Scotland | QED *n*=164 | PL: Not stated FU: Not stated | A: 11-16 *M*=16 S: not stated | Not stated | Internalising and externalising behaviour problems. At risk of removal from family | nil | 1. emotional distress, behavioural difficulties, hyperactivity/ inattention, peer problems, prosocial behaviour 2. overall functioning: symptom distress, interpersonal relations, social roles. 3. General well-being, problem behaviour, communication, conflicts. | 1. SDQ Added Value Score 2. COM (P&A) | 1. Significant decrease psychosocial distress scores on all measures from pre to post. Self mean SDQ scores in all domains below clinical cut-off. Significant decreases: total difficulties, emotional distress, behavioural and hyperactivity/ attention difficulties, peer problems. Significant increase in parent & self-rated prosocial behaviour. Significant reduction self-reported psychosocial stress: emotional distress, behavioural problems attention difficulties. 2. Some/lot better parent rating: family change & communication skills 93%, adolescent problem behaviour 88%, family conflict 93%. Some/lot better self-rated: family change 93%, communication skills 86%, problem behaviour 84%, family conflict 87%. | Critical |
| Kretschmar, Tossone, Butcher, & Marsh (2018) USA | QED *n*=530 | PL: Not stated FU: 12 months: Between 18 and 19 years of age. | *M*=21 end Data Collection m=214 f=316 | White 8%, Non white 52%, Black 68% Multi racial 5% Asian 3% | Juvenile Justice diversion program for behavioural health issues *M*=3 charges. | 1. Suitable but did not enrol (Group A, *n*=120) 2. Suitable, enrolled, failed to complete (Group B, *n*=223) . | 1. Early adulthood Offending 2. Time to first adult charge 3. Time to recurrent early adulthood charges | 1. Justice system records 2. Days from 18th birthday to day of offending 3. Days from previous charge to day of the new charge measured from 18th birthday. | Aged 19: 68% never charged, 32% 1-16 charge. FFT highest amount never charged (76%) and lowest percentage of 3 or more charges (10%), and lower odds of early adulthood recidivism compared to non-starters and non-completers. Non-starters and non- | Critical |

| Author/s, Year, Location | Design, Sample size | Program Length (PL), Follow up (FU) | Age (A), Sex (S) | Ethnicity | Sample Characteristics | Comparison Groups/s | Outcome Variable/s | Outcome Measure/s | Outcome/s | Risk of Bias |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Not stated 26% | | | | | completers: no significant difference odds of early adulthood recidivism. FFT completers with new early adulthood charge, took >30 extra days compared to both other groups. | |
| Celinska, Sung, Kim, & Valdimarsdottir (2019) USA | QED *n*=155 | PL: FU: 12 months | 11 to 17 *M*=15.5 m=87 f=68 | Majority White: 88 Black: 44 Asian: 6 Other: 17 | *n*=107. Juvenile offenders. Court supervision. Trauma history 25, mental health issues 50, Convictions:2. Risk delinquency behaviour, aggression people and/or animals, property damage, truancy, theft and deceit. | *n*=48. Youth Case Management (YCM): referrals to community service providers: 45, individual therapy Trauma history 27%, mental health issues: 24, 50% Prior convictions: 1.73 | 1. Family functioning resilience 2. Recidivism | 1. SNA 2. Juvenile justice system records | 1. Both groups significantly improved in 5/7 domains: life domain functioning, child strengths, caregiver strengths, child behavioural/emotional needs, and child risk behaviours. 2. Both groups improved significantly in delinquency behaviour. FFT group significant reduced likelihood of reconvictions for drug & property offences. | Critical |
| Robbins, Waldron, Turner, Brody, Hops, & Ozechowski (2019) USA | QED *n=164* | PL: Not stated FU: Baseline, 5 & 12 months | A: 11-18 *M*=15 S: m=97 f=67 | Hispanic 62%, White 12%, African American 19%, Native American 4%, other 3%. | n=105. Behaviour problems (e.g., delinquency, substance use). Standard therapist supervision as usual (SAU): 2 hours of supervision per week (1 group, 1 individual). | Building Outcomes with Observation-Based Supervision of Therapy (BOOST): audio recording of all therapy sessions with review all recordings and weekly supervision meetings. | 1. Externalising /internalising problem behaviour 2. Effects of family functioning changes on behaviours | 1. CBCL/YSR 2. FES | 1. Both significant improvements on majority of variables. Both groups were effective in improving, externalising, internalising & offending behaviours. 2. Parents and youth in both conditions reported improvements in family functioning | Moderate |
| Vardanian, Scavenius, Granski, & Chacko (2020) Denmark | QED *n*=576 | PL: M=17 x 75-minute sessions FU: Not stated | 11-18 *M*=14 m=281 f=295 | Not stated | Child welfare services, social worker contact with parent self-referral, and school teachers or school psychologists. Moderate to severe behaviour problems (truancy, verbal | nil | 1a. Mental health 1b. CU traits 2. Family functioning. 3a. School attendance/ performance | 1a. SDQ 1b. ICU 2. SCORE 3. Adolescents/parents reports | 1. Significant improvements on all SDQ scores for both parents' and adolescents. Only adolescents' self-evaluation of prosocial skills insignificant. Significant improvement in Parents reports of adolescents ICU traits. | Critical |

| Author/s, Year, Location | Design, Sample size | Program Length (PL), Follow up (FU) | Age (A), Sex (S) | Ethnicity | Sample Characteristics | Comparison Groups/s | Outcome Variable/s | Outcome Measure/s | Outcome/s | Risk of Bias |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | aggression, violence, criminal behaviour, drug use) | | 3b. Substance use | | 2. Significant improvements on all SCORE-15 scales for parents' and adolescents' evaluations of family functioning. | |
| Gan, Zhou, Abdul Wahab, Ruby, & Hoo (2021) Singapore | RCT *n*=120 | PL: *M*=12 sessions over 4.7 months FU: Post & probation end | A: 13-18 *M*=16 S: m=107 f=13 | Not stated | On probation. 90% public housing. 45.8% low-income/ social support. Baseline YOQSR2.0 scores at or above the clinical: FFT 34%, TAU 49% | TAU. Probation services attending programs addressing offense or family needs | 1. Mental well-being 2. Probation completion. 3. Adolescents perceived family functioning | 1. YOQSR2.0 2. Juvenile Justice records 3. MFAD: General Functioning subscale | 1. FFT only: higher levels of well-being to probation end. YOQ clinical recovery: FFT 30%, TAU 16%. FFT 43%, TAU 18%. Significant association between treatment group and clinical recovery rates in YOQSR2.0 scores. 2. FFT significantly more likely to complete probation successfully. Probation completion: FFT 89%, TAU 70%. 3. No group differences in family functioning scores over time. Significant trend for FFT higher rates change and clinical recovery family functioning scale. | Some concerns |
| Hukkelberg, Ogden, & Thøgersen (2022) Norway | RCT *n*=2471 | PL: *M*=159 days FU: Baseline post & 6 & 12-month | A: 11-19 *M*=15 S: m=1497 f=992 | Norwegian 2128, (86%) Immigrant 195 (8%) | *n*=453. Serious problem behaviours referred by Child Welfare. **FFT significant more treatment days than MST.* | MST (*n*=2018) Larger % of males and higher levels of YLS/CMI risk factors | Criminal risk factors: 1. live at home, 2. attends school or work (min 50%), 3. No violence/ threats, 4. law-abiding, 5. drug-free. | Youth Level of Service/Case Management Inventory: 1. prior/current offenses/dispositions, 2. family circumstances/ parenting, 3. education/ employment, 4. peer relations, 5. substance abuse, 6. leisure/recreation, 7. personality/ behaviour, 8. attitudes/ orientation/beliefs | Both groups behavioural problems and YLS/CMI risk factors all significantly reduced after completion and remained low at follow-up, compared to baseline. Assessments of YLS/CMI predicted youths' achievement of the ultimate treatment goal at post-test and 6- and 12-month follow-up. | High |
| Thøgersen, Elmose, Viding, McCrory, & Bjørnebekk | RCT *n*=159 | PL: 3–6 months FU: Baseline, | A: 11-19 *M*=15 S: m=86 | 130 (82%) Parents born in Norway. 28 (18%) | Risk of delinquency, aggression/violence, threats, truancy, school problem behaviour, drug use. | Family counselling service + other treatment: MST 29 (40.3%), Parent. Training/ Guidance 2 | 1a. CU traits 1b. Impulsivity 1c. Emotional hypo or hyperreactivity | 1a. ICU (Parent) 1b. SNAP-IV (Parent) 1c. Parent CBCL: anxious-depressed subscale | 1. Improvements in all active treatment groups. CU traits related to large reductions in aggressive and rule breaking behaviour, and increases in | High |

| Author/s, Year, Location | Design, Sample size | Program Length (PL), Follow up (FU) | Age (A), Sex (S) | Ethnicity | Sample Characteristics | Comparison Groups/s | Outcome Variable/s | Outcome Measure/s | Outcome/s | Risk of Bias |
|---|---|---|---|---|---|---|---|---|---|---|
| (2022) Norway | | 6 & 18 months | f=73 | Parents minority/ immigrant backgroun d. 0.6% missing data | Mean salary slightly lower than average. 14% financial welfare support. | (2.8%), Child Mental Health Services 2. (2.8%), No Services 10 (13.8%) | 2a. Antisocial behaviours 2b. Prosocial behaviours 3. Parent–youth relationship quality | 2a. Parent CBCL: Rule-breaking and aggressive behaviour subscales 2b. SSRS 3a. FES 3b. IPPA | prosocial behaviour, parent reported family cohesion and perceived maternal support for both groups. CU traits did not negatively predict short- or long-term gains on measures in either treatment groups. | |

**PL=Program Length, FU=Follow up, A=Age, S=Sex, M = mean, m = male, f= female*

Table 2

*Assessment Tools and Acronyms*

| Assessment Tool | Acronym | Reference |
|---|---|---|
| Adolescent Parent Account of Child Symptoms | APACS | Taylor, Chadwick, Heptinstall, & Danckaerts, 1996 |
| Alabama Parenting Questionnaire, short version | APQ-15 | Elgar, Waschbusch, Dadds, & Sigvaldason, 2007 |
| Behavior Assessment Scale for Children, 2nd edition | BASC-2 | Reynolds & Kamphaus, 2006 |
| Child Behavior Checklist - Parents | CBCL | Achenbach and Rescorla 2001 |
| Children's Depression Rating Scale–Revised | CDRS–R | Poznanski & Mokros, 1995 |
| Client Outcome Measure - A | COM-A | Sexton & Alexander, 1999 |
| Client Outcome Measure - P | COM-P | Sexton & Alexander, 1999 |
| Computerized Diagnostic Interview Schedule for Children | CDISC | Shaffer, 1992 |
| Family Environment Scale | FES | Moos & Moos, 1986 |
| Form 90 | | Miller, 1996 |
| Multisystemic Therapy | MST | Henggeler et al., 1998 |
| Nijmeegse Ouderlijke Stress Index | NOSI-R | De Brock et al. 2004 |
| Outcome Questionnaire | OQ-45.2 | Lambert, Gregersen, & Burlingame, 2004 |
| Problem Oriented Screening Instrument for Teenagers | POSIT | McLaney, Del Boca, and Babor, 1994 |
| Prosocial Behaviors: parental responses to the Social Skills Rating Scale | SSRS | Gresham and Elliott, 1990 |
| Schedule for Affective Disorders and Schizophrenia for School-Age Children–Present and Life Version | K-SADS-PL | Kaufman, Birmaher, Brent, Rao, & Ryan, 1996 |
| Self Report Delinquency Inventory | SRD | Elliott & Huizinga, 1989 |
| Strengths and Needs Assessment | SNA | Lyons, 2009 |
| Swanson, Nolan, and Pelham—IV questionnaire | SNAPIV | Swanson et al., 2001 |
| The Beck Depression Inventory | BDI | Beck, Ward, Mendelson, Mock, & Erlbaugh, 1961 |
| The Child Behavior Checklist-The Youth Self-Report | YSR | Achenbach & Rescorla, 2001 |
| The Conflict Tactic Scale | CTS | Straus, 1979 |
| The Family Environment Scale | FES | Moos & Moos, 1986 |
| The General Functioning subscale of the McMaster Family Assessment Device | FAD | Epstein, Baldwin, & Bishop, 1983 |
| The Inventory of Callous–Unemotional Traits short version | ICU-12 | Hawes et al., 2014 |
| The Inventory of Parent and Peer Attachment-Revised | IPPA | Gullone and Robinson, 2005 |
| The National Youth Survey Delinquency Scale | NYSDS | Elliot, Huizinga, & Ageton, 1985 |
| The Parental Bonding Instrument | PBI | Parker, Tupling, & Brown, 1979 |
| The strengths and difficulties questionnaire | SDQ | Goodman, 2001 |
| The Systemic Clinical Outcome and Routine Evaluation | SCORE-15 | Stratton, Bland, Janes, & Lask, 2010 |
| Therapist Adherence Measure | TAM | Sexton et al., 2004 |
| Therapist Outcome Measure - P | TOM | Sexton & Alexander, 1999 |
| Timeline Followback Interview | TLFB | Miller & Del Boca, 1994 |
| Youth Level of Service/Case Management Inventory - Part I | YLS/CMI | Hoge & Andrews, 2011 |
| Youth Outcome Questionnaire Self-Report Version 2.0 | YOQSR2.0 | Wells, Burlingame, & Rose, 2003 |
| Youth Self-Report - Child Behavior Checklist | YSR/CBCL | Achenbach & Edelbrock, 1982 |

### *Mental Health Outcomes*

Overall, a wide range of mental health variables were examined in 17 studies. The studies used a range of assessment tool domains to explore these variables. The mental health related domains were explored using the SDQ (*n* = 4; Goodman, 1997, 2001), the Inventory of Callous–Unemotional Traits (ICU-12; *n* = 3; Hawes et al., 2014), the Child Behaviour Checklist (CBCL; *n* = 4; Achenbach and Rescorla 2001), the SNA (*n* = 3; Lyons, 2009), and the OQ-45.2 (*n* = 2; Lambert, Gregersen, & Burlingame, 2004). Other variables examined related to mental health included alcohol and other drugs (AOD; *n* = 5; Slesnick & Prestopnik, 2009; Waldron et al., 2001), depression (*n* = 2; Slesnick & Prestopnik, 2009; Rohde et al. 2015), and CD and ODD (*n* = 2; Heywood & Fergusson, 2016; Humayun, et al., 2017).

#### *Strengths and Difficulties Questionnnaire.*

The 25-item self-report SDQ was used to measure mental health variables in four studies. There are youth, parent, and teacher versions of the SDQ that measures positive and negative aspects of a child or adolescent's behaviour, emotions, and peer relationships. There are five subscales, each with items that assess emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behaviours. Participants respond to each item on a 3-point Likert scale which ranges from 0 ("Not True") to 2 ("Certainly True"). Each subscale scale score ranges from 0 to 10 and the total difficulty score ranges from 0 to 40. Higher scores indicate problem behaviour. The SDQ has been demonstrated to have acceptable reliability (Goodman, 2001) and convergent validity with other comparible measures (Hill & Hughes, 2007). Of the four studies that used the SDQ, all of them reported significant improvements from pre to post intervention on all SDQ domain scores for both parents' and adolescents' evaluations (Hartnett, et al., 2016; Marshall, et al., 2018), with the exception of peer problems (self-rated) in one study (Graham, et al., 2014)

and prosocial skills (self-rated) in another (Vardanian, et al., 2020). Only one of the studies used a comparison group and found FFT resulted in significantly immproved scores on the SDQ compared to a wait list control group (Hartnett et al., 2016).

***Inventory of Callous-Unemotional Traits.***

The ICU-12 (Hawes et al., 2014) was used in three studies to assess callousness, lack of guilt, empathy, and emotional expression. This 12-item parent-report scale was derived from the 24-item ICU (Frick et al., 2014). The total ICU contains a 7-item callousness subscale ("Other's feelings are unimportant to him/her") and a 5-item uncaring subscale ("Does things to make others feel good"). Participants rate each item on a 4-point Likert scale from 0 ("not at all true") to 4 ("definitely true"). Higher scores indicate more problematic traits. The 12-item scale demonstrated high internal consistency, good discrimination of the callous-unemotional construct, good test-retest reliability, and convergent and discriminant validity (Hawes et al., 2014). Only one of the studies who examined CU traits did so with comparison groups (Thøgersen, et al., 2022) and found improvements in aggressive and rule breaking behavior, and increases in prosocial behavior, parent reported family cohesion and perceived maternal support for FFT as well as a range of active treatment control conditions, including MST, parent training, and child mental health services. The other two studies who examined CU traits and FFT reported similar results. White et al. (2013) found particpants with CU traits improved in behavioural, emotional, and social adjustment and Vardanian et al. (2020) reported significant improvements on all parent reported scales of callous-unemotional traits.

***Child Behavior Checklist.***

The CBCL (Achenbach, 2009) was used in four studies to assess behavioural, social, and emotional problems within the preceeding six months (Eeren et al., 2018; Robbins et al., 2019; Slesnick & Prestopnik, 2009; Thøgersen et al., 2022). The CBCL has eight domains:

aggressive behaviour, anxious/depressed, attention problems, rule-breaking behaviour, somatic complaints, social problems, thought problems, and withdrawn/depressed. There are also three overall scales: internalising problems (anxious/depressed, withdrawn/ depressed, and somatic complaints), externalising problems (rule-breaking and aggressive behaviour) and a total problems score (all the problem items; Achenbach, 2009). All fours studies used comparison groups and reported improvements in their FFT and comparision groups . One study reported FFT, Ecologically-Based Family Therapy, and Services as Usual (SAU - runaway shelter case management and individual therapy) all reduced internalising problems and depression (Slesnick & Prestopnik, 2009). However this varied by age. For the younger adolescents, internalising problems and depression decreased significantly over time for both EBFT and FFT, while SAU did not. Another reported adolescents in both conditions showed significant improvements in externalizing and internalising behaviors (Robbins et al., 2019). One study reported only on externalising behaviours only and found no difference between FFT and MST (Eeren et al., 2018). The fourth study reported large reductions in aggressive and rule breaking behaviour, and large increases in prosocial behaviour, parent reported family cohesion and perceived maternal support for FFT and MST, parent training, and child mental health services (Thøgersen et al., 2022).

***Strengths and Needs Assessment.***

Three studies used the SNA (Lyons, 2009) to rate the strengths and needs of adolescents and their parents in seven domains: life domain functioning (vocation, school, and family life), child strengths (Family life, personal achievements and community involvement), acculturation (culture and language), caregiver strengths (stability in the home and involvement with the child), caregiver needs (physical and mental health), child behavioural/emotional needs (depression, anxiety, impulsivity, substance abuse, and anger), and child risk behaviours (danger to others, suicide risk, self-harm, running away, sexual

aggression, delinquency, and fire setting). Clinicians rated families on a scale ranging from 0 (no evidence of problem) to 3 (severe) with lower scores represented improvement (Lyons, 2009). Studies have suggested that the SNA has both validity and reliability (Lyons, 2009). Of the three studies who used the SNA, two used a comparison group and both found that FFT as well as case management with individual therapy and mentoring significantly improved in three domains: life domain functioning, child strengths, and child risk behaviours (Celinska et al., 2013; 2019). One study reported improvements in caregiver strengths, child behavioural/emotional needs (Celinska et al., 2019), and another reported that significant pre-post improvments across four SNA domains occurred for both males and females (Celinska & Cheng, 2017).

***Behavior Assessment Scale for Children.***

One study used the Behavior Assessment Scale for Children, 2nd edition (BASC-2; Reynolds & Kamphaus, 2006) to assess emotional and behavioural functioning and self-perceptions (White et al., 2013). There is extensive evidence to support the reliability of the BASC-2 in adolescent samples (Frick et al., 2010). In this study no significant improvements from pre-test to post-test were found for the indicators of emotional or social adjustment (Emotional Symptoms Index, Relationship with Parents, and Interpersonal Relations). However, significant improvements were found in both the parent reported Aggression subscale, and the Conduct Problems subscale (White et al., 2013).

***Youth Outcome Questionnaire.***

One study used the the Youth Outcome Questionnaire Self-Report Version 2.0 (YOQSR2.0; Wells et al., 2003) to measure psychosocial well-being (Gan et al., 2021). The 64-item questionnaire assesses behavioural and emotional problems in six domains: Intrapersonal Distress, Somatic Complaints, Interpersonal Relations, Social Problems, Behavioural Dysfunction, and Critical Items. Each item is scored on a 5-point scale with five

potential responses: “Never or Almost Never,” “Rarely,” “Sometimes,” “Frequently,” and “Always or Almost Always” (Wells et al., 2003). Each an answer is given a numerical score which are all added to give each domain a score as well as an overall distress score. Higher scores indicated lower levels of well-being (Wells et al., 2003). The YOQSR2.0 has good reliability and validity as well as appropriate sensitivity to change, and good internal consistency ($\alpha = 0.94$; Ridge et al., 2009). Gan et al. (2021) found the FFT group reported significantly higher levels of well-being immediately following the intervention and at the end of each participants probation period, compared to the TAU group receiving standard probation services. They also reported a significant association between FFT and clinicial recovery as measured by their YOQSR2.0 scores (Gan et al., 2021)

***AOD Use.***

Six studies explored AOD use as an outcome measure and again, there were wide variations within this variable. There was variation in the type of substance used (e.g., cannabis, alcohol) as well as in the range of ways use was measured. Some studies measured success with days of no use, or days with minimal use (Slesnick & Prestopnik 2009; Waldron, 2001). There were three studies who measured AOD use, and used a comparison group. One found both FFT and the comparison group significantly reduced use (Cannabis; Waldron et al., 2001). Another reported the comparison group had greater reduced use of alcohol but both groups decreased their cannabis use (Slesnick & Prestopnik, 2009). One found the best outcomes from FFT supplemented with a coping with depression course (Rhode et al., 2014). Of the studies who examined substance use with no comparison group, one found no change in AOD use from pre to post FFT (Heywood & Fergusson, 2016) and another had mixed results with significant increase in alcohol use and a signficant reduction in cannabis use (Vardanian, 2020).

### *Depression/Mood.*

Depression/Mood disorders were explored in two studies. One used the The Beck Depression Inventory (BDI; Beck et al., 1961; Slesnick & Prestopnik 2009) and the other used the Children's Depression Rating Scale–Revised (CDRS–R; Poznanski & Mokros, 1995), as well as the Schedule for Affective Disorders and Schizophrenia for School-Age Children–Present and Life Version (K-SADS-PL; Kaufman et al., 1996; Rhode et al., 2014). Both studies reported significant decreases in depression symptoms. Among alcohol using participants residing in a runaway shelter, only the younger adolescents (12-15 years) in both the FFT and EBFT groups (not the SAU group) had a significant decrease in depression (BDI) and internalising problems (Slesnick & Prestopnik, 2009). In the group with similar alcohol problems, as well as cannabis use problems and major depressive disorder (54%) or Dysthymia (18%), depressive symptoms decreased significantly in all three of their treatment groups by 40% during treatment and by 60% one year later (Rhode et al., 2014).

### *CD and ODD.*

Two studies examined CD and ODD: one with (Humayun et al., 2017) and one without a comparison group (Heywood & Fergusson, 2016). Humayun et al. (2017) reported both FFT and youth offending management as usual (MAU) resulted in significant reduction in CD and ODD symptoms that were maintained at six and 18 month follow-up. Heywood & Fergusson (2016) reported significant reductions in parent and teacher reported CD and ODD behaviours with two of five conduct problem measures maintained at six and 12 month follow-up.

## *Recidivism*

The most common outcome measured was recidivism ($n$ = 13), however, the definition, time frames and measurement of recidivism was inconsistent. Some studies used reconvictions for certain offences only (e.g., Celinska, et al., 2019) and others used arrest

(e.g., Baglivio et al., 2014), offense severity (e.g., Gordon et al., 1988), removal from the family home (e.g., Darnell & Schuler, 2015), or any contact at all with police (e.g., Eeren et al., 2018). Recidivism was also measured at different time frames across studies. Some considered recidivism at the conclusion of the intervention (e.g., Alexander & Parsons, 1973), after 12-15 months (e.g., Alexander et al., 1976), or 2.5 years post intervention. (e.g., Gordon et al., 1988). Of the 13 studies who measured some form of recidivism, 10 found significant reductions in recidivism rates pre to post intervention. Of the studies who examined recividism, 10 did so with a comparison group and of those, six found no significant differences between the groups.

#### *Recidividism and Mental Health.*

Of the 13 studies who explored recividism, there were five who also measured an outcome variable relating to mental health. One found no significant pre to post improvements on an emotional symptoms index (White et al., 2013). Four found significant improvements from pre to post intervention on a range of mental health variables. These include the SNA domains of life functioning, child behavioral emotional needs, child strengths, and child risk behavior scales (Celinksa, 2017; 2019), as well as CD and ODD symptoms or diagnoses (Humayan et al., 2017; Gan et al. 2021).

### *Family Functioning*

Family functioning ($n$ = 12) was also defined and measured in a variety of ways. Early studies focussed on family communication frequency, duration, periods of silence, and verbal reciprocity (e.g., Alexander & Parsons, 1973), while others used rates of supportive and defensive communication (e.g., Alexander, et al., 1976). More recent studies used assessment tools to measure family functioning such as the Family Environment Scale (Moos & Moos, 1986; Slesnick & Prestopnik 2009; Robbins, et al. 2019) or The Systemic Clinical Outcome and Routine Evaluation, 15-item version (SCORE-15; Stratton et al., 2010; Vardanian, et al.,

2020). Of the 12 studies who explored family functioning in some form, nine found significant improvements from pre to post intervention. Nine of the studies who explored family functioning used a comparison group and seven of those showed no significant differences between the groups.

### *Family Functioning and Mental Health.*

Of the 12 studies who explored family functioning, nine explored variables related to mental health and five did so with a comparison group. All nine studies reported significant positive change from pre to post intervention on their mental health variable/s and only one (Hartnett et al., 2016) reported FFT performed significantly better than the (waitlist) control group.

## Risk of Bias

Assessments of the risk of bias within the RCT studies with the ROB-2 tool included bias arising from the randomisation process; from deviations from intended interventions; from missing outcome data; from measurement of the outcome; and from selection of reported results (Higgins et al., 2011). An overall risk of bias rating as either Low; Some Concerns; or High was given. Low ratings indicate assessment of a low risk of bias for all domains. Some Concerns indicates there were some concerns in at least one domain, but not to be at high risk of bias for any domain and a High rating indicates risk of bias in at least one domain or some concerns for multiple domains in a way that substantially lowers confidence in the result (Higgins et al., 2011). Of the 12 studies that used a RCT design, three were rated as having Some Concerns, and nine were rated as having High risk of bias. Assessments of the risk of bias within the QED studies with the ROBINS tool included bias due to confounding domains, selection of participants, classification of interventions, deviations from intended interventions, missing data, measurement of outcomes, and selection of the reported result (Sterne et al., 2016). Of the nine QEDs with a comparison group, six were rated as Critical,

two Serious and one Moderate. Of the seven QEDs without a comparison group, five were rated as critical and two serious.

## Discussion

The primary aim of this study was to review the available literature to ascertain whether FFT resulted in improvments in mental health issues in adolescents with antisocial behaviours and comorbid mental health issues or diagnoses. The secondary aim was to ascertain whether FFT is effective for adolescents with antisocial behaviours in reducing recidivism and improving their family functioning. This study provided a comprehensive narrative of the impact of FFT on key mental health outcomes (e.g., CD, ODD, AOD use, CU traits, depression, emotional symptoms, intrapersonal distress, and cognitive problems). Results indicated that FFT was an efficacious treatment for improving mental health issues with significant pre-post-treatment improvements reported in most studies across a range of mental health related outcomes. Results also indicated that FFT was an efficacious treatment for reducing recidivism and improving family functioning for most of the adolescents. However, the majority of studies used a comparison group and also reported similar pre-post-treatment changes across a range of comparison group types. This suggests that while FFT was effective in improving mental health issues, reducing recidivism, and improving family functioning, so were the interventions to which it was being compared.

### Mental Health Outcomes

Most studies reported significant mental health symptom reduction from pre to post intervention, however again, FFT was as effective in this regard as EBFT (Slesnick & Prestopnik, 2009), Case worker support (Celinska et al., 2013; 2019: Humayun et al., 2017), MST (Eeran et al., 2018; Robbins et al., 2019), and family counselling services (Thøgersen., 2022). In contrast, Hartnett et al. (2016) found significantly better outcomes in the SDQ domains of emotional symptoms, conduct problems, hyperactivity/inattention, peer

relationship problems, and prosocial behaviours for the FFT group only. Half of the participants of the FFT group moved from clinical/high-risk to non-clinical or low-risk range on these domains, compared with 18.2% in the control group and this was maintained at a 3-month follow-up (Hartnett et al., 2016). However, FFT participants were compared to a waitlist control group who did not receive any intervention. Further, the participants were younger adolescents, with almost half from two parent households and almost all from the dominant Irish culture. Younger adolescents are known to engage in milder antisocial behaviours (Dishion & Patterson, 2015), and being separated from a biological parent is known to be a contributing factor to the development of antisocial behaviour (McGee, 2021). There is also known to be a significant disparity in behavioural outcomes between white and minority racial populations (Gregory et al., 2010).

Studies who examined the effect of FFT on AOD use and depression produced mixed results. A reduction in cannabis use was reported with greater frequency (Slesnick & Prestopnik, 2009; Vardanian et al., 2020; Waldron et al., 2001), compared to alcohol use, which either did not change (Heywood & Fergusson, 2016) or increased (Vardanian et al., 2020). This is in contrast to a 2017 systematic review of 14 studies by Hartnett et al. who found support for the effectiveness of FFT in reducing both cannabis and alcohol use.

One study in this review found the best outcomes for substance use reductions from FFT supplemented with a coping with depression course (Rhode et al., 2014). These authors found depressive symptoms decreased significantly by 40% during treatment and by 60% one year later (Rhode et al., 2014). Depressive symptoms were also significantly reduced among 12 to 15 year old alcohol using participants, however this was true for both the FFT and EBFT intervention (Slesnick & Prestopnik, 2009). Both FFT and the usual juvenile justice case management also resulted in CD and ODD symptom reduction (Humayun et al., 2017).

Other authors have also found the application of evidenced based strategies from Cognitive Behaviour Therapy (CBT) integrated with family therapy produced successful outcomes. In an RCT of CBT or family therapy, and combined CBT and family therapy for substance abusing adolescents, Waldron et al. (2001) found that significantly fewer days of use for the family therapy alone and the combined intervention. However, at 7 months follow up, it was the combined intervention adolescents who had significant reductions in percentage of days of use.

## Recidivism and Family Functioning

This review found that most of the FFT interventions were equally effective in reducing recidivism as usual probation services (Sexton & Turner, 2011), CBT (van der Put et al., 2013), MST (Balivio et al., 2014), and justice system case management (Humayun et al., 2017). A similar pattern emerged with regard to family functioning domains. There were significant improvements from pre to post intervention in FFT and EBFT (Slesnick & Prestopnik, 2009), and caseworker support (Celinska et al., 2019; Humayun et al., 2017; Robbins et al., 2019). Similarly, Littell et al. (2023) statistical analysis of pooled effects showed FFT was not consistently superior to active comparisons, however, in contrast to this review, Littell et al. did not find sufficient evidence of significant pre to post improvments.

Littell et al. (2023) also highlighted the substantial heterogeneity between studies with regard to definitions of recidivism and family functioning. Similarly, this review found most studies define recidivism and family functioning in a wide range of ways. Recidivism included arrest, conviction, truancy, or possession of tobacco, or any encounter with police which can skew outcomes and invalidate results (Weisman & Montgomery, 2019). Family functioning was also measured as cohesion, adaptability, resilience, or conflict. Further, the early studies that formed the enduring empirical foundation for FFT reported family functioning variables that are highly subjective and not based on valid peer-reviewed

measures such as supportive and defensive communication (Alexander, et al., 1976) or periods of silence (Alexander & Parsons, 1973). In addition to multiple definitions of outcomes, data came from a range of sources with varying validity and across multiple follow periods. Measurement of outcomes came from administrative justice systems and parent or adolescent self-reports, and the highly variable follow-up periods ranged from completion (e.g., Gan et al., 2021) to three years post intervention (Darnell et al. 2015). Some were followed up at individual milestones such as probation completion (Gan et al., 2021) or after their eighteenth birthday (Kretschmar et al., 2018). There were also different follow up times between groups within studies (Gan et al., 2021: Gordon et al., 1988). The impact of variations across a wide range of follow up periods makes comparison of results difficult in reviews and meta-analysis (Higgins et al., 2011). Longer follow up periods can also result in reduced participation which can compromise a study's validity (Higgins et al., 2011).

## Methodological Concerns

While these findings provide support for the efficacy of FFT for adolescents with mental health issues, there are significant concerns about the poor quality of the literature. In a recent systematic review and meta-analysis, Littell et al. (2023) found that the overall quality of evidence for FFT is weak, and they reported that the available evidence does not support claims that FFT has consistent, positive effects across studies. Similar to Littell et al. (2023), there were a range of significant methodological problems in the studies included in this review. These included conflict of interest, risk of bias, baseline differences, performance bias, detection bias, attrition bias, inconsistent follow up periods, selective reporting, inconsistent sources of data and problematic comparison conditions.

The most notable concerns with this body of research are conflicts of interest and risk of bias. Almost half of the studies included in this review had authors who are known to be associated with FFT in some capacity, which raises issues around conflict of interest. Studies

have reported that results of trials conducted by authors with a financial conflict of interest were more likely to be positive (Ahn et al., 2017).

Risk of bias was a significant concern for the studies in this review. Most of the RCT's were rated as having High risk of bias, while only three were rated as having some concerns. According to the authors of the ROB Assessment 2.0 Tool (Higgins et al., 2011), studies with a high risk of bias substantially lowers confidence in their results. All of the QEDs were rated as having either Critical or Serious risk of bias. The authors of the ROBINS ROB tool (Sterne et al., 2016) assert that studies with a critical risk of bias are too problematic to provide useful evidence on the effects of the intervention. These problematic high levels of risk of bias reduce the validity of the findings of this review. The overall findings of significant pre to post improvments, as well as support for the efficacy of FFT for adolescents with mental health issues should not be used to support claims that FFT has consistent, positive effects across studies.

Random allocation was also a concern. Although 12 studies reported random allocation of participants, in some cases, random allocation was altered to fit the availability of services (e.g., Alexander & Parson 1973). Further, not all studies who reported random allocation described how the allocation occurred. The selection of comparison groups also resulted in some bias. This was illustrated in the Kretschmar et al. (2018) study. They reported significantly better results between FFT and two alternatives on recidivism, however their comparison groups were families who, after being offered FFT, chose not to enrol or failed to complete. This reduces the credibility of the results, given that the remainers likely differed in key characteristics from those who did not participate.

Another common problem within the FFT literature base is the significant differences between the groups at baseline. Some studies (e.g., Baglivio, 2014) used propensity score matching to create statistically equivalent groups regarding some characteristics however,

other unmeasured characteristics may have remained significantly different. For example, Celinska's (2013) study had significant between group differences on gender, race/ethnicity, and pre-treatment criminal history. Eeran et al. (2018) had problematic baseline characteristic differences in that higher risk offenders who were ordered by the court to attend were assigned to the comparison MST group. Other studies did not implement statistical controls for baseline differences between groups (e.g., Van der Put 2013). Some studies had differences between groups in levels of care or attention, or in exposure to factors other than the interventions of interest. For example, Humayun et al. (2017) reported significant reduction in CD and ODD symptoms, however participants received FFT supplemented with caseworker support and Hukkleberg et al., (2022) reported their FFT intervention had significantly more treatment days than their MST comparison group.

Attrition bias also likely influenced the results in some studies. For example, Waldron et al., (2001) had 13 families who did not complete post-treatment assessments and were dropped from analysis. Removing those who do not complete treatment compromises external validity because the effect of the treatment in participants with missing data could not be established. Other methodological concerns involve the wide variations in alternative comparison treatments. Alternative interventions included: eclectic-psychodynamic family counselling, or client-centred family groups (Alexander, 1973); individual CBT or psychoeducation (Waldron, 2001); ecologically-based family therapy (Slesnick, 2004); Individual counselling or mentoring for youth (Celinska, 2013); Multisystemic Therapy (Baglivio, 2014). The nature of TAU also varied across the studies. TAU included: services provided by runaway shelters (Slesnick, 2009), usual probation services (Darnell et al., 2015), and casework services (Humayun, 2010). The impact of this wide variety of interventions to which FFT has been compared, is that there was considerable diversity in the duration, intensity, and amount of services families received. Therefore, any difference in

outcomes cannot reliably be attributed to the superiority of the intervention (Littell et al., 2023)

Other reviews have reported they found the overall quality of FFT reviews was low, which they claim makes any certainties about FFT inconclusive (Weisman & Montgomery, 2018). In contrast, Hartnett et al. (2017) conducted a review of 14 FFT studies and concluded that FFT is more effective than alternative treatments and no-treatment, but not more effective than TAU. These authors reported a low risk of bias assessment in the half of the studies and asserted that FFT was superior to other models of family therapy (including client-centered and psychodynamic), individual and group therapy, cognitive behaviour therapy, probation and mental health services, and parenting education groups. However, our results are consistent with those of Littell et al. (2023). Differences between these reviews could be a result of Littell et al, (2023) stringent exclusion of QEDs that lacked statistical controls for baseline differences, which may result in selection bias. Littell et al. (2023) also used robust analytic methods than those used in previous meta-analyses and collapsed results across study designs to increase statistical power.

Understanding the mental health variables that may influence the effectiveness of FFT is important given the high prevalence of mental health issues among adolescents who engage in antisocial behaviour. Knowledge of the impact of FFT on participants' mental health issues can assist with future planning to maximise the chance of the adolescent and their families' improved outcomes. While these studies collectively show continued support for the utility of FFT, it has also been shown that FFT is unlikely to be more effective than other forms of interventions for adolescents who engage in antisocial behaviour. There is some evidence to suggest that FFT in combination with other individual interventions designed to ameliorate mental health issues, such as CBT or depression specific individual programs, might be optimal treatment for families. Rhode et al. (2014) reported significant

depression remission from FFT combined with a Coping with Depression program, and Waldron et al. (2001) demonstrated that FFT and CBT produced significant reductions in percentage of days of substance use seven months after the intervention. It may be that treating an individual's mental health individually, in conjunction with family therapy, produces optimal outcomes.

## Limitations

While this review provides an overview of a gap in the existing FFT literature regarding mental health issues, there are several limitations that challenge the utility of this review. First, only peer reviewed studies were included which means other important published literature about FFT may have been excluded. It is also narrative in style which has fundamental shortcomings regarding potential bias in the appraisal of included articles and interpretation of their findings. Further, the inclusion of studies with high or critical risk of bias, while necessary for appraisal of all of the peer reviewed literature, was not ideal. Future research on the mental health of adolescents who participate in FFT could perform a meta-analysis to provide more precise estimates of the effects of FFT than those derived from the individual studies included within this review.

## Conclusion

Results indicate FFT improved some of the mental health issues of adolescents who engage in antisocial behaviours with significant pre-post-treatment changes being reported across a range of mental health related outcomes. Results also indicated that FFT is an efficacious treatment in reducing recidivism and improving family functioning. However, this was also true for many of the wide range comparison treatment groups, which suggests the improvements were not FFT specific. The FFT literature base is significantly limited by a range of methodological problems, particularly the risk of bias. Given the very high prevalence of mental health issues among adolescents who engage in antisocial behaviour,

their needs often extend beyond family functioning and risk factors associated with antisocial behaviours (Kang et al., 2018). Untreated mental health issues can affect adolescents detrimentally in a range of important ways including their physical health, school performance, relationships with peers and family, with some problems persisting into adulthood (World Health Organization, 2021). Thus, continued investigations into FFT with high methodological standards, and consistent quality measures of mental health issues, as well as for recidivism and family functioning is imperative.

**References**

Abram, K. M., Teplin, L. A., King, D. C., Longworth, S. L., Emanuel, K. M., Romero, E. G., & Olson, N. D. (2013). *PTSD, trauma, and comorbid psychiatric disorders in detained youth.* Washington, DC. Retrieved from https://ojjdp.ojp.gov/sites/g/files/xyckuh176/files/pubs/239603.pdf

Achenbach, T. M., & Rescorla, L. A. (2001). *Child behavior checklist for ages 6-18* (pp. 6-1). Burlington, VT: University of Vermont. Retrieved from https://reachoutandread.org/wp-content/uploads/2023/06/CBCL-1-Summary.pdf

Ahn, R., Woodbridge, A., Abraham, A., Saba, S., Korenstein, D., Madden, E., ... & Keyhani, S. (2017). Financial ties of principal investigators and randomized controlled trial outcomes: Cross sectional study. *British Medical Journal* 356. https://doi.org/10.1136/bmj.i6770

Alexander, J. F., Barton, C., Schiaro, R. S., & Parsons, B. V. (1976). Systems-behavioral intervention with families of delinquents: Therapist characteristics, family behavior, and outcome. *Journal of Consulting and Clinical Psychology*, *44*(4), 656. https://doi.org/10.1037/0022-006x.44.4.656

Alexander, J. F., & Parsons, B. V. (1973). Short-term behavioral intervention with delinquent families: Impact on family process and recidivism. *Journal of Abnormal Psychology, 81*(3), 219–225. https://doi.org/10.1037/h0034537

Alexander, J., & Parsons, B. V. (1982). *Functional Family Therapy.* Brooks/Cole Publishing Company. https://doi.org/10.1037/11621-000

Alexander, J. F., Robbins, M. S., & Sexton, T. L. (2000). Family-based interventions with older, at-risk youth: From promise to proof to practice. *Journal of Primary Prevention, 21*(2), 185-205. https://doi.org/10.1037/10438-002

Alexander, J. F., Waldron, H. B., Robbins, M. S., & Neeb, A. A. (2013). *Functional Family Therapy for adolescent behavior problems*. American Psychological Association. https://doi.org/10.1037/14139-000

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Baglivio, M. T., Jackowski, K., Greenwald, M. A., & Wolff, K. T. (2014). Comparison of multisystemic therapy and functional family therapy effectiveness: A multiyear statewide propensity score matching analysis of juvenile offenders. *Criminal Justice and Behavior*, *41*(9), 1033-1056. https://doi.org/10.1177/0093854814543272

Beck, A. T., Ward, C., Mendelson, M., Mock, J., & Erbaugh, J. J. A. G. P. (1961). Beck depression inventory (BDI). *Archives Of General Psychiatry*, *4*(6), 561-571. https://doi.org/10.1037/t00741-000

Carr, A. (2014). The evidence base for family therapy and systemic interventions for child-focused problems. *Journal of Family Therapy*, *36*(2), 107-157. https://doi.org/10.1111/1467-6427.12032

Celinska, K., & Cheng, C. C. (2017). Gender comparisons in the processes and outcomes of functional family therapy. *Journal of Juvenile Justice, 6*(1), 82. Retrieved from http://ezproxy.newcastle.edu.au/login?url=https://www.proquest.com/scholarly-journals/gender-comparisons-processes-outcomes-functional/docview/1916119391/se-2?accountid=10499

Celinska, K., Furrer, S., & Cheng, C. C. (2013). An outcome-based evaluation of functional family therapy for youth with behavioral problems. *Journal of Juvenile Justice, 2*(2), 23. https://doi.org/10.1111/1467-6427.12224

Celinska, K., Sung, H. E., Kim, C., & Valdimarsdottir, M. (2019). An outcome evaluation of functional family therapy for court-involved youth. *Journal of Family Therapy, 41*(2), 251-276. https://doi.org/10.1111/1467-6427.12224

Colman, I., Murray, J., Abbott, R. A., Maughan, B., Kuh, D., Croudace, T. J., & Jones, P. B. (2009). Outcomes of conduct problems in adolescence: 40 year follow-up of national cohort. *British Medical Journal*, *338*. https://doi.org/10.1136/bmj.a2981

Cox, E., Leung, R., Baksheev, G., Day, A., Toumbourou, J. W., Miller, P., ... & Walker, A. (2016). Violence prevention and intervention programmes for adolescents in Australia: A systematic review. *Australian Psychologist, 51*(3), 206-222. https://doi.org/10.1111/ap.12168

Darnell, A. J., & Schuler, M. S. (2015). Quasi-experimental study of Functional Family Therapy effectiveness for juvenile justice aftercare in a racially and ethnically diverse community sample. *Children and Youth Services Review*, *50*, 75-82. https://doi.org/10.1016/j.childyouth.2015.01.013

Dishion, T. J., & Patterson, G. R. (2015). The development and ecology of antisocial behavior in children and adolescents. *Developmental psychopathology: Volume three: Risk, disorder, and adaptation*, 503-541. https://doi.org/10.1002/9780470939406.ch13

Dopp, A. R., Borduin, C. M., White II, M. H., & Kuppens, S. (2017). Family-based treatments for serious juvenile offenders: A multilevel meta-analysis. *Journal of Consulting and Clinical Psychology, 85*(4), 335. https://doi.org/10.1037/ccp0000183

Eeren, H. V., Goossens, L. M., Scholte, R. H., Busschbach, J. J., & Van der Rijken, R. E. (2018). Multisystemic therapy and functional family therapy compared on their effectiveness using the propensity score method. *Journal of Abnormal Child Psychology*, *46*, 1037-1050. https://doi.org/10.1007/s10802-017-0392-4

Eisner, M. (2009). No ewffects in independent prevention trials: Can we reject the cynical view?. *Journal of Experimental Criminology*, *5*, 163-183. https://doi.org/10.1007/s11292-009-9071-y

Fairchild, G., Hawes, D. J., Frick, P. J., Copeland, W. E., Odgers, C. L., Franke, B., ... & De Brito, S. A. (2019). Conduct disorder. *Nature Reviews Disease Primers*, *5*(1), 43. https://doi.org/10.1038/s41572-019-0095-y

Frick, P. J., Barry, C. T., & Kamphaus, R. W. (2010). *Clinical assessment of child and adolescent personality and behavior* (pp. 413-435). New York: Springer. Retrieved from http://197.136.134.32/bitstream/handle/123456780/6207/2010_Book_ClinicalAssessmentOfChildAndAd.pdf?sequence=1&isAllowed=y

Frick, P. J., Ray, J. V., Thornton, L. C., & Kahn, R. E. (2014). Can callous-unemotional traits enhance the understanding, diagnosis, and treatment of serious conduct problems in children and adolescents? A comprehensive review. *Psychological Bulletin*, *140*(1), 1. https://doi.org/10.1037/a0033076

Functional Family Therapy Limited liability company (2022) https://www.fftllc.com/

Gan, D. Z., Zhou, Y., Abdul Wahab, N. D. B., Ruby, K., & Hoo, E. (2021). Effectiveness of Functional Family Therapy in a non-western context: Findings from a randomized-controlled evaluation of youth offenders in Singapore. *Family Process*, *60*(4), 1170-1184. https://doi.org/10.1111/famp.12630

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(11), 1337-1345. https://doi.org/10.1097/00004583-200111000-00015

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a

community sample. *The British Journal of Psychiatry*, *177*(6), 534-539. https://doi.org/10.1192/bjp.177.6.534

Gordon, D. A., Arbuthnot, J., Gustafson, K. E., & McGreen, P. (1988). Home-based behavioral-systems family therapy with disadvantaged juvenile delinquents. *American Journal of Family Therapy*, *16*(3), 243-255. https://doi.org/10.1080/01926188808250729

Graham, C., Carr, A., Rooney, B., Sexton, T., & Wilson Satterfield, L. R. (2014). Evaluation of functional family therapy in an Irish context. *Journal of Family Therapy*, *36*(1), 20-38. https://doi.org/10.1111/1467-6427.12028

Gregory, A., & Thompson, A. R. (2010). African American high school students and variability in behavior across classrooms. *Journal of Community Psychology*, *38*(3), 386-402. https://doi.org/10.1002/jcop.20370

Hartnett, D., Carr, A., Hamilton, E., & O'Reilly, G. (2017). The effectiveness of functional family therapy for adolescent behavioral and substance misuse problems: A meta-analysis. *Family Process*, *56*(3), 607-619. https://doi.org/10.1111/famp.12256

Hartnett, D., Carr, A., & Sexton, T. (2016). The Effectiveness of Functional Family Therapy in Reducing Adolescent Mental Health Risk and Family Adjustment Difficulties in an Irish Context. *Family Process*, *55*(2), 287-304. https://doi.org/10.1111/famp.12195

Hawes, S. W., Byrd, A. L., Henderson, C. E., Gazda, R. L., Burke, J. D., Loeber, R., & Pardini, D. A. (2014). Refining the parent-reported Inventory of callous–unemotional traits in boys with conduct problems. *Psychological Assessment*, *26*(1), 256. https://doi.org/10.1037/a0034718

Heerde, J. A., Toumbourou, J. W., Hemphill, S. A., Le, H., Herrenkohl, T. I., & Catalano, R. F. (2018). Prevent crime and save money: Return-on-investment models in

Australia. *Trends and Issues in Crime and Criminal Justice*, (545), 1-19. Retrieved from https://www.aic.gov.au/publications/tandi/tandi545

Hemphill, S. A., Heerde, J. A., Herrenkohl, T. I., & Farrington, D. P. (2015). Within-individual versus between-individual predictors of antisocial behaviour: A longitudinal study of young people in Victoria, Australia. *Australian & New Zealand Journal of Criminology*, *48*(3), 429-445. https://doi.org/10.1177/0004865815589829

Heywood, C., & Fergusson, D. (2016). A pilot study of functional family therapy in New Zealand. *New Zealand Journal of Psychology (Online)*, *45*(3), 12**.** Retrieved from https://web-s-ebscohost-com.ezproxy.newcastle.edu.au/ehost/pdfviewer/pdfviewer?vid=0&sid=55cad5cd-95de-47d9-91be-16ecbebd1b0d%40redis

Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., ... & Sterne, J. A. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal,* 343. https://doi.org/10.1136/bmj.d5928

Hill, C. R., & Hughes, J. N. (2007). An examination of the convergent and discriminant validity of the Strengths and Difficulties Questionnaire. *School Psychology Quarterly*, *22*(3), 380. https://doi.org/10.1037/1045-3830.22.3.380

Hukkelberg, S., Ogden, T., & Thøgersen, D. M. (2022). Youth level of service/case management inventory assessments as predictors of behavioral change in multisystemic therapy and functional family therapy in Norway. *Research on Social Work Practice*, *32*(7), 784-792. https://doi.org/10.1177/10497315221086641

Humayun, S., Herlitz, L., Chesnokov, M., Doolan, M., Landau, S., & Scott, S. (2017). Randomized controlled trial of Functional Family Therapy for offending and antisocial behavior in UK youth. *Journal of Child Psychology and Psychiatry*, *58*(9), 1023-1032. https://doi.org/10.1111/jcpp.12743

Kang, T., Wood, J. M., Eno Louden, J., & Ricks, E. P. (2018). Prevalence of internalizing, externalizing, and psychotic disorders among low-risk juvenile offenders. *Psychological Services*, *15*(1), 78. https://doi.org/10.1037/ser0000152

Kaufman, J., Birmaher, B., Brent, D., Rao, U. M. A., Flynn, C., Moreci, P., ... & Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(7), 980-988. https://doi.org/10.1097/00004583-199707000-00021

Klein, N. C., Alexander, J. F., & Parsons, B. V. (1977). Impact of family systems intervention on recidivism and sibling delinquency: A model of primary prevention and program evaluation. *Journal of Consulting and Clinical Psychology*, *45*(3), 469. https://doi.org/10.1037/0022-006x.45.3.469

Kretschmar, J. M., Tossone, K., Butcher, F., & Marsh, B. (2018). Examining the impact of a juvenile justice diversion program for youth with behavioral health concerns on early adulthood recidivism. *Children and Youth Services Review, 91*, 168-176. https://doi.org/10.1016/j.childyouth.2018.06.010

Lambert, M. J., Lunnen, K., Umphress, V., Hansen, N., & Burlingame, G. M. (1994). Outcome Questionnaire 45.2. PsycTESTS Dataset. https://doi.org/10.1037/t00830-000

Littell, J. H., Pigott, T. D., Nilsen, K. H., Roberts, J., & Labrum, T. K. (2023). Functional Family Therapy for families of youth (age 11–18) with behaviour problems: A systematic review and meta-analysis. *Campbell Systematic Reviews, 19(*3), e1324. https://doi.org/10.1002/cl2.1324

Lyons, J. S. (2009). Measurement as communication. *Communimetrics: A Communication Theory of Measurement in Human Service Settings*, 19-44. https://doi.org/10.1007/978-0-387-92822-7_2

Marshall, J., Hamilton, R., & Cairns, N. (2018). Application of the Strengths and Difficulties Questionnaire Added Value Score in evaluating the effectiveness of Functional Family Therapy within local authority social work services. *Child and Adolescent Mental Health*, *23*(2), 85-91. https://doi.org/10.1111/camh.12190

McGee, T. R., Whitten, T., Williams, C., Jolliffe, D., & Farrington, D. P. (2021). Classification of patterns of offending in developmental and life-course criminology, with special reference to persistence. *Aggression and Violent Behavior, 59*, 101460. https://doi.org/10.1016/j.avb.2020.101460

McLeod, J. D., Uemura, R., & Rohrman, S. (2012). Adolescent mental health, behavior problems, and academic achievement. *Journal of Health and Social Behavior*, *53*(4), 482-497. https://doi.org/10.1177/0022146512462888

McMahon, R. J., & Frick, P. J. (2005). Evidence-based assessment of conduct problems in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, *34*(3), 477-505. https://doi.org/10.1207/s15374424jccp3403_6

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS med, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

National Institute for Health and Care Excellence (2017). Antisocial behaviour and conduct disorders in children and young people. Recognition, intervention, and management. NICE Clinical Guidelines, No. 158. Leicester: British Psychological Society. Retrieved from https://pubmed.ncbi.nlm.nih.gov/26065062/

Paradis, A. D., Koenen, K. C., Fitzmaurice, G. M., & Buka, S. L. (2016). Impact of persistent and adolescent-limited antisocial behaviour on adult health outcomes. *Journal Epidemiology and Community Health*. *70*(10), 1004–1010. https://doi.org/10.1136/jech-2014-205140

Parsons, B. V., & Alexander, J. F. (1973). Short-term family intervention: a therapy outcome study. *Journal of Consulting and Clinical Psychology*, *41*(2), 195. https://doi.org/10.1037/h0035181

Poznanski, E. O., Grossman, J. A., Buchsbaum, Y., Banegas, M., Freeman, L., & Gibbons, R. (1995). Children's depression rating scale--revised. *Journal of the American Academy of Child Psychiatry*. https://doi.org/10.1037/t55280-000

Reynolds, C. R., & Kamphaus, R. W. (2006). Behavior Assessment System for Children--Second Edition. *PsycTESTS Dataset.* https://doi.org/10.1037/t15079-000

Ridge, N. W., Warren, J. S., Burlingame, G. M., Wells, M. G., & Tumblin, K. M. (2009). Reliability and validity of the youth outcome questionnaire self-report. *Journal of Clinical Psychology*, *65*(10), 1115-1126. https://doi.org/10.1002/jclp.20620

Robbins, M. S., Waldron, H. B., Turner, C. W., Brody, J., Hops, H., & Ozechowski, T. (2019). Evaluating supervision models in functional family therapy: Does adding observation enhance outcomes?. *Family Process*, *58*(4), 873-890. https://doi.org/10.1111/famp.12399

Rohde, P., Waldron, H. B., Turner, C. W., Brody, J., & Jorgensen, J. (2014). Sequenced versus coordinated treatment for adolescents with comorbid depressive and substance use disorders. *Journal of Consulting and Clinical Psychology*, *82*(2), 342. https://doi.org/10.1037/a0035808

Sawyer, A. M., Borduin, C. M., & Dopp, A. R. (2015). Long-term effects of prevention and treatment on youth antisocial behavior: A meta-analysis. *Clinical Psychology Review*, *42*, 130-144. https://doi.org/10.1016/j.cpr.2015.06.009

Scavenius, C., Granski, M., Lindberg, M. R., Vardanian, M. M., & Chacko, A. (2020). Adolescent gender and age differences in responsiveness to functional family therapy. *Family Process*, *59*(4), 1465-1482. https://doi.org/10.1111/famp.12512

Stratton, P., Bland, J., Janes, E., & Lask, J. (2010). Developing an indicator of family function and a practicable outcome measure for systemic family and couple therapy: the SCORE. *Journal of Family Therapy*, *32*(3), 232-258. https://doi.org/10.1111/j.1467-6427.2010.00507.x

Sexton, T. L & Alexander, J. F (2000). Juvenile Justice Bulletin. *Office of Juvenile Justice and Delinquency Prevention.* Retrieved from https://www.ncjrs.gov/pdffiles1/ojjdp/184743.pdf

Sexton, T., & Turner, C. W. (2010). The effectiveness of functional family therapy for youth with behavioral problems in a community practice setting. *Journal of Family Psychology*, *24*(3), 339. https://doi.org/10.1037/a0019406

Shaw, D. S., Dishion, T. J., Supplee, L., Gardner, F., & Arnds, K. (2006). Randomized trial of a family-centered approach to the prevention of early conduct problems: 2-year effects of the family check-up in early childhood. *Journal of Consulting and Clinical Psychology*, *74*(1), 1. https://doi.org/10.1037/0022-006x.74.1.1

Shufelt, J. L., & Cocozza, J. J. (2006). *Youth with mental health disorders in the juvenile justice system: Results from a multi-state prevalence study* (pp. 1-6). Delmar, NY: National Center for Mental Health and Juvenile Justice. Retrieved from https://www.njjn.org/uploads/digital-library/Youth-with-Mental-Health-Disorders-in-the-JJ-%20System_Shufelt-and-Cocozza-NCMHJJ-6.2006.pdf

Slesnick, N., & Prestopnik, J. L. (2009). Comparison of family therapy outcome with alcohol-abusing, runaway adolescents. *Journal of Marital and Family Therapy*, *35*(3), 255-277. https://doi.org/10.1111/j.1752-0606.2009.00121.x

Smith, R. G., Jorna, P., Sweeney, J., & Fuller, G. (2014). Counting the costs of crime in Australian Institute of Criminology (2014) Australia: A 2011 estimate with rates of crime highest among those adolescents and young adults aged 15 to 24 years. *Australian crime: facts & figures 2013*. Retrieved from https://www.aic.gov.au/publications/facts/facts-14

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., ... & Higgins, J. P. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *British Medical Journal. 355*. https://doi.org/10.1136/bmj.i4919

Stratton, P., Bland, J., Janes, E., & Lask, J. (2010). Developing an indicator of family function and a practicable outcome measure for systemic family and couple therapy: The SCORE. *Journal of Family Therapy*, *32*(3), 232-258. https://doi.org/10.1111/j.1467-6427.2010.00507.x

Thøgersen, D. M., Bjørnebekk, G., Scavenius, C., & Elmose, M. (2021). Callous-unemotional traits do not predict functional family therapy outcomes for adolescents with behavior problems. *Frontiers in Psychology*, *11*, 537706. https://doi.org/10.3389/fpsyg.2020.537706

Thøgersen, D. M., Elmose, M., Viding, E., McCrory, E., & Bjørnebekk, G. (2022). Behavioral improvements but limited change in callous-unemotional traits in adolescents treated for conduct problems. *Journal of Child and Family Studies*, *31*(12), 3342-3358. https://doi.org/10.1007/s10826-022-02435-6

Uink, B. N., Modecki, K. L., & Barber, B. L. (2017). Disadvantaged youth report less negative emotion to minor stressors when with peers: An experience sampling study. *International Journal of Behavioral Development*, *41*(1), 41-51. https://doi.org/10.1177/0165025415626516

Van der Put, C. E., Stams, G. J. J., Hoeve, M., Deković, M., Spanjaard, H. J. M., van der Laan, P. H., & Barnoski, R. P. (2012). Changes in the relative importance of dynamic risk factors for recidivism during adolescence. *International Journal of Offender Therapy and Comparative Criminology*, *56*(2), 296-316. https://doi.org/10.1177/0306624x11398462

Vardanian, M. M., Scavenius, C., Granski, M., & Chacko, A. (2020). An international examination of the effectiveness of functional family therapy (FFT) in a Danish community sample. *Journal of Marital and Family Therapy*, *46*(2), 289-303. https://doi.org/10.1111/jmft.12405

Waldron, H. B., Slesnick, N., Brody, J. L., Peterson, T. R., & Turner, C. W. (2001). Treatment outcomes for adolescent substance abuse at 4- and 7-month assessments. *Journal of Consulting and Clinical Psychology*, *69*(5), 802–813. https://doi.org/10.1037/0022-006x.69.5.802

Weisman, C. B., & Montgomery, P. (2019). Functional Family Therapy (FFT) for behavior disordered youth aged 10–18: An overview of reviews. *Research on Social Work Practice*, *29*(3), 333-346. https://doi.org/10.1177/1049731518792309

Weisz, J. R., Kuppens, S., Ng, M. Y., Vaughn-Coaxum, R. A., Ugueto, A. M., Eckshtain, D., & Corteselli, K. A. (2019). Are psychotherapies for young people growing stronger? Tracking trends over time for youth anxiety, depression, attention-deficit/hyperactivity disorder, and conduct problems. *Perspectives on Psychological Science, 14*(2), 216-237. https://doi.org/10.1177/1745691618805436

Wells, M. G., Burlingame, G. M., Lambert, M. J., Hoag, M. J., & Hope, C. A. (1996). Conceptualization and measurement of patient change during psychotherapy: Development of the Outcome Questionnaire and Youth Outcome Questionnaire. *Psychotherapy: Theory, Research, Practice, Training*, *33*(2), 275. https://doi.org/10.1037/0033-3204.33.2.275

White, S. F., Frick, P. J., Lawing, K., & Bauer, D. (2013). Callous–unemotional traits and response to Functional Family Therapy in adolescent offenders. *Behavioral Sciences & the Law*, *31*(2), 271-285. https://doi.org/10.1002/bsl.2041

World Health Organization (2021) Mental health of adolescents. Retrieved from https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health

# Appendices

Appendix A: Instructions to Authors for the Journal of Family Psychology

**Submission**
To submit to the Editorial Office of Arin M. Connell, please submit manuscripts electronically through the Manuscript Submission Portal in Microsoft Word or Open Office format. Prepare manuscripts according to the *Publication Manual of the American Psychological Association* using the 7th edition. Manuscripts may be copyedited for bias-free language (see Chapter 5 of the *Publication Manual*). APA Style and Grammar Guidelines for the 7th edition are available. In addition to addresses and phone numbers, please supply electronic mail addresses and fax numbers, if available, for potential use by the editorial office and later by the production office. Keep a copy of the manuscript to guard against loss.

*Journal of Family Psychology* is now using a software system to screen submitted content for similarity with other published content. The system compares the initial version of each submitted manuscript against a database of 40+ million scholarly documents, as well as content appearing on the open web. This allows APA to check submissions for potential overlap with material previously published in scholarly journals (e.g., lifted, or republished material).

**Editor's Choice**
Each issue of the *Journal of Family Psychology* will highlight one manuscript with the designation as an "Editor's Choice" paper. Selection is based on the recommendations of the associate editors, who consider the paper's potential impact to the field, the distinction of expanding the contributors to, or the focus of, our science, or its discussion of an important future direction for science.

**Article requirements**
For general guidelines to style, authors should study articles previously published in the journal. All manuscripts must include an abstract containing a maximum of 250 words typed on a separate page. After the abstract, please supply up to five keywords or brief phrases.

The manuscript title should be accurate, fully explanatory, and preferably no longer than 12 words. The title should reflect the content and population studied (e.g., "family therapy for depression in children"). If the paper reports a randomized clinical trial, this should be indicated in the title, and the CONSORT criteria must be used for reporting purposes.

Research manuscripts and review and theoretical manuscripts that provide creative and integrative summaries of an area of work relevant to family psychology should not exceed 30–35 pages, all inclusive (including cover page, abstract, text, references, tables, figures), with margins of at least 1 inch on all sides and a standard font (e.g., Times New Roman) of 12 points (no smaller). The entire paper (text, references, tables, figures, etc.,) must be double spaced. References should not exceed 8 pages.

Brief reports are encouraged for innovative work that may be premature for publication as a full research report because of small sample size, novel methodologies, etc. Brief reports also are an appropriate format for replications and for clinical case studies (note that replication submissions should include "A Replication of XX Study" in the subtitle of the manuscript as well as in the abstract). Authors of brief reports should indicate in the cover letter that the full

report is not under consideration for publication elsewhere. Brief reports should be designated as such and should not exceed a total of 20 pages, all-inclusive. References should not exceed eight pages. Manuscripts exceeding the space requirement will be returned to the author for shortening prior to peer review. All research involving human participants must describe oversight of the research process by the relevant Institutional Review Boards and should describe consent and assent procedures briefly in the Method section.

It is important to highlight the significance and novel contribution of the work. The translation of research into practice must be evidenced in all manuscripts. Authors should incorporate a meaningful discussion of the clinical and/or policy implications of their work throughout the manuscript, rather than simply providing a separate section for this material.

**Masked review**
The *Journal of Family Psychology*® uses a masked reviewing system for all submissions. The cover letter should include all authors' names and institutional affiliations. However, in order to permit anonymous review, the first page of text should omit this information. This cover page should only include the title of the manuscript and the date it is submitted.

Please make every effort to see that the manuscript itself contains no clues to the authors' identities, including grant numbers, names of institutions providing IRB approval, self-citations, and links to online repositories for data, materials, code, or preregistrations (e.g., Create a View-only Link for a Project).

Please ensure that the final version for production includes a byline and full author note for typesetting.

**Cover letter**
Authors should indicate in their cover letter that the work has not been published previously and is not under consideration for publication elsewhere. The relationship of the submitted manuscript with other publications and/or submissions of the author, if any, should be explained. The cover letter should include a statement indicating that the manuscript has been seen and reviewed by all authors and that all authors have contributed to it in a meaningful way. The cover letter must include the full mailing address, telephone, fax, and email address for the corresponding author.

**CONSORT criteria**
The *Journal of Family Psychology* requires the use of the CONSORT reporting standards (i.e., a checklist and flow diagram) for randomized clinical trials, consistent with the policy established by the Publications and Communications Board of APA.

CONSORT (Consolidated Standards of Reporting Trials) offers a standard way to improve the quality of such reports and to ensure that readers have the information necessary to evaluate the quality of a clinical trial. Manuscripts that report randomized clinical trials are required to include a flow diagram of the progress through the phases of the trial and a checklist that identifies where in the manuscript the various criteria are addressed. The checklist should be placed in an appendix of the manuscript for review purposes. When a study is not fully consistent with the CONSORT statement, the limitations should be acknowledged and discussed in the text of the manuscript. For follow-up studies of previously published clinical trials, authors should submit a flow diagram of the progress through the phases of the trial and follow-up. The above checklist information should be

completed to the extent possible, especially for the Results and Discussion sections of the manuscript. Visit the CONSORT Statement Web site for more details and resources.

**Journal Article Reporting Standards**
Authors are encouraged to consult the APA Style Journal Article Reporting Standards (JARS) for quantitative, qualitative, and mixed methods research. Updated in 2018, the standards offer ways to improve transparency in reporting to ensure that readers have the information necessary to evaluate the quality of the research and to facilitate collaboration and replication.

**The new JARS:**
recommend the division of hypotheses, analyses, and conclusions into primary, secondary, and exploratory groupings to allow for a full understanding of quantitative analyses presented in a manuscript and to enhance reproducibility; offer modules for authors reporting on N-of-1 designs, replications, clinical trials, longitudinal studies and observational studies, as well as the analytic methods of structural equation modeling and Bayesian analysis; and include guidelines on reporting of study preregistration (including making protocols public); participant characteristics (including demographic characteristics); inclusion and exclusion criteria; psychometric characteristics of outcome measures and other variables; and planned data diagnostics and analytic strategy.

**Transparency and openness**
APA endorses the Transparency and Openness Promotion (TOP) Guidelines by a community working group in conjunction with the Center for Open Science (Nosek et al. 2015). Effective July 1, 2021, empirical research, including meta-analyses, submitted to the *Journal of Family Psychology* must meet the "disclosure" level for all eight aspects of research planning and reporting. Authors should include a subsection in the method section titled "Transparency and openness." This subsection should detail the efforts the authors have made to comply with the TOP guidelines. For example: We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study, and we follow JARS (Kazak, 2018). All data, analysis code, and research materials are available at [stable link to repository]. Data were analyzed using R, version 4.0.0 (R Core Team, 2020) and the package *ggplot*, version 3.2.1 (Wickham, 2016). This study's design and its analysis were not pre-registered. Links to preregistrations and data, code, and materials should also be included in the author note.

**Data, materials, and code**
Authors must state whether data and study materials are available and, if so, where to access them. Recommended repositories include APA's repository on the Open Science Framework (OSF), or authors can access a full list of other recommended repositories.
In both the Author Note and at the end of the Method section, specify whether and where the data and material will be available or include a statement noting that they are not available. For submissions with quantitative or simulation analytic methods, state whether the study analysis code is available, and, if so, where to access it. For example: All data have been made publicly available at the [repository name] and can be accessed at [persistent URL or DOI]. Materials and analysis code for this study are available by emailing the corresponding author. Materials and analysis code for this study are not available. The code behind this analysis/simulation has been made publicly available at the [repository name] and can be accessed at [persistent URL or DOI]. Preregistration of studies and analysis plans

Preregistration of studies and specific hypotheses can be a useful tool for making strong theoretical claims. Likewise, preregistration of analysis plans can be useful for distinguishing confirmatory and exploratory analyses. Investigators are encouraged to preregister their studies and analysis plans prior to conducting the research (e.g., ClinicalTrials.gov or the Preregistration for Quantitative Research in Psychology template) via a publicly accessible registry system (e.g., OSF, ClinicalTrials.gov, or other trial registries in the WHO Registry Network). Articles must state whether or not any work was preregistered and, if so, where to access the preregistration. If any aspect of the study is preregistered, include the registry link in the method section and the author note. For example: This study's design was preregistered; see [STABLE LINK OR DOI]. This study's design and hypotheses were preregistered; see [STABLE LINK OR DOI]. This study's analysis plan was preregistered; see [STABLE LINK OR DOI]. This study was not preregistered.

**Manuscript preparation**

Prepare manuscripts according to the *Publication Manual of the American Psychological Association* using the 7th edition. Manuscripts may be copyedited for bias-free language (see Chapter 5 of the *Publication Manual*). Review APA's Journal Manuscript Preparation Guidelines before submitting your article. Double-space all copy. Other formatting instructions, as well as instructions on preparing tables, figures, references, metrics, and abstracts, appear in the *Manual*. Additional guidance on APA Style is available on the APA Style website. Below are additional instructions regarding the preparation of display equations, computer code, and tables.

**Display equations**

We strongly encourage you to use MathType (third-party software) or Equation Editor 3.0 (built into pre-2007 versions of Word) to construct your equations, rather than the equation support that is built into Word 2007 and Word 2010. Equations composed with the built-in Word 2007/Word 2010 equation support are converted to low-resolution graphics when they enter the production process and must be rekeyed by the typesetter, which may introduce errors. To construct your equations with MathType or Equation Editor 3.0: Go to the Text section of the Insert tab and select Object. Select MathType or Equation Editor 3.0 in the drop-down menu. If you have an equation that has already been produced using Microsoft Word 2007 or 2010 and you have access to the full version of MathType 6.5 or later, you can convert this equation to MathType by clicking on MathType Insert Equation. Copy the equation from Microsoft Word and paste it into the MathType box. Verify that your equation is correct, click File, and then click Update. Your equation has now been inserted into your Word file as a MathType Equation. Use Equation Editor 3.0 or MathType only for equations or for formulas that cannot be produced as Word text using the Times or Symbol font.

**Computer code**

Because altering computer code in any way (e.g., indents, line spacing, line breaks, page breaks) during the typesetting process could alter its meaning, we treat computer code differently from the rest of your article in our production process. To that end, we request separate files for computer code.

**In online supplemental material**

We request that runnable source code be included as supplemental material to the article. For more information, visit Supplementing Your Article With Online Material.

**In the text of the article**

If you would like to include code in the text of your published manuscript, please submit a separate file with your code exactly as you want it to appear, using Courier New font with a type size of 8 points. We will make an image of each segment of code in your article that exceeds 40 characters in length. (Shorter snippets of code that appear in text will be typeset in Courier New and run in with the rest of the text.) If an appendix contains a mix of code and explanatory text, please submit a file that contains the entire appendix, with the code keyed in 8-point Courier New.

**Tables**
Use Word's insert table function when you create tables. Using spaces or tabs in your table will create problems when the table is typeset and may result in errors.
Academic writing and English language editing services

Authors who feel that their manuscript may benefit from additional academic writing or language editing support prior to submission are encouraged to seek out such services at their host institutions, engage with colleagues and subject matter experts, and/or consider several vendors that offer discounts to APA authors.

Please note that APA does not endorse or take responsibility for the service providers listed. It is strictly a referral service.

Use of such service is not mandatory for publication in an APA journal. Use of one or more of these services does not guarantee selection for peer review, manuscript acceptance, or preference for publication in any APA journal.

**Submitting supplemental materials**
APA can place supplemental materials online, available via the published article in the PsycArticles® database. Please see Supplementing Your Article With Online Material for more details.

**Abstract and keywords**
All manuscripts must include an abstract containing a maximum of 250 words typed on a separate page. After the abstract, please supply up to five keywords or brief phrases.

**References**
List references in alphabetical order. Each listed reference should be cited in text, and each text citation should be listed in the references section.

Examples of basic reference formats:

*Journal article*
McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, *126*(1), 1–51. https://doi.org/10.1037/rev0000126

*Authored book*
Brown, L. S. (2018). *Feminist therapy* (2nd ed.). American Psychological Association. https://doi.org/10.1037/0000092-000

*Chapter in an edited book*

Balsam, K. F., Martell, C. R., Jones. K. P., & Safren, S. A. (2019). Affirmative cognitive behavior therapy with sexual and gender minority people. In G. Y. Iwamasa & P. A. Hays (Eds.), *Culturally responsive cognitive behavior therapy: Practice and supervision* (2nd ed., pp. 287–314). American Psychological Association. https://doi.org/10.1037/0000119-012

*Data set citation*
Alegria, M., Jackson, J. S., Kessler, R. C., & Takeuchi, D. (2016). Collaborative Psychiatric Epidemiology Surveys (CPES), 2001–2003 [Data set]. Inter-university Consortium for Political and Social Research. https://doi.org/10.3886/ICPSR20240.v8

*Software/Code citation*
Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. https://www.jstatsoft.org/v36/i03/

Wickham, H. et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686, https://doi.org/10.21105/joss.01686

All data, program code, and other methods should be appropriately cited in the text and listed in the references section.

**Figures**
Preferred formats for graphics files are TIFF and JPG, and preferred format for vector-based files is EPS. Graphics downloaded or saved from web pages are not acceptable for publication. Multipanel figures (i.e., figures with parts labeled a, b, c, d, etc.) should be assembled into one file. When possible, please place symbol legends below the figure instead of to the side.

**Resolution**
All color line art and halftones: 300 DPI
Black and white line tone and gray halftone images: 600 DPI

**Line weights**
Adobe Photoshop images
Color (RGB, CMYK) images: 2 pixels
Grayscale images: 4 pixels
Adobe Illustrator Images
Stroke weight: 0.5 points

APA offers authors the option to publish their figures online in color without the costs associated with print publication of color figures.

The same caption will appear on both the online (color) and print (black and white) versions. To ensure that the figure can be understood in both formats, authors should add alternative wording (e.g., “the red (dark gray) bars represent”) as needed.

For authors who prefer their figures to be published in color both in print and online, original color figures can be printed in color at the editor's and publisher's discretion provided the author agrees to pay:
$900 for one figure
An additional $600 for the second figure

An additional $450 for each subsequent figure

**Permissions**
Authors of accepted papers must obtain and provide to the editor on final acceptance all necessary permissions to reproduce in print and electronic form any copyrighted work, including test materials (or portions thereof), photographs, and other graphic images (including those used as stimuli in experiments).

On advice of counsel, APA may decline to publish any image whose copyright status is unknown. Download Permissions Alert Form (PDF, 13KB)

**Publication policies**
APA policy prohibits an author from submitting the same manuscript for concurrent consideration by two or more publications.

See also APA Journals® Internet Posting Guidelines. APA requires authors to reveal any possible conflict of interest in the conduct and reporting of research (e.g., financial interests in a test or procedure, funding by pharmaceutical companies for drug research).

Download Disclosure of Interests Form (PDF, 38KB)

In light of changing patterns of scientific knowledge dissemination, APA requires authors to provide information on prior dissemination of the data and narrative interpretations of the data/research appearing in the manuscript (e.g., if some or all were presented at a conference or meeting, posted on a listserv, shared on a website, including academic social networks like ResearchGate, etc.). This information (2–4 sentences) must be provided as part of the author note.

Authors of accepted manuscripts are required to transfer the copyright to APA.
For manuscripts **not** funded by the Wellcome Trust or the Research Councils UK
Publication Rights (Copyright Transfer) Form (PDF, 83KB)
For manuscripts funded by the Wellcome Trust or the Research Councils UK
Wellcome Trust or Research Councils UK Publication Rights Form (PDF, 34KB)

**Ethical Principles**
It is a violation of APA Ethical Principles to publish "as original data, data that have been previously published" (Standard 8.13).

In addition, APA Ethical Principles specify that "after research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through reanalysis and who intend to use such data only for that purpose, provided that the confidentiality of the participants can be protected and unless legal rights concerning proprietary data preclude their release" (Standard 8.14). APA expects authors to adhere to these standards. Specifically, APA expects authors to have their data available throughout the editorial review process and for at least 5 years after the date of publication.

Authors are required to state in writing that they have complied with APA ethical standards in the treatment of their sample, human or animal, or to describe the details of treatment.

Download Certification of Compliance with APA Ethical Principles Form (PDF, 26KB)

The APA Ethics Office provides the full Ethical Principles of Psychologists and Code of Conduct electronically on its website in HTML, PDF, and Word format. You may also request a copy by emailing or calling the APA Ethics Office (202-336-5930). You may also read "Ethical Principles," December 1992, *American Psychologist*, Vol. 47, pp. 1597–1611.

**Other information**
Visit the Journals Publishing Resource Center for more resources for writing, reviewing, and editing articles for publishing in APA journals.

Appendix B: Prospero Application

Functional Family Therapy (FFT) for Adolescents with Antisocial Behaviours: Impacts on Mental Health Issues, Family Functioning and Recidivism

**Citation**
Debra Delaney, Dr Renate Thienel, and Dr Tanya Hanstock.
PROSPERO registration number: CRD42023451290
Available from: https://www.crd.york.ac.uk/prospero/

**Review questions**

1. Is FFT effective in improving mental health issues in adolescents with antisocial behaviour and comorbid mental health issues or diagnoses.

2. Is FFT effective in reducing recidivism and improving family functioning?

**Searches**
We have developed a search strategy to search for relevant research in a number of databases including Ovid hosted databases: APA PsycArticles, APA PsycINFO, Medline, and Embase; and EBSCO hosted databases: Academic Search Ultimate, Humanities Source Ultimate, Psychology and Behavioral Sciences Collection. Also searched were Proquest hosted databases: Social sciences, Psychology, and Criminal justice; as well as Scopus.

No restrictions were placed on the time period of publication, but searches were limited to those published in English, conducted with humans, and in peer reviewed journals. Databases were searched on 13 July 2021. The techniques of 'snowballing' and 'pearling' will also be used to assist in identifying all relevant research. Unpublished studies will not be sought.

**Types of study to be included**
Randomised controlled trials and quasi-experimental studies will be included in this Systematic Review.

**Condition or domain being studied**
We are interested in the mental health of adolescents who engage in antisocial behaviours and take our definitions of antisocial behaviours from the American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR), criteria on Conduct Disorder and Oppositional Defiance Disorder, as well as the International Classification of Diseases 11th Revision (World Health Organisation). The over-arching common theme of CD type behaviours is the violation of the basic rights of others or age-appropriate societal norms, rules, or laws. This can include aggression to people or animals, destruction of property, and deceitfulness or theft. ODD type behaviours include argumentative/defiant behaviour, or vindictiveness that can be present in multiple settings and impair social functioning. Antisocial behaviours can also involve running away from home, school truancy, and alcohol or other drug (AOD) use (APA, 2022).

Mental Health domains include mood, anxiety, or any behaviour, emotion, and social domains that contribute to impairment as stated in the DSM-5-TR.

**Inclusion/Exclusion Criteria**

Studies were included if they:

1) Described an FFT intervention applied to the broad range of antisocial behaviours.
2) Were published in English in a peer reviewed academic journal.
3) Distinctly used an FFT therapeutic technique.
4) Were delivered to any type of adolescents including those with mental health diagnoses.

Studies were excluded if they:

1) Described a hypothetical or non FFT intervention.
2) Described an intervention that is a variation of FFT such as FFT-CW (child welfare).
3) Were delivered to a population younger than 11 years of age or older than 18 years of age.
4) Were conference abstracts or unpublished work such as doctoral theses.

**Intervention(s), exposure(s)**

Functional Family Therapy (FFT)

FFT is an intervention program targeting families with an adolescent demonstrating antisocial behaviours. The intervention consists of 12–14, one-hour, weekly sessions with a therapist and can be conducted in both clinic and home settings. Source: https://fftllc.com/

**Comparison/control groups**

The comparison or control groups used in the reviewed studies vary. They include no treatment control groups, treatment as usual groups (such as probation and parole services), or an alternate group, family, or individual therapy treatment.

**Main outcome(s)**

We have not restricted our search strategy by a specific outcome variable as we are interested in all outcomes relating directly and indirectly to reductions in, and prevention of antisocial behaviours which may include improvements in externalising behaviours, and family functioning.

Outcomes may be measured using self-report, other observer report such as parents or teachers, or measures of wellbeing such as the Strengths and Needs Assessment or Strengths and Difficulties Questionnaire, as well as objective measures, such as criminal justice system contact records.

**Measures of effect**

Measures of effect may include odds ratios, relative risks, and standardised mean difference where possible.

**Data extraction (selection and coding)**

*Study selection*

Two researchers will apply the eligibility criteria to screen titles and abstracts of potentially relevant studies for inclusion. Full text reviews will be conducted by one researcher. Disagreements over inclusion will be resolved through discussion with one or more additional researchers. Each researcher's decisions will be tracked using Covidence software for managing systematic reviews.

*Data extraction*
Data extraction will be completed by one researcher, with a second reviewer checking 20% of extractions. Disagreements over extraction will be resolved through discussion with one or more additional researchers. Each researcher's decisions will be tracked using Covidence software for managing systematic reviews.

Data to be extracted from each study include study year and authors, design, aims, program length, location, sample size and age range, sample characteristics, ethnicity, sex, comparison group description, outcome variable, outcome measure, results, and quality rating.

**Risk of bias (quality) assessment**
Assessment of risk of bias in randomised controlled trials will be done using the Cochrane Collaboration Risk of Bias (ROB) Assessment 2.0 Tool and risk of bias for non-randomised studies will be done with the Cochrane Collaboration Risk of Bias in of Interventions (ROBINS). Two researchers will independently assess risk of bias for each study. Disagreements will be resolved by third researcher.

**Strategy for data synthesis**
A narrative synthesis will be conducted, with greater emphasis given to studies with lower risk of bias.

**Analysis of subgroups or subsets**
Where possible given the number of relevant studies identified, we will analyse outcomes by antisocial behaviour type, e.g., interventions targeting adolescents in contact with the justice system, or problematic alcohol or substance use. Where possible we will stratify our analysis within these groups according to sample size, sex, ethnicity, and comparison group characteristics, as well as risk of bias (e.g., low vs. medium risk).

**Contact details for further information**
Debra Delaney: Debra.Delaney@uon.edu.au

**Organisational affiliation of the review**
University of Newcastle https://www.newcastle.edu.au/

**Review team members and their organisational affiliations**
1. Debra Delaney, student, Master of Clinical Psychology, University of Newcastle.
2. Dr Renate Thienel, Research Associate, School of Medicine and Public Health, University of Newcastle.
3. Dr Tanya Hanstock, Senior Lecturer, School of Psychological Sciences, University of Newcastle.

**Type and method of review**
Systematic review

**Anticipated or actual start date**
13 July 2023

**Anticipated completion date**
1 December 2023

**Funding sources/sponsors**
This Systematic Review has no funding sources.

**Conflicts of interest**
Debra Delaney worked for a FFT program for three years and is aware of any potential bias from this. The other authors have no conflicts of interest to declare.

**Language**
English

**Country**
Australia

**Stage of review**
Review Ongoing

**Subject index terms status**
Subject indexing assigned by CRD

**Subject index terms**
Adolescent; Antisocial Personality Disorder; Child; Child Abuse; Family Therapy; Humans; Longitudinal Studies

**Stage of review at time of this submission**

| Stage | Started | Completed |
|---|---|---|
| Preliminary searches | Yes | Yes |
| Piloting of the study selection process | Yes | No |
| Formal screening of search results against eligibility criteria | Yes | No |
| Data extraction | No | No |
| Risk of bias (quality) assessment | No | No |
| Data analysis | No | No |

*The record owner confirms that the information they have supplied for this submission is accurate and complete and they understand that deliberate provision of inaccurate information or omission of data may be construed as scientific misconduct.*
*The record owner confirms that they will update the status of the review when it is completed and will add publication details in due course.*

# Appendix C: Revised Cochrane Risk-of-Bias Tool for Randomised Trials (RoB 2)

## Revised Cochrane risk-of-bias tool for randomized trials (RoB 2)
## SHORT VERSION (CRIBSHEET)

Edited by Julian PT Higgins, Jelena Savović, Matthew J Page, Jonathan AC Sterne
on behalf of the RoB 2 Development Group

**Version of 22 August 2019**

The development of the RoB 2 tool was supported by the MRC Network of Hubs for Trials Methodology Research (MR/L004933/2- N61), with the support of the host MRC ConDuCT-II Hub (Collaboration and innovation for Difficult and Complex randomised controlled Trials In Invasive procedures - MR/K025643/1), by MRC research grant MR/M025209/1, and by a grant from The Cochrane Collaboration.

Preliminary considerations

**Study design**

- ☐ Individually-randomized parallel-group trial
- ☐ Cluster-randomized parallel-group trial
- ☐ Individually randomized cross-over (or other matched) trial

**For the purposes of this assessment, the interventions being compared are defined as**

Experimental: Comparator:

**Specify which outcome is being assessed for risk of bias**

**Specify the numerical result being assessed.** In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed.

**Is the review team's aim for this result…?**

- ☐ to assess the effect of *assignment to intervention* (the 'intention-to-treat' effect)
- ☐ to assess the effect of *adhering to intervention* (the 'per-protocol' effect)

**If the aim is to assess the effect of *adhering to intervention*,** select the deviations from intended intervention that should be addressed (at least one must be checked):

- ☐ occurrence of non-protocol interventions
- ☐ failures in implementing the intervention that could have affected the outcome
- ☐ non-adherence to their assigned intervention by trial participants

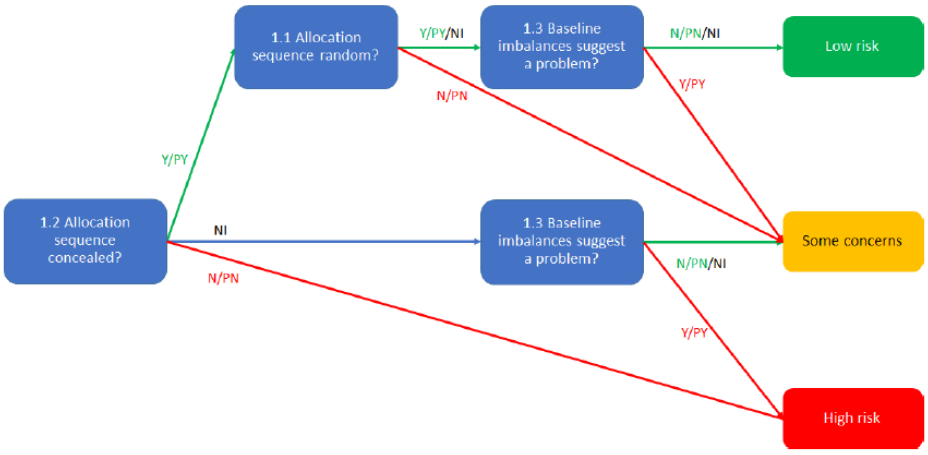
**Which of the following sources were <u>obtained</u> to help inform the risk-of-bias assessment? (tick as many as apply)**

- ☐ Journal article(s)
- ☐ Trial protocol
- ☐ Statistical analysis plan (SAP)
- ☐ Non-commercial trial registry record (e.g. ClinicalTrials.gov record)
- ☐ Company-owned trial registry record (e.g. GSK Clinical Study Register record)
- ☐ "Grey literature" (e.g. unpublished thesis)
- ☐ Conference abstract(s) about the trial
- ☐ Regulatory document (e.g. Clinical Study Report, Drug Approval Package)
- ☐ Research ethics application
- ☐ Grant database summary (e.g. NIH RePORTER or Research Councils UK Gateway to Research)
- ☐ Personal communication with trialist
- ☐ Personal communication with the sponsor

Domain 1: Risk of bias arising from the randomization process

| Signalling questions | Elaboration | Response options |
|---|---|---|
| **1.1 Was the allocation sequence random?** | Answer 'Yes' if a random component was used in the sequence generation process. Examples include computer-generated random numbers; reference to a random number table; coin tossing; shuffling cards or envelopes; throwing dice; or drawing lots. Minimization is generally implemented with a random element (at least when the scores are equal), so an allocation sequence that is generated using minimization should generally be considered to be random.<br><br>Answer 'No' if no random element was used in generating the allocation sequence or the sequence is predictable. Examples include alternation; methods based on dates (of birth or admission); patient record numbers; allocation decisions made by clinicians or participants; allocation based on the availability of the intervention; or any other systematic or haphazard method.<br><br>Answer 'No information' if the only information about randomization methods is a statement that the study is randomized.<br><br>In some situations a judgement may be made to answer 'Probably no' or 'Probably yes'. For example, , in the context of a large trial run by an experienced clinical trials unit, absence of specific information about generation of the randomization sequence, in a paper published in a journal with rigorously enforced word count limits, is likely to result in a response of 'Probably yes' rather than 'No information'. Alternatively, if other (contemporary) trials by the same investigator team have clearly used non-random sequences, it might be reasonable to assume that the current study was done using similar methods. | Y/PY/PN/N/NI |
| **1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?** | Answer 'Yes' if the trial used any form of remote or centrally administered method to allocate interventions to participants, where the process of allocation is controlled by an external unit or organization, independent of the enrolment personnel (e.g. independent central pharmacy, telephone or internet-based randomization service providers).<br><br>Answer 'Yes' if envelopes or drug containers were used appropriately. Envelopes should be opaque, sequentially numbered, sealed with a tamper-proof seal and opened only after the envelope has been irreversibly assigned to the participant. Drug containers should be sequentially numbered and of identical appearance, and dispensed or administered only after they have been irreversibly assigned to the participant. This level of detail is rarely provided in reports, and a judgement may be required to justify an answer of 'Probably yes' or 'Probably no'.<br><br>Answer 'No' if there is reason to suspect that the enrolling investigator or the participant had knowledge of the forthcoming allocation. | Y/PY/PN/N/NI |
| **1.3 Did baseline differences between intervention groups suggest a problem with the randomization process?** | *Note that differences that are compatible with chance do not lead to a risk of bias. A small number of differences identified as 'statistically significant' at the conventional 0.05 threshold should usually be considered to be compatible with chance.*<br><br>Answer 'No' if no imbalances are apparent or if any observed imbalances are compatible with chance.<br><br>Answer 'Yes' if there are imbalances that indicate problems with the randomization process, including:<br><br>(1) substantial differences between intervention group sizes, compared with the intended allocation ratio;<br>or<br>(2) a substantial excess in statistically significant differences in baseline characteristics between intervention groups, beyond that expected by chance; or<br>(3) imbalance in one or more key prognostic factors, or baseline measures of outcome variables, that is very unlikely to be due to chance and for which the between-group difference is big enough to result in bias in the intervention effect estimate.<br><br>Also answer 'Yes' if there are other reasons to suspect that the randomization process was problematic:<br><br>(4) excessive similarity in baseline characteristics that is not compatible with chance.<br><br>Answer 'No information' when there is no *useful* baseline information available (e.g. abstracts, or studies that reported only baseline characteristics of participants in the final analysis).<br><br>The answer to this question should not influence answers to questions 1.1 or 1.2. For example, if the trial has large baseline imbalances, but authors report adequate randomization methods, questions 1.1 and 1.2 should still be answered on the basis of the reported adequate methods, and any concerns about the imbalance should be raised in the answer to the question 1.3 and reflected in the domain-level risk-of-bias judgement.<br><br>Trialists may undertake analyses that attempt to deal with flawed randomization by controlling for imbalances in prognostic factors at baseline. To remove the risk of bias caused by problems in the randomization process, it would be necessary to know, and measure, all the prognostic factors that were imbalanced at baseline. It is unlikely that all important prognostic factors are known and measured, so such analyses will at best reduce the risk of bias. If review authors wish to assess the risk of bias in a trial that controlled for baseline imbalances in order to mitigate failures of randomization, the study should be assessed using the ROBINS-I tool. | Y/PY/PN/N/NI |
| **Risk-of-bias judgement** | See algorithm. | Low / High / Some concerns |

| | | |
|---|---|---|
| Optional: What is the predicted direction of bias arising from the randomization process? | If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions. | NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |



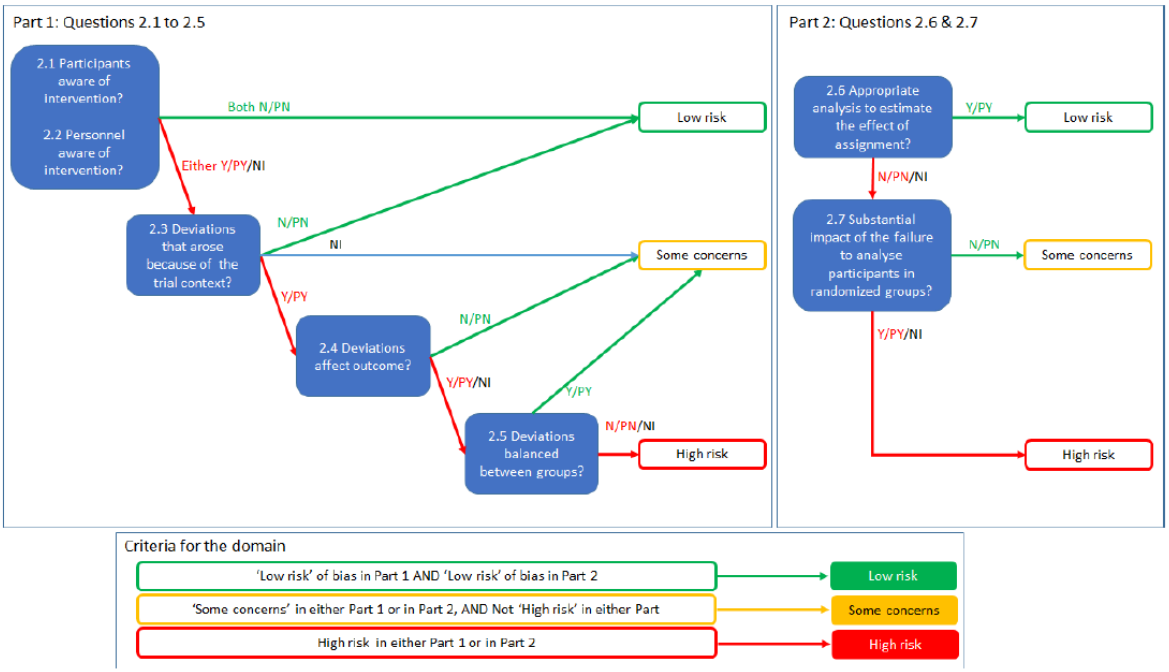

**Algorithm for suggested judgement of risk of bias arising from the randomization process**

Domain 2: Risk of bias due to deviations from the intended interventions (*effect of assignment to intervention*)

| Signalling questions | Elaboration | Response options |
|---|---|---|
| **2.1. Were participants aware of their assigned intervention during the trial?** | If participants are aware of their assigned intervention it is more likely that health-related behaviours will differ between the intervention groups. Blinding participants, most commonly through use of a placebo or sham intervention, may prevent such differences. If participants experienced side effects or toxicities that they knew to be specific to one of the interventions, answer this question 'Yes' or 'Probably yes'. | Y/PY/PN/N/NI |
| **2.2. Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?** | If carers or people delivering the interventions are aware of the assigned intervention then its implementation, or administration of non-protocol interventions, may differ between the intervention groups. Blinding may prevent such differences. If participants experienced side effects or toxicities that carers or people delivering the interventions knew to be specific to one of the interventions, answer question 'Yes' or 'Probably yes'. If randomized allocation was not concealed, then it is likely that carers and people delivering the interventions were aware of participants' assigned intervention during the trial. | Y/PY/PN/N/NI |

| | | |
|---|---|---|
| **2.3. If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention that arose because of the trial context?** | For the effect of assignment to intervention, this domain assesses problems that arise when changes from assigned intervention that are inconsistent with the trial protocol arose because of the trial context. We use the term **trial context** to refer to effects of recruitment and engagement activities on trial participants and when trial personnel (carers or people delivering the interventions) undermine the implementation of the trial protocol in ways that would not happen outside the trial. For example, the process of securing informed consent may lead participants subsequently assigned to the comparator group to feel unlucky and therefore seek the experimental intervention, or other interventions that improve their prognosis.<br><br>Answer 'Yes' or 'Probably yes' **only** if there is evidence, or strong reason to believe, that the trial context led to failure to implement the protocol interventions or to implementation of interventions not allowed by the protocol.<br><br>Answer 'No' or 'Probably no' if there were changes from assigned intervention that are inconsistent with the trial protocol, such as non-adherence to intervention, but these are consistent with what could occur outside the trial context.<br><br>Answer 'No' or 'Probably no' for changes to intervention that are consistent with the trial protocol, for example cessation of a drug intervention because of acute toxicity or use of additional interventions whose aim is to treat consequences of one of the intended interventions.<br><br>If blinding is compromised because participants report side effects or toxicities that are specific to one of the interventions, answer 'Yes' or 'Probably yes' only if there were changes from assigned intervention that are inconsistent with the trial protocol and arose because of the trial context.<br><br>The answer 'No information' may be appropriate, because trialists do not always report whether deviations arose because of the trial context. | NA/Y/PY/PN/N/NI |
| **2.4 If Y/PY to 2.3: Were these deviations likely to have affected the outcome?** | Changes from assigned intervention that are inconsistent with the trial protocol and arose because of the trial context will impact on the intervention effect estimate if they affect the outcome, but not otherwise. | NA/Y/PY/PN/N/NI |

| | | |
|---|---|---|
| **2.5. If Y/PY/NI to 2.4: Were these deviations from intended intervention balanced between groups?** | Changes from assigned intervention that are inconsistent with the trial protocol and arose because of the trial context are more likely to impact on the intervention effect estimate if they are not balanced between the intervention groups. | NA/Y/PY/PN/N/NI |
| **2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?** | Both intention-to-treat (ITT) analyses and modified intention-to-treat (mITT) analyses excluding participants with missing outcome data should be considered appropriate. Both naïve 'per-protocol' analyses (excluding trial participants who did not receive their assigned intervention) and 'as treated' analyses (in which trial participants are grouped according to the intervention that they received, rather than according to their assigned intervention) should be considered inappropriate. Analyses excluding eligible trial participants post-randomization should also be considered inappropriate, but post-randomization exclusions of ineligible participants (when eligibility was not confirmed until after randomization, and could not have been influenced by intervention group assignment) can be considered appropriate. | Y/PY/PN/N/NI |
| **2.7 If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomized?** | This question addresses whether the number of participants who were analysed in the wrong intervention group, or excluded from the analysis, was sufficient that there could have been a substantial impact on the result. It is not possible to specify a precise rule: there may be potential for substantial impact even if fewer than 5% of participants were analysed in the wrong group or excluded, if the outcome is rare or if exclusions are strongly related to prognostic factors. | NA/Y/PY/PN/N/NI |
| **Risk-of-bias judgement** | See algorithm. | Low / High / Some concerns |
| Optional: What is the predicted direction of bias due to deviations from intended interventions? | If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions. | NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

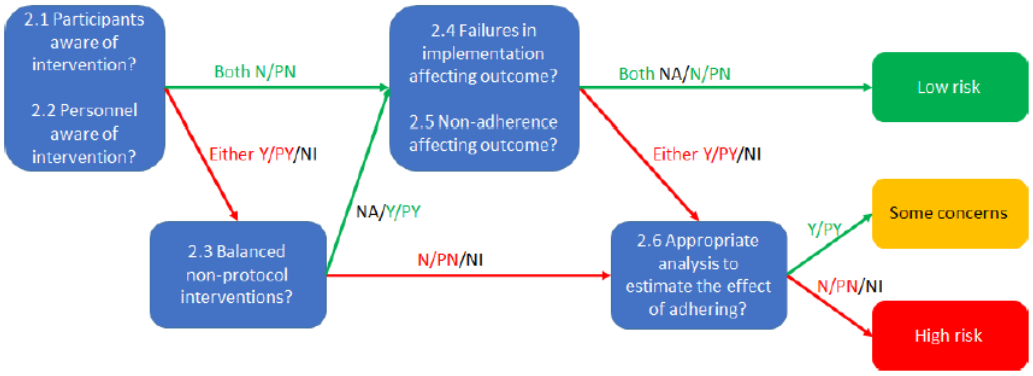



**Algorithm for suggested judgement of risk of bias due to deviations from the intended interventions (effect of assignment to intervention)**

Domain 2: Risk of bias due to deviations from the intended interventions (*effect of adhering to intervention*)

| Signalling questions | Elaboration | Response options |
|---|---|---|
| **2.1. Were participants aware of their assigned intervention during the trial?** | If participants are aware of their assigned intervention it is more likely that health-related behaviours will differ between the intervention groups. Blinding participants, most commonly through use of a placebo or sham intervention, may prevent such differences. If participants experienced side effects or toxicities that they knew to be specific to one of the interventions, answer this question 'Yes' or 'Probably yes'. | Y/PY/PN/N/NI |
| **2.2. Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?** | If carers or people delivering the interventions are aware of the assigned intervention then its implementation, or administration of non-protocol interventions, may differ between the intervention groups. Blinding may prevent such differences. If participants experienced side effects or toxicities that carers or people delivering the interventions knew to be specific to one of the interventions, answer 'Yes' or 'Probably yes'. If randomized allocation was not concealed, then it is likely that carers and people delivering the interventions were aware of participants' assigned intervention during the trial. | Y/PY/PN/N/NI |
| **2.3. [If applicable:] If Y/PY/NI to 2.1 or 2.2: Were important non-protocol interventions balanced across intervention groups?** | This question is asked only if the preliminary considerations specify that the assessment will address imbalance of important non-protocol interventions between intervention groups. Important non-protocol interventions are the additional interventions or exposures that: (1) are inconsistent with the trial protocol; (2) trial participants might receive with or after starting their assigned intervention; and (3) are prognostic for the outcome. Risk of bias will be higher if there is imbalance in such interventions between the intervention groups. | NA/Y/PY/PN/N/NI |
| **2.4. [If applicable:] Were there failures in implementing the intervention that could have affected the outcome?** | This question is asked only if the preliminary considerations specify that the assessment will address failures in implementing the intervention that could have affected the outcome. Risk of bias will be higher if the intervention was not implemented as intended by, for example, the health care professionals delivering care. Answer 'No' or 'Probably no' if implementation of the intervention was successful for most participants. | NA/Y/PY/PN/N/NI |
| **2.5. [If applicable:] Was there non-adherence to the assigned intervention regimen that could have affected participants' outcomes?** | This question is asked only if the preliminary considerations specify that the assessment will address non-adherence that could have affected participants' outcomes. Non-adherence includes imperfect compliance with a sustained intervention, cessation of intervention, crossovers to the comparator intervention and switches to another active intervention. Consider available information on the proportion of study participants who continued with their assigned intervention throughout follow up, and answer 'Yes' or 'Probably yes' if the proportion who did not adhere is high enough to raise concerns. Answer 'No' for studies of interventions that are administered once, so that imperfect adherence is not possible, and all or most participants received the assigned intervention. | NA/Y/PY/PN/N/NI |

| | | |
|---|---|---|
| **2.6. If N/PN/NI to 2.3, or Y/PY/NI to 2.4 or 2.5: Was an appropriate analysis used to estimate the effect of adhering to the intervention?** | Both ' naïve 'per-protocol' analyses (excluding trial participants who did not receive their allocated intervention) and 'as treated' analyses (comparing trial participants according to the intervention they actually received) will usually be inappropriate for estimating the effect of adhering to intervention (the 'per-protocol' effect). However, it is possible to use data from a randomized trial to derive an unbiased estimate of the effect of adhering to intervention. Examples of appropriate methods include: (1) instrumental variable analyses to estimate the effect of receiving the assigned intervention in trials in which a single intervention, administered only at baseline and with all-or-nothing adherence, is compared with standard care; and (2) inverse probability weighting to adjust for censoring of participants who cease adherence to their assigned intervention, in trials of sustained treatment strategies. These methods depend on strong assumptions, which should be appropriate and justified if the answer to this question is 'Yes' or 'Probably yes'. It is possible that a paper reports an analysis based on such methods without reporting information on the deviations from intended intervention, but it would be hard to judge such an analysis to be appropriate in the absence of such information.<br>If an important non-protocol intervention was administered to all participants in one intervention group, adjustments cannot be made to overcome this.<br>Some examples of analysis strategies that would not be appropriate to estimate the effect of adhering to intervention are (i) 'Intention to treat (ITT) analysis', (ii) 'per protocol analysis', (iii) 'as-treated analysis', (iv) 'analysis by treatment received'. | NA/Y/PY/PN/N/NI |
| **Risk-of-bias judgement** | See algorithm. | Low / High / Some concerns |
| Optional: What is the predicted direction of bias due to deviations from intended interventions? | If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions. | NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

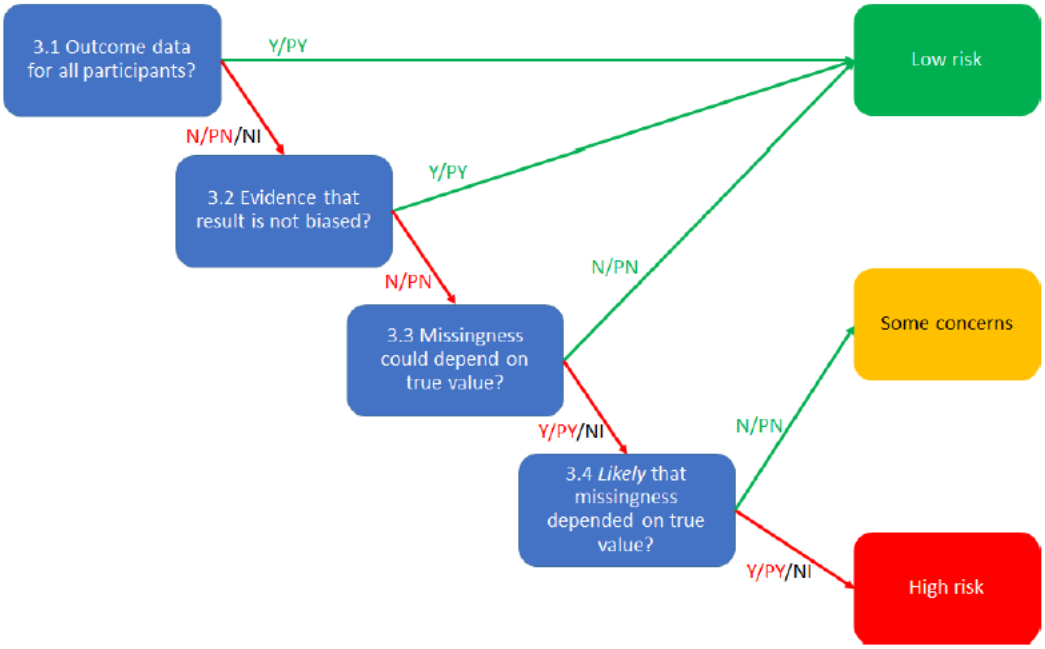



**Algorithm for suggested judgement of risk of bias due to deviations from the intended interventions (effect of adhering to intervention)**

Domain 3: Risk of bias due to missing outcome data

| Signalling questions | Elaboration | Response options |
|---|---|---|
| **3.1 Were data for this outcome available for all, or nearly all, participants randomized?** | The appropriate study population for an analysis of the intention to treat effect is all randomized participants.<br>"Nearly all" should be interpreted as that the number of participants with missing outcome data is sufficiently small that their outcomes, whatever they were, could have made no important difference to the estimated effect of intervention.<br>For continuous outcomes, availability of data from 95% of the participants will often be sufficient. For dichotomous outcomes, the proportion required is directly linked to the risk of the event. If the observed number of events is much greater than the number of participants with missing outcome data, the bias would necessarily be small.<br>Only answer 'No information' if the trial report provides no information about the extent of missing outcome data. This situation will usually lead to a judgement that there is a high risk of bias due to missing outcome data.<br>Note that imputed data should be regarded as missing data, and not considered as 'outcome data' in the context of this question. | Y/PY/PN/N/NI |
| **3.2 If N/PN/NI to 3.1: Is there evidence that the result was not biased by missing outcome data?** | Evidence that the result was not biased by missing outcome data may come from: (1) analysis methods that correct for bias; or (2) sensitivity analyses showing that results are little changed under a range of plausible assumptions about the relationship between missingness in the outcome and its true value. However, imputing the outcome variable, either through methods such as 'last-observation-carried-forward' or via multiple imputation based only on intervention group, should not be assumed to correct for bias due to missing outcome data. | NA/Y/PY/PN/N |
| **3.3 If N/PN to 3.2: Could missingness in the outcome depend on its true value?** | If loss to follow up, or withdrawal from the study, could be related to participants' health status, then it is possible that missingness in the outcome was influenced by its true value. However, if all missing outcome data occurred for documented reasons that are unrelated to the outcome then the risk of bias due to missing outcome data will be low (for example, failure of a measuring device or interruptions to routine data collection).<br>In time-to-event analyses, participants censored during trial follow-up, for example because they withdrew from the study, should be regarded as having missing outcome data, even though some of their follow up is included in the analysis. Note that such participants may be shown as included in analyses in CONSORT flow diagrams. | NA/Y/PY/PN/N/NI |

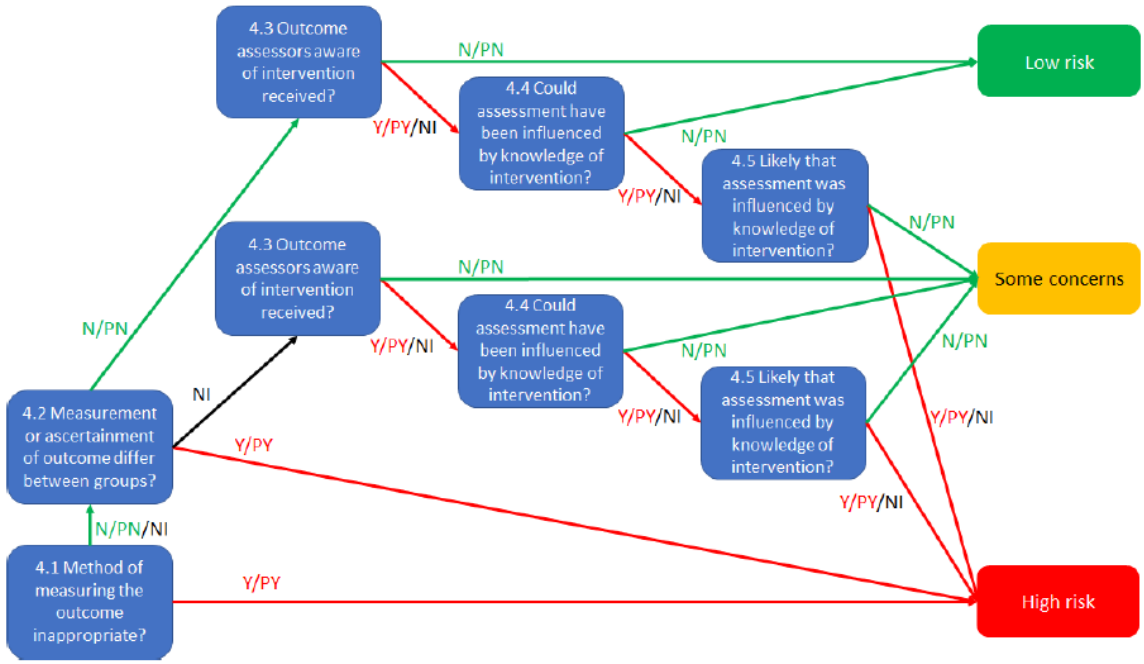
| | | |
|---|---|---|
| **3.4 If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?** | This question distinguishes between situations in which (i) missingness in the outcome could depend on its true value (assessed as 'Some concerns') from those in which (ii) it is likely that missingness in the outcome depended on its true value (assessed as 'High risk of bias'). Five reasons for answering 'Yes' are:<br>1. Differences between intervention groups in the proportions of missing outcome data. If there is a difference between the effects of the experimental and comparator interventions on the outcome, and the missingness in the outcome is influenced by its true value, then the proportions of missing outcome data are likely to differ between intervention groups. Such a difference suggests a risk of bias due to missing outcome data, because the trial result will be sensitive to missingness in the outcome being related to its true value. For time-to-event-data, the analogue is that rates of censoring (loss to follow-up) differ between the intervention groups.<br>2. Reported reasons for missing outcome data provide evidence that missingness in the outcome depends on its true value;<br>3. Reported reasons for missing outcome data differ between the intervention groups;<br>4. The circumstances of the trial make it likely that missingness in the outcome depends on its true value. For example, in trials of interventions to treat schizophrenia it is widely understood that continuing symptoms make drop out more likely.<br>5. In time-to-event analyses, participants' follow up is censored when they stop or change their assigned intervention, for example because of drug toxicity or, in cancer trials, when participants switch to second-line chemotherapy.<br>Answer 'No' if the analysis accounted for participant characteristics that are likely to explain the relationship between missingness in the outcome and its true value. | NA/Y/PY/PN/N/NI |
| **Risk-of-bias judgement** | See algorithm. | Low / High / Some concerns |
| Optional: What is the predicted direction of bias due to missing outcome data? | If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions. | NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |




**Algorithm for suggested judgement of risk of bias due to missing outcome data**

Domain 4: Risk of bias in measurement of the outcome

| Signalling questions | Elaboration | Response options |
|---|---|---|
| **4.1 Was the method of measuring the outcome inappropriate?** | This question aims to identify methods of outcome measurement (data collection) that are unsuitable for the outcome they are intended to evaluate. The question *does not* aim to assess whether the choice of outcome being evaluated was sensible (e.g. because it is a surrogate or proxy for the main outcome of interest). In most circumstances, for pre-specified outcomes, the answer to this question will be 'No' or 'Probably no'.<br>Answer 'Yes' or 'Probably yes' if the method of measuring the outcome is inappropriate, for example because:<br>(1) it is unlikely to be sensitive to plausible intervention effects (e.g. important ranges of outcome values fall outside levels that are detectable using the measurement method); or<br>(2) the measurement instrument has been demonstrated to have poor validity. | Y/PY/PN/N/NI |
| **4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?** | Comparable methods of outcome measurement (data collection) involve the same measurement methods and thresholds, used at comparable time points. Differences between intervention groups may arise because of 'diagnostic detection bias' in the context of passive collection of outcome data, or if an intervention involves additional visits to a healthcare provider, leading to additional opportunities for outcome events to be identified. | Y/PY/PN/N/NI |
| **4.3 If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware of the intervention received by study participants?** | Answer 'No' if outcome assessors were blinded to intervention status. For participant-reported outcomes, the outcome assessor is the study participant. | NA/Y/PY/PN/N/NI |
| **4.4 If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?** | Knowledge of the assigned intervention could influence participant-reported outcomes (such as level of pain), observer-reported outcomes involving some judgement, and intervention provider decision outcomes. They are unlikely to influence observer-reported outcomes that do not involve judgement, for example all-cause mortality. | NA/Y/PY/PN/N/NI |
| **4.5 If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?** | This question distinguishes between situations in which (i) knowledge of intervention status could have influenced outcome assessment but there is no reason to believe that it did (assessed as 'Some concerns') from those in which (ii) knowledge of intervention status was likely to influence outcome assessment (assessed as 'High'). When there are strong levels of belief in either beneficial or harmful effects of the intervention, it is more likely that the outcome was influenced by knowledge of the intervention received. Examples may include patient-reported symptoms in trials of homeopathy, or assessments of recovery of function by a physiotherapist who delivered the intervention. | NA/Y/PY/PN/N/NI |
| **Risk-of-bias judgement** | See algorithm. | Low / High / Some concerns |
| Optional: What is the predicted direction of bias in measurement of the outcome? | If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions. | NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

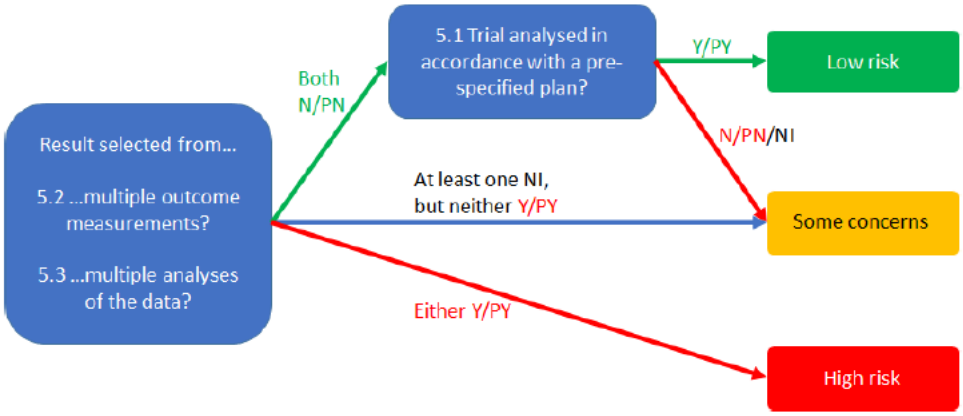



**Algorithm for suggested judgement of risk of bias in measurement of the outcome**

Domain 5: Risk of bias in selection of the reported result

| Signalling questions | Elaboration | Response options |
|---|---|---|
| **5.1 Were the data that produced this result analysed in accordance with a pre-specified analysis plan that was finalized before unblinded outcome data were available for analysis?** | If the researchers' pre-specified intentions are available in sufficient detail, then planned outcome measurements and analyses can be compared with those presented in the published report(s). To avoid the possibility of selection of the reported result, finalization of the analysis intentions must precede availability of unblinded outcome data to the trial investigators.<br>Changes to analysis plans that were made before unblinded outcome data were available, or that were clearly unrelated to the results (e.g. due to a broken machine making data collection impossible) do not raise concerns about bias in selection of the reported result. | Y/PY/PN/N/NI |
| **Is the numerical result being assessed likely to have been selected, on the basis of the results, from...** | | |
| **5.2. ... multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?** | A particular outcome domain (i.e. a true state or endpoint of interest) may be **measured** in multiple ways. For example, the domain pain may be measured using multiple scales (e.g. a visual analogue scale and the McGill Pain Questionnaire), each at multiple time points (e.g. 3, 6 and 12 weeks post-treatment). If multiple measurements were made, but only one or a subset is reported on the basis of the results (e.g. statistical significance), there is a high risk of bias in the fully reported result. Attention should be restricted to outcome measurements that are eligible for consideration by the RoB 2 tool user. For example, if only a result using a specific measurement scale is eligible for inclusion in a meta-analysis (e.g. Hamilton Depression Rating Scale), and this is reported by the trial, then there would not be an issue of selection even if this result was reported (on the basis of the results) in preference to the result from a different measurement scale (e.g. Beck Depression Inventory).<br>Answer 'Yes' or 'Probably yes' if:<br>There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that a domain was measured in multiple eligible ways, but data for only one or a subset of measures is fully reported (without justification), and the fully reported result is likely to have been selected on the basis of the results. Selection on the basis of the results can arise from a desire for findings to be newsworthy, sufficiently noteworthy to merit publication, or to confirm a prior hypothesis. For example, trialists who have a preconception, or vested interest in showing, that an experimental intervention is beneficial may be inclined to report outcome measurements selectively that are favourable to the experimental intervention.<br>Answer 'No' or 'Probably no' if:<br>There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that all eligible reported results for the outcome domain correspond to all intended outcome measurements.<br>or<br>There is only one possible way in which the outcome domain can be measured (hence there is no opportunity to select from multiple measures).<br>or<br>Outcome measurements are inconsistent across different reports on the same trial, but the trialists have provided the reason for the inconsistency and it is not related to the nature of the results.<br>Answer 'No information' if:<br>Analysis intentions are not available, or the analysis intentions are not reported in sufficient detail to enable an assessment, and there is more than one way in which the outcome domain could have been measured. | Y/PY/PN/N/NI |
| **5.3 ... multiple eligible analyses of the data?** | A particular outcome measurement may be analysed in multiple ways. Examples include: unadjusted and adjusted models; final value vs change from baseline vs analysis of covariance; transformations of variables; different definitions of composite outcomes (e.g. 'major adverse event'); conversion of continuously scaled outcome to categorical data with different cut-points; different sets of covariates for adjustment; and different strategies for dealing with missing data. Application of multiple methods generates multiple effect estimates for a specific outcome measurement. If multiple estimates are generated but only one or a subset is reported on the basis of the results (e.g. statistical significance), there is a high risk of bias in the fully reported result. Attention should be restricted to analyses that are eligible for consideration by the RoB 2 tool user. For example, if only the result from an analysis of post-intervention values is eligible for inclusion in a meta-analysis (e.g. at 12 weeks after randomization), and this is reported by the trial, then there would not be an issue of selection even if this result was reported (on the basis of the results) in preference to the result from an analysis of changes from baseline.<br>Answer 'Yes' or 'Probably yes' if: | Y/PY/PN/N/NI |

| | | |
|---|---|---|
| | There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that a measurement was analysed in multiple eligible ways, but data for only one or a subset of analyses is fully reported (without justification), and the fully reported result is likely to have been selected on the basis of the results. Selection on the basis of the results arises from a desire for findings to be newsworthy, sufficiently noteworthy to merit publication, or to confirm a prior hypothesis. For example, trialists who have a preconception or vested interest in showing that an experimental intervention is beneficial may be inclined to selectively report analyses that are favourable to the experimental intervention.<br>Answer 'No' or 'Probably no' if:<br>There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that all eligible reported results for the outcome measurement correspond to all intended analyses.<br>or<br>There is only one possible way in which the outcome measurement can be analysed (hence there is no opportunity to select from multiple analyses).<br>or<br>Analyses are inconsistent across different reports on the same trial, but the trialists have provided the reason for the inconsistency and it is not related to the nature of the results.<br>Answer 'No information' if:<br>Analysis intentions are not available, or the analysis intentions are not reported in sufficient detail to enable an assessment, and there is more than one way in which the outcome measurement could have been analysed. | |
| **Risk-of-bias judgement** | See algorithm. | Low / High / Some concerns |
| Optional: What is the predicted direction of bias due to selection of the reported result? | If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions. | NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |




**Algorithm for suggested judgement of risk of bias in selection of the reported result**

Overall risk of bias

| | | |
|---|---|---|
| **Risk-of-bias judgement** | | Low / High / Some concerns |
| Optional: What is the overall predicted direction of bias for this outcome? | | Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable / NA |

| **Overall risk-of-bias judgement** | **Criteria** |
|---|---|
| Low risk of bias | The study is judged to be at **low risk of bias for all domains** for this result. |
| Some concerns | The study is judged to raise **some concerns** in at least one domain for this result, but not to be at high risk of bias for any domain. |
| High risk of bias | The study is judged to be at **high risk of bias** in at least one domain for this result.<br>Or<br>The study is judged to have **some concerns** for **multiple domains** in a way that substantially lowers confidence in the result. |

The development of the RoB 2 tool was supported by the MRC Network of Hubs for Trials Methodology Research (MR/L004933/2- N61), with the support of the host MRC ConDuCT-II Hub (Collaboration and innovation for Difficult and Complex randomised controlled Trials In Invasive procedures - MR/K025643/1), by MRC research grant MR/M025209/1, and by a grant from The Cochrane Collaboration.

# Appendix D: The Risk of Bias in Non-Randomised Studies – of Interventions (ROBINS-I) Assessment Tool

## The Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS-I) assessment tool

(version for cohort-type studies)

Developed by: Jonathan AC Sterne, Miguel A Hernán, Barnaby C Reeves, Jelena Savović, Nancy D Berkman, Meera Viswanathan, David Henry, Douglas G Altman, Mohammed T Ansari, Isabelle Boutron, James Carpenter, An-Wen Chan, Rachel Churchill, Asbjørn Hróbjartsson, Jamie Kirkham, Peter Jüni, Yoon Loke, Terri Pigott, Craig Ramsay, Deborah Regidor, Hannah Rothstein, Lakhbir Sandhu, Pasqualina Santaguida, Holger J Schünemann, Beverly Shea, Ian Shrier, Peter Tugwell, Lucy Turner, Jeffrey C Valentine, Hugh Waddington, Elizabeth Waters, Penny Whiting and Julian PT Higgins

**Version 1 August 2016**



### ROBINS-I tool (Stage I): At protocol stage

Specify the review question

| | |
|---|---|
| Participants | |
| Experimental intervention | |
| Comparator | |
| Outcomes | |

List the confounding domains relevant to all or most studies

| |
|---|
| |

List co-interventions that could be different between intervention groups and that could impact on outcomes

| |
|---|
| |

### ROBINS-I tool (Stage II): For each study

Specify a target randomized trial specific to the study

| | |
|---|---|
| Design | Individually randomized / Cluster randomized / Matched (e.g. cross-over) |
| Participants | |
| Experimental intervention | |
| Comparator | |

Is your aim for this study...?

- ☐ to assess the effect of *assignment to* intervention
- ☐ to assess the effect of *starting and adhering to* intervention

Specify the outcome

Specify which outcome is being assessed for risk of bias (typically from among those earmarked for the Summary of Findings table). Specify whether this is a proposed benefit or harm of intervention.

| |
|---|
| |

Specify the numerical result being assessed

In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed.

| |
|---|
| |

Preliminary consideration of confounders

Complete a row for each important confounding domain (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as potentially important.

*"Important" confounding domains are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the intervention. "Validity" refers to whether the confounding variable or variables fully measure the domain, while "reliability" refers to the precision of the measurement (more measurement error means less reliability).*

| **(i) Confounding domains listed in the review protocol** | | | | |
|---|---|---|---|---|
| Confounding domain | Measured variable(s) | Is there evidence that controlling for this variable was unnecessary?* | Is the confounding domain measured validly and reliably by this variable (or these variables)? | OPTIONAL: Is failure to adjust for this variable (alone) expected to favour the experimental intervention or the comparator? |
| | | | Yes / No / No information | Favour experimental / Favour comparator / No information |
| | | | | |
| | | | | |
| | | | | |

| **(ii) Additional confounding domains relevant to the setting of this particular study, or which the study authors identified as important** | | | | |
|---|---|---|---|---|
| Confounding domain | Measured variable(s) | Is there evidence that controlling for this variable was unnecessary?* | Is the confounding domain measured validly and reliably by this variable (or these variables)? | OPTIONAL: Is failure to adjust for this variable (alone) expected to favour the experimental intervention or the comparator? |
| | | | Yes / No / No information | Favour experimental / Favour comparator / No information |
| | | | | |
| | | | | |
| | | | | |

* In the context of a particular study, variables can be demonstrated not to be confounders and so not included in the analysis: (a) if they are not predictive of the outcome; (b) if they are not predictive of intervention; or (c) because adjustment makes no or minimal difference to the estimated effect of the primary parameter. Note that "no statistically significant association" is not the same as "not predictive".

Preliminary consideration of co-interventions

Complete a row for each important co-intervention (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as important.

*"Important" co-interventions are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the intervention.*

| **(i) Co-interventions listed in the review protocol** | | |
|---|---|---|
| Co-intervention | Is there evidence that controlling for this co-intervention was unnecessary (e.g. because it was not administered)? | Is presence of this co-intervention likely to favour outcomes in the experimental intervention or the comparator |
| | | Favour experimental / Favour comparator / No information |
| | | Favour experimental / Favour comparator / No information |
| | | Favour experimental / Favour comparator / No information |

| **(ii) Additional co-interventions relevant to the setting of this particular study, or which the study authors identified as important** | | |
|---|---|---|
| Co-intervention | Is there evidence that controlling for this co-intervention was unnecessary (e.g. because it was not administered)? | Is presence of this co-intervention likely to favour outcomes in the experimental intervention or the comparator |
| | | Favour experimental / Favour comparator / No information |
| | | Favour experimental / Favour comparator / No information |
| | | Favour experimental / Favour comparator / No information |

Risk of bias assessment (cohort-type studies)

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

| Bias domain | Signalling questions | Elaboration | Response options |
|---|---|---|---|
| Bias due to confounding | 1.1 Is there potential for confounding of the effect of intervention in this study?<br>**If N/PN to 1.1**: the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered | In rare situations, such as when studying harms that are very unlikely to be related to factors that influence treatment decisions, no confounding is expected and the study can be considered to be at low risk of bias due to confounding, equivalent to a fully randomized trial. There is no NI (No information) option for this signalling question. | Y / PY / PN / N |
| | **If Y/PY to 1.1**: determine whether there is a need to assess time-varying confounding: | | |
| | 1.2. Was the analysis based on splitting participants' follow up time according to intervention received?<br>**If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6)<br>**If Y/PY**, proceed to question 1.3. | If participants could switch between intervention groups then associations between intervention and outcome may be biased by time-varying confounding. This occurs when prognostic factors influence switches between intended interventions. | NA / Y / PY / PN / N / NI |
| | 1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome?<br>**If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6)<br>**If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8) | If intervention switches are unrelated to the outcome, for example when the outcome is an unexpected harm, then time-varying confounding will not be present and only control for baseline confounding is required. | NA / Y / PY / PN / N / NI |
| | **Questions relating to baseline confounding only** | | |
| | 1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains? | Appropriate methods to control for measured confounders include stratification, regression, matching, standardization, and inverse probability weighting. They may control for individual variables or for the estimated propensity score. Inverse probability weighting is based on a function of the propensity score. Each method depends on the assumption that there is no unmeasured or residual confounding. | NA / Y / PY / PN / N / NI |
| | 1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study? | Appropriate control of confounding requires that the variables adjusted for are valid and reliable measures of the confounding domains. For some topics, a list of valid and reliable measures of confounding domains will be specified in the review protocol but for others such a list may not be available. Study authors may cite references to support the use of a particular measure. If authors control for confounding variables with no indication of their validity or reliability pay attention to the subjectivity of the measure. Subjective measures (e.g. based on self-report) may have lower validity and reliability than objective measures such as lab findings. | NA / Y / PY / PN / N / NI |
| | 1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention? | Controlling for post-intervention variables that are affected by intervention is not appropriate. Controlling for mediating variables estimates the direct effect of intervention and may introduce bias. Controlling for common effects of intervention and outcome introduces bias. | NA / Y / PY / PN / N / NI |
| | **Questions relating to baseline and time-varying confounding** | | |
| | 1.7. Did the authors use an appropriate analysis method that adjusted for all the important confounding domains and for time-varying confounding? | Adjustment for time-varying confounding is necessary to estimate the effect of starting and adhering to intervention, in both randomized trials and NRSI. Appropriate methods include those based on inverse probability weighting. Standard regression models that include time-updated confounders may be problematic if time-varying confounding is present. | NA / Y / PY / PN / N / NI |
| | 1.8. **If Y/PY to 1.7**: Were confounding domains that were adjusted for measured validly and reliably by the variables available in this study? | See 1.5 above. | NA / Y / PY / PN / N / NI |
| | **Risk of bias judgement** | See Table 1. | Low / Moderate / Serious / Critical / NI |
| | Optional: What is the predicted direction of bias due to confounding? | Can the true effect estimate be predicted to be greater or less than the estimated effect in the study because one or more of the important confounding domains was not controlled for? Answering this question will be based on expert knowledge and results in other studies and therefore can only be completed after all of the studies in the body of evidence have been reviewed. Consider the potential effect of each of the unmeasured domains and whether all important confounding domains not controlled for in the analysis would be likely to change the estimate in the same direction, or if one important confounding domain that was not controlled for in the analysis is likely to have a dominant impact. | Favours experimental / Favours comparator / Unpredictable |

| | | | |
|---|---|---|---|
| Bias in selection of participants into the study | 2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention?<br>**If N/PN to 2.1**: go to 2.4<br>2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?<br>2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome? | This domain is concerned only with selection into the study based on participant characteristics observed *after* the start of intervention. Selection based on characteristics observed *before* the start of intervention can be addressed by controlling for imbalances between experimental intervention and comparator groups in baseline characteristics that are prognostic for the outcome (baseline confounding).<br>Selection bias occurs when selection is related to an effect of either intervention or a cause of intervention **and** an effect of either the outcome or a cause of the outcome. Therefore, the result is at risk of selection bias if selection into the study is related to both the intervention and the outcome. | Y / PY / PN / N / NI<br>NA / Y / PY / PN / N / NI<br>NA / Y / PY / PN / N / NI |
| | 2.4. Do start of follow-up and start of intervention coincide for most participants? | If participants are not followed from the start of the intervention then a period of follow up has been excluded, and individuals who experienced the outcome soon after intervention will be missing from analyses. This problem may occur when prevalent, rather than new (incident), users of the intervention are included in analyses. | Y / PY / PN / N / NI |
| | 2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases? | It is in principle possible to correct for selection biases, for example by using inverse probability weights to create a pseudo-population in which the selection bias has been removed, or by modelling the distributions of the missing participants or follow up times and outcome events and including them using missing data methodology. However such methods are rarely used and the answer to this question will usually be "No". | NA / Y / PY / PN / N / NI |
| | **Risk of bias judgement** | See Table 1. | Low / Moderate / Serious / Critical / NI |
| | Optional: What is the predicted direction of bias due to selection of participants into the study? | If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions. | Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

| | | | |
|---|---|---|---|
| Bias in classification of interventions | 3.1 Were intervention groups clearly defined? | A pre-requisite for an appropriate comparison of interventions is that the interventions are well defined. Ambiguity in the definition may lead to bias in the classification of participants. For individual-level interventions, criteria for considering individuals to have received each intervention should be clear and explicit, covering issues such as type, setting, dose, frequency, intensity and/or timing of intervention. For population-level interventions (e.g. measures to control air pollution), the question relates to whether the population is clearly defined, and the answer is likely to be 'Yes'. | Y / PY / PN / N / NI |
| | 3.2 Was the information used to define intervention groups recorded at the start of the intervention? | In general, if information about interventions received is available from sources that could not have been affected by subsequent outcomes, then differential misclassification of intervention status is unlikely. Collection of the information at the time of the intervention makes it easier to avoid such misclassification. For population-level interventions (e.g. measures to control air pollution), the answer to this question is likely to be 'Yes'. | Y / PY / PN / N / NI |
| | 3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome? | Collection of the information at the time of the intervention may not be sufficient to avoid bias. The way in which the data are collected for the purposes of the NRSI should also avoid misclassification. | Y / PY / PN / N / NI |
| | **Risk of bias judgement** | See Table 1. | Low / Moderate / Serious / Critical / NI |
| | Optional: What is the predicted direction of bias due to measurement of outcomes or interventions? | If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions. | Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

| | | | |
|---|---|---|---|
| Bias due to deviations from intended interventions | **If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2** | | |
| | 4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice? | Deviations that happen in usual practice following the intervention (for example, cessation of a drug intervention because of acute toxicity) are part of the intended intervention and therefore do not lead to bias in the effect of assignment to intervention.<br><br>Deviations may arise due to expectations of a difference between intervention and comparator (for example because participants feel unlucky to have been assigned to the comparator group and therefore seek the active intervention, or components of it, or other interventions). Such deviations are not part of usual practice, so may lead to biased effect estimates. However these are not expected in observational studies of individuals in routine care. | Y / PY / PN / N / NI |
| | 4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome? | Deviations from intended interventions that do not reflect usual practice will be important if they affect the outcome, but not otherwise. Furthermore, bias will arise only if there is imbalance in the deviations across the two groups. | NA / Y / PY / PN / N / NI |
| | **If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6** | | |
| | 4.3. Were important co-interventions balanced across intervention groups? | Risk of bias will be higher if unplanned co-interventions were implemented in a way that would bias the estimated effect of intervention. Co-interventions will be important if they affect the outcome, but not otherwise. Bias will arise only if there is imbalance in such co-interventions between the intervention groups. Consider the co-interventions, including any pre-specified co-interventions, that are likely to affect the outcome and to have been administered in this study. Consider whether these co-interventions are balanced between intervention groups. | Y / PY / PN / N / NI |
| | 4.4. Was the intervention implemented successfully for most participants? | Risk of bias will be higher if the intervention was not implemented as intended by, for example, the health care professionals delivering care during the trial. Consider whether implementation of the intervention was successful for most participants. | Y / PY / PN / N / NI |
| | 4.5. Did study participants adhere to the assigned intervention regimen? | Risk of bias will be higher if participants did not adhere to the intervention as intended. Lack of adherence includes imperfect compliance, cessation of intervention, crossovers to the comparator intervention and switches to another active intervention. Consider available information on the proportion of study participants who continued with their assigned intervention throughout follow up, and answer 'No' or 'Probably No' if this proportion is high enough to raise concerns. Answer 'Yes' for studies of interventions that are administered once, so that imperfect adherence is not possible.<br>We distinguish between analyses where follow-up time after interventions switches (including cessation of intervention) is assigned to (1) the new intervention or (2) the original intervention. (1) is addressed under time-varying confounding, and should not be considered further here. | Y / PY / PN / N / NI |
| | 4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention? | It is possible to conduct an analysis that corrects for some types of deviation from the intended intervention. Examples of appropriate analysis strategies include inverse probability weighting or instrumental variable estimation. It is possible that a paper reports such an analysis without reporting information on the deviations from intended intervention, but it would be hard to judge such an analysis to be appropriate in the absence of such information. Specialist advice may be needed to assess studies that used these approaches.<br><br>If everyone in one group received a co-intervention, adjustments cannot be made to overcome this. | NA / Y / PY / PN / N / NI |
| | **Risk of bias judgement** | See Table 2 | |
| | Optional: What is the predicted direction of bias due to deviations from the intended interventions? | If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions. | |

| | | | |
|---|---|---|---|
| Bias due to missing data | 5.1 Were outcome data available for all, or nearly all, participants? | "Nearly all" should be interpreted as "enough to be confident of the findings", and a suitable proportion depends on the context. In some situations, availability of data from 95% (or possibly 90%) of the participants may be sufficient, providing that events of interest are reasonably common in both intervention groups. One aspect of this is that review authors would ideally try and locate an analysis plan for the study. | Y / PY / PN / N / NI |
| | 5.2 Were participants excluded due to missing data on intervention status? | Missing intervention status may be a problem. This requires that the *intended* study sample is clear, which it may not be in practice. | Y / PY / PN / N / NI |
| | 5.3 Were participants excluded due to missing data on other variables needed for the analysis? | This question relates particularly to participants excluded from the analysis because of missing information on confounders that were controlled for in the analysis. | Y / PY / PN / N / NI |
| | 5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions? | This aims to elicit whether either (i) differential proportion of missing observations or (ii) differences in reasons for missing observations could substantially impact on our ability to answer the question being addressed. "Similar" includes some minor degree of discrepancy across intervention groups as expected by chance. | NA / Y / PY / PN / N / NI |
| | 5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data? | Evidence for robustness may come from how missing data were handled in the analysis and whether sensitivity analyses were performed by the investigators, or occasionally from additional analyses performed by the systematic reviewers. It is important to assess whether assumptions employed in analyses are clear and plausible. Both content knowledge and statistical expertise will often be required for this. For instance, use of a statistical method such as multiple imputation does not guarantee an appropriate answer. Review authors should seek naïve (complete-case) analyses for comparison, and clear differences between complete-case and multiple imputation-based findings should lead to careful assessment of the validity of the methods used. | NA / Y / PY / PN / N / NI |
| | **Risk of bias judgement** | See Table 2 | Low / Moderate / Serious / Critical / NI |
| | Optional: What is the predicted direction of bias due to missing data? | If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions. | Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

| | | | |
|---|---|---|---|
| Bias in measurement of outcomes | 6.1 Could the outcome measure have been influenced by knowledge of the intervention received? | Some outcome measures involve negligible assessor judgment, e.g. all-cause mortality or non-repeatable automated laboratory assessments. Risk of bias due to measurement of these outcomes would be expected to be low. | Y / PY / PN / N / NI |
| | 6.2 Were outcome assessors aware of the intervention received by study participants? | If outcome assessors were blinded to intervention status, the answer to this question would be 'No'. In other situations, outcome assessors may be unaware of the interventions being received by participants despite there being no active blinding by the study investigators; the answer this question would then also be 'No'. In studies where participants report their outcomes themselves, for example in a questionnaire, the outcome assessor is the study participant. In an observational study, the answer to this question will usually be 'Yes' when the participants report their outcomes themselves. | Y / PY / PN / N / NI |
| | 6.3 Were the methods of outcome assessment comparable across intervention groups? | Comparable assessment methods (i.e. data collection) would involve the same outcome detection methods and thresholds, same time point, same definition, and same measurements. | Y / PY / PN / N / NI |
| | 6.4 Were any systematic errors in measurement of the outcome related to intervention received? | This question refers to differential misclassification of outcomes. Systematic errors in measuring the outcome, if present, could cause bias if they are related to intervention or to a confounder of the intervention-outcome relationship. This will usually be due either to outcome assessors being aware of the intervention received or to non-comparability of outcome assessment methods, but there are examples of differential misclassification arising despite these controls being in place. | Y / PY / PN / N / NI |
| | **Risk of bias judgement** | See Table 2 | Low / Moderate / Serious / Critical / NI |
| | Optional: What is the predicted direction of bias due to measurement of outcomes? | If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions. | Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

| | | | |
|---|---|---|---|
| Bias in selection of the reported result | Is the reported effect estimate likely to be selected, on the basis of the results, from...<br>7.1. ... multiple outcome *measurements* within the outcome domain? | For a specified outcome domain, it is possible to generate multiple effect estimates for different measurements. If multiple measurements were made, but only one or a subset is reported, there is a risk of selective reporting on the basis of results. | Y / PY / PN / N / NI |
| | 7.2 ... multiple *analyses* of the intervention-outcome relationship? | Because of the limitations of using data from non-randomized studies for analyses of effectiveness (need to control confounding, substantial missing data, etc), analysts may implement different analytic methods to address these limitations. Examples include unadjusted and adjusted models; use of final value vs change from baseline vs analysis of covariance; different transformations of variables; a continuously scaled outcome converted to categorical data with different cut-points; different sets of covariates used for adjustment; and different analytic strategies for dealing with missing data. Application of such methods generates multiple estimates of the effect of the intervention versus the comparator on the outcome. If the analyst does not pre-specify the methods to be applied, and multiple estimates are generated but only one or a subset is reported, there is a risk of selective reporting on the basis of results. | Y / PY / PN / N / NI |
| | 7.3 ... different *subgroups*? | Particularly with large cohorts often available from routine data sources, it is possible to generate multiple effect estimates for different subgroups or simply to omit varying proportions of the original cohort. If multiple estimates are generated but only one or a subset is reported, there is a risk of selective reporting on the basis of results. | Y / PY / PN / N / NI |
| | **Risk of bias judgement** | See Table 2 | Low / Moderate / Serious / Critical / NI |
| | Optional: What is the predicted direction of bias due to selection of the reported result? | If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions. | Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

| | | | |
|---|---|---|---|
| Overall bias | **Risk of bias judgement** | See Table 3. | Low / Moderate / Serious / Critical / NI |
| | Optional:<br>What is the overall predicted direction of bias for this outcome? | | Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |



Table 1. Reaching risk of bias judgements in ROBINS-I: pre-intervention and at-intervention domains

| Judgement | Bias due to confounding | Bias in selection of participants into the study | Bias in classification of interventions |
|---|---|---|---|
| Low risk of bias (the study is comparable to a well-performed randomized trial with regard to this domain) | No confounding expected. | (i) All participants who would have been eligible for the target trial were included in the study;<br>*and*<br>(ii) For each participant, start of follow up and start of intervention coincided. | (i) Intervention status is well defined;<br>*and*<br>(ii) Intervention definition is based solely on information collected at the time of intervention. |
| Moderate risk of bias (the study is sound for a non-randomized study with regard to this domain but cannot be considered comparable to a well-performed randomized trial): | (i) Confounding expected, all known important confounding domains appropriately measured and controlled for;<br>*and*<br>(ii) Reliability and validity of measurement of important domains were sufficient, such that we do not expect serious residual confounding. | (i) Selection into the study may have been related to intervention and outcome;<br>*and*<br>The authors used appropriate methods to adjust for the selection bias;<br>*or*<br>(ii) Start of follow up and start of intervention do not coincide for all participants;<br>*and*<br>(a) the proportion of participants for which this was the case was too low to induce important bias;<br>*or*<br>(b) the authors used appropriate methods to adjust for the selection bias;<br>*or*<br>(c) the review authors are confident that the rate (hazard) ratio for the effect of intervention remains constant over time. | (i) Intervention status is well defined;<br>*and*<br>(ii) Some aspects of the assignments of intervention status were determined retrospectively. |

| | | | |
|---|---|---|---|
| Serious risk of bias (the study has some important problems); | (i) At least one known important domain was not appropriately measured, or not controlled for;<br>*or*<br>(ii) Reliability or validity of measurement of an important domain was low enough that we expect serious residual confounding. | (i) Selection into the study was related (but not very strongly) to intervention and outcome;<br>*and*<br>This could not be adjusted for in analyses;<br>*or*<br>(ii) Start of follow up and start of intervention do not coincide;<br>*and*<br>A potentially important amount of follow-up time is missing from analyses;<br>*and*<br>The rate ratio is not constant over time. | (i) Intervention status is not well defined;<br>*or*<br>(ii) Major aspects of the assignments of intervention status were determined in a way that could have been affected by knowledge of the outcome. |
| Critical risk of bias (the study is too problematic to provide any useful evidence on the effects of intervention); | (i) Confounding inherently not controllable<br>*or*<br>(ii) The use of negative controls strongly suggests unmeasured confounding. | (i) Selection into the study was very strongly related to intervention and outcome;<br>*and*<br>This could not be adjusted for in analyses;<br>*or*<br>(ii) A substantial amount of follow-up time is likely to be missing from analyses;<br>*and*<br>The rate ratio is not constant over time. | (Unusual) An extremely high amount of misclassification of intervention status, e.g. because of unusually strong recall biases. |
| No information on which to base a judgement about risk of bias for this domain. | No information on whether confounding might be present. | No information is reported about selection of participants into the study or whether start of follow up and start of intervention coincide. | No definition of the intervention or no explanation of the source of information about intervention status is reported. |



Table 2. Reaching risk of bias judgements in ROBINS-I: post-intervention domains

| Judgement | Bias due to deviations from intended intervention | Bias due to missing data | Bias in measurement of outcomes | Bias in selection of the reported result |
|---|---|---|---|---|
| Low risk of bias (the study is comparable to a well-performed randomized trial with regard to this domain) | **Effect of assignment to intervention**:<br>(i) Any deviations from intended intervention reflected usual practice;<br>*or*<br>(ii) Any deviations from usual practice were unlikely to impact on the outcome.<br><br>**Effect of starting and adhering to intervention**:<br>The important **co-interventions** were balanced across intervention groups, and there were no deviations from the intended interventions (in terms of **implementation or adherence**) that were likely to impact on the outcome. | (i) Data were reasonably complete;<br>*or*<br>(ii) Proportions of and reasons for missing participants were similar across intervention groups;<br>*or*<br>(iii) The analysis addressed missing data and is likely to have removed any risk of bias. | (i) The methods of outcome assessment were comparable across intervention groups;<br>*and*<br>(ii) The outcome measure was unlikely to be influenced by knowledge of the intervention received by study participants (i.e. is objective) or the outcome assessors were unaware of the intervention received by study participants;<br>*and*<br>(iii) Any error in measuring the outcome is unrelated to intervention status. | There is clear evidence (usually through examination of a pre-registered protocol or statistical analysis plan) that all reported results correspond to all intended outcomes, analyses and sub-cohorts. |

| | | | | |
|---|---|---|---|---|
| Moderate risk of bias (the study is sound for a non-randomized study with regard to this domain but cannot be considered comparable to a well-performed randomized trial): | **Effect of assignment to intervention**:<br>There were deviations from usual practice, but their impact on the outcome is expected to be slight.<br><br>**Effect of starting and adhering to intervention**:<br>(i) There were deviations from intended intervention, but their impact on the outcome is expected to be slight.<br>*or*<br>(ii) The important co-interventions were not balanced across intervention groups, or there were deviations from the intended interventions (in terms of implementation and/or adherence) that were likely to impact on the outcome;<br>*and*<br>The analysis was appropriate to estimate the effect of starting and adhering to intervention, allowing for deviations (in terms of implementation, adherence and co-intervention) that were likely to impact on the outcome. | (i) Proportions of and reasons for missing participants differ slightly across intervention groups;<br>*and*<br>(ii) The analysis is unlikely to have removed the risk of bias arising from the missing data. | (i) The methods of outcome assessment were comparable across intervention groups;<br>*and*<br>(ii) The outcome measure is only minimally influenced by knowledge of the intervention received by study participants;<br>*and*<br>(iii) Any error in measuring the outcome is only minimally related to intervention status. | (i) The outcome measurements and analyses are consistent with an *a priori* plan; or are clearly defined and both internally and externally consistent;<br>*and*<br>(ii) There is no indication of selection of the reported analysis from among multiple analyses;<br>*and*<br>(iii) There is no indication of selection of the cohort or subgroups for analysis and reporting on the basis of the results. |

| | | | | |
|---|---|---|---|---|
| Serious risk of bias (the study has some important problems); | **Effect of assignment to intervention**:<br>There were deviations from usual practice that were unbalanced between the intervention groups and likely to have affected the outcome.<br><br>**Effect of starting and adhering to intervention**:<br>(i) The important co-interventions were not balanced across intervention groups, or there were deviations from the intended interventions (in terms of implementation and/or adherence) that were likely to impact on the outcome;<br>*and*<br>(ii) The analysis was not appropriate to estimate the effect of starting and adhering to intervention, allowing for deviations (in terms of implementation, adherence and co-intervention) that were likely to impact on the outcome. | (i) Proportions of missing participants differ substantially across interventions;<br>*or*<br>Reasons for missingness differ substantially across interventions;<br>*and*<br>(ii) The analysis is unlikely to have removed the risk of bias arising from the missing data;<br>*or*<br>Missing data were addressed inappropriately in the analysis;<br>*or*<br>The nature of the missing data means that the risk of bias cannot be removed through appropriate analysis. | (i) The methods of outcome assessment were not comparable across intervention groups;<br>*or*<br>(ii) The outcome measure was subjective (i.e. vulnerable to influence by knowledge of the intervention received by study participants);<br>*and*<br>The outcome was assessed by assessors aware of the intervention received by study participants;<br>*or*<br>(iii) Error in measuring the outcome was related to intervention status. | (i) Outcomes are defined in different ways in the methods and results sections, or in different publications of the study;<br>*or*<br>(ii) There is a high risk of selective reporting from among multiple analyses;<br>*or*<br>(iii) The cohort or subgroup is selected from a larger study for analysis and appears to be reported on the basis of the results. |

| | | | | |
|---|---|---|---|---|
| Critical risk of bias (the study is too problematic to provide any useful evidence on the effects of intervention); | **Effect of assignment to intervention:**<br>There were substantial deviations from usual practice that were unbalanced between the intervention groups and likely to have affected the outcome.<br><br>**Effect of starting and adhering to intervention:**<br>(i) There were substantial imbalances in important co-interventions across intervention groups, or there were substantial deviations from the intended interventions (in terms of implementation and/or adherence) that were likely to impact on the outcome;<br>*and*<br>(ii) The analysis was not appropriate to estimate the effect of starting and adhering to intervention, allowing for deviations (in terms of implementation, adherence and co-intervention) that were likely to impact on the outcome. | (i) (Unusual) There were critical differences between interventions in participants with missing data;<br>*and*<br>(ii) Missing data were not, or could not, be addressed through appropriate analysis. | The methods of outcome assessment were so different that they cannot reasonably be compared across intervention groups. | (i) There is evidence or strong suspicion of selective reporting of results;<br>*and*<br>(ii) The unreported results are likely to be substantially different from the reported results. |
| No information on which to base a judgement about risk of bias for this domain. | No information is reported on whether there is deviation from the intended intervention. | No information is reported about missing data or the potential for data to be missing. | No information is reported about the methods of outcome assessment. | There is too little information to make a judgement (for example, if only an abstract is available for the study). |



Table 3. Interpretation of domain-level and overall risk of bias judgements in ROBINS-I

| Judgement | Within each domain | Across domains | Criterion |
|---|---|---|---|
| Low risk of bias | The study is comparable to a well-performed randomized trial with regard to this domain | The study is comparable to a well-performed randomized trial | The study is judged to be at **low risk of bias for all domains**. |
| Moderate risk of bias | The study is sound for a non-randomized study with regard to this domain but cannot be considered comparable to a well-performed randomized trial | The study provides sound evidence for a non-randomized study but cannot be considered comparable to a well-performed randomized trial | The study is judged to be at **low or moderate risk of bias for all domains**. |
| Serious risk of bias | the study has some important problems in this domain | The study has some important problems | The study is judged to be at **serious risk of bias** in at least one domain, but not at critical risk of bias in any domain. |
| Critical risk of bias | the study is too problematic in this domain to provide any useful evidence on the effects of intervention | The study is too problematic to provide any useful evidence and should not be included in any synthesis | The study is judged to be at **critical risk of bias in at least one domain**. |
| No information | No information on which to base a judgement about risk of bias for this domain | No information on which to base a judgement about risk of bias | There is no clear indication that the study is at serious or critical risk of bias *and* there is a lack of information in one or more key domains of bias (*a judgement is required for this*). |