Clinical CBT treatment (Targeted / High-risk)

Toolkit technical report

Ciara Keenan, Iman Idrees, Karen Allen, Lauren Murphy, Jessica Hinks, Ofelia Torres, Jenny Towers, Jaimee S. Mallion, Sophie Hedges

June 2026

About National Children's Bureau

This report has been produced by the National Children's Bureau on behalf of the Youth Endowment Fund. The National Children's Bureau works collaboratively across the issues affecting children to influence policy and get services working together to deliver a better childhood. They were commissioned by the Youth Endowment Fund (YEF) as their Toolkit Partner 2023-2026. The Toolkit partnership is managed on behalf of NCB by Dr Ciara Keenan. Any queries relating to the methods should be directed to ckeenan@ncb.org.uk.

About Youth Endowment Fund

The Youth Endowment Fund's mission is to prevent children and young people becoming involved in violence. They do this by finding out what works and building a movement to put this knowledge into practice. The fund was established in March 2019 by children's charity Impetus, with a £200m endowment and ten-year mandate from the Home Office. For more information, please visit www.youthendowmentfund.org.uk.

Acknowledgements

Our deepest appreciation goes to all those who shared their insights and expertise to enhance the quality of this research. Special acknowledgement is due to Dr. Laura Knight, Head of Toolkit, Nigel Tungwarara and Dr. Chad Hemady, Senior Research Managers at the Youth Endowment Fund, for their diligent work and unwavering support throughout this project. Our sincere appreciation also goes to Professor James Thomas and the entire team at the EPPI-Centre, whose dedication and countless hours have been instrumental in the development of this work.

YOUTH ENDOWMENT FUND

NATIONAL CHILDREN'S BUREAU

Abstract/Plain Language Summary

The primary objective of this report is to evaluate the evidence regarding the effectiveness of **clinical cognitive behavioural therapy (CBT) for high-risk children and young people** in reducing and/or preventing violence and offending.

Cognitive behavioural therapy involves working with the inter-relationship of thoughts, feelings and behaviour through the recognition of cognitive distortions, the improvement of executive functions such as self-regulation, and the development of coping, communication, and social problem-solving skills. These models aim to support young people in identifying appropriate behavioural goals, developing awareness and understanding of emotional experiences, and regulating behavioural responses as required.

Targeted interventions indicative of this strand include clinical cognitive behavioural therapy for high-risk children and young people, particularly those presenting with offending histories, externalising difficulties, neurodevelopmental conditions, and other psychiatric diagnoses.

Key findings

- Clinical CBT interventions appear to reduce violence by 69%, based on a meta-analysis of 28 measured outcomes across nine studies. There is an overall evidence security rating of Level 3 (Moderate) for this finding.
- Clinical CBT interventions may reduce crime and offending by 88%, based on a meta-analysis of 35 measured outcomes across six studies. There is an overall evidence security rating of Level 2 for this finding.
- Across all related outcomes, Clinical CBT interventions are estimated to have an overall beneficial effect (g=-0.48) equating to a high YEF impact rating, based on a meta-analysis of 128 measured outcomes across 14 studies. There is an overall evidence security rating of Level 3 (Moderate) for this finding.
- Subgroup analyses provided limited evidence of moderation effects. Most moderators showed no statistically significant subgroup effects, including programme intensity, inclusion criteria, family/parent involvement,

publication decade, age range, and ethnicity, suggesting that the apparent effectiveness of Clinical CBT interventions was generally similar across these study and participant characteristics.

- The clearest signals of variation were by intervention setting ($p < .001$), delivery format ($p = .011$), and possibly gender composition ($p = .003$), but these findings should be interpreted cautiously as several subgroup estimates were based on very small numbers of studies, with some statistically significant findings coming from only one or two studies.
- Adaptations were essential for successful implementation, with programmes commonly modified to meet the developmental needs, co-occurring presentations, and contextual realities of clinical services.
- Engagement was strengthened by family involvement and interactive or flexible intervention features, which helped maintain participation and support skill use across settings.
- Structured protocols supported consistent delivery, but organisational factors (e.g., staffing, communication, service constraints) influenced how reliably interventions were implemented.
- None of the reviewed studies provided detailed economic evaluations or cost data relating to the implementation of Clinical CBT programmes, limiting the ability to assess affordability, cost-effectiveness, or resource requirements for wider service implementation.

Conclusion

Our meta-analysis finds that Clinical CBT programmes reduce violence and offending among high-risk children and young people and appear to have an overall beneficial effect across related outcomes overall. Subgroup analyses provided limited evidence of consistent moderation, suggesting that effects were broadly similar across participant and study characteristics.

Clinical CBT should be implemented as a structured, targeted intervention for high-risk children and young people, with delivery adapted to developmental and clinical needs. Clear protocols, staff training, organisational support, family involvement, and engaging delivery approaches appear important for successful implementation. However, the lack of economic evidence means conclusions about affordability and cost-effectiveness cannot yet be drawn.

YOUTH ENDOWMENT FUND

NATIONAL CHILDREN'S BUREAU

Table of Contents

Abstract/Plain Language Summary 3

List of tables 6

List of figures 8

Preface on Terminology 9

Objective and Approach 10

Description of the Intervention 12

How Effective is the Intervention? 20

How Secure is the Evidence? 57

Who does it work for? 59

What factors affect implementation? 62

How much does it cost? 73

Conclusion and Takeaway Messages 74

References 83

Appendix 1. Methods of the systematic review 83

Appendix 2. Location Details 101

Appendix 3. Characteristics of included studies for effectiveness 102

Appendix 4. Characteristics of included studies for implementation 115

Appendix 5. Availability of evidence according to each of Proctor et al.'s (2011) implementation outcomes 121

Appendix 6. Moderator categories 123

Appendix 7. Measured outcomes across included studies for effectiveness 134

List of tables

Table 1: Summary of findings on violence and crime outcomes 25

Table 2: RVE Output for meta-analysis on violence outcomes 26

Table 3: RVE Output for meta-analysis on crime and offending outcomes 28

Table 4: RVE Output for meta-analysis of all related outcomes 30

Table 5: Subgroup analysis on programme intensity 31

Table 6: Subgroup analysis on inclusion criteria 32

Table 7: Subgroup analysis on intervention setting 33

Table 8: Subgroup analysis on mode of delivery 34

Table 9: Subgroup analysis on family involvement 35

Table 10: Subgroup analysis on outcome domain 36

Table 11: Subgroup analysis on country 39

Table 12: Subgroup analysis on publication year 40

Table 13: Subgroup analysis on developmental age band 41

Table 14: Subgroup analysis on ethnicity 42

Table 15: Subgroup analysis on gender 43

Table 16: Meta-regression (with robust inference) on quality moderators 44

Table 17: Meta-regression (with robust inference) on setting-level moderators 46

Table 18: Meta-regression (with robust inference) on intervention-level moderators 47

Table 19: Meta-regression (with robust inference) on population-level moderators 49

Table 20: Meta-regression (with robust inference) on outcome-level moderators 50

Table 21: Eligibility criteria for Clinical CBT studies 87

Table 22: Full-text screening results 91

Table 23: Reasons for exclusion after full text screening 92

Table 24: List of linked studies 92

Table 25: Quality appraisal ratings for studies included in the Clinical CBT strand 95

YOUTH ENDOWMENT FUND

NATIONAL CHILDREN'S BUREAU

List of figures

Figure 1: Absolute risk reduction in violence from Clinical CBT interventions compared to business as usual 27

Figure 2: Absolute risk reduction in crime and offending from Clinical CBT interventions compared to business as usual 29

Figure 3: Funnel plot of individual studies 53

Figure 4: Influence diagnostics for individual studies 54

Figure 5: Leave-one-out sensitivity analysis 56

Figure 6: PRISMA Flow Diagram 94

YOUTH ENDOWMENT FUND

NATIONAL CHILDREN'S BUREAU

Preface on Terminology

This review draws on evidence spanning over half a century, during which language around personal characteristics has evolved significantly. At times, we may have to reproduce original terminology used in studies which we recognize today as being outdated and unacceptable offensive terms. This only occurs when the terminology is used in direct quotations or refers to an outcome that the author measured that remains relevant to our analysis. The wider narrative will adhere to current inclusive-language standards guided by the National Children's Bureau, Youth Endowment Fund, and Race Equality Foundation. These guiding principles include using capitalisation to acknowledge shared identities (e.g., Black, Asian), whilst not capitalising white due to its association with white supremacy. The review also avoids deficit framing and respects individuals' self-identification. Person-first language will generally be used when referring to children and young people, except for Deaf and autistic communities, who widely prefer identity-first language. The team acknowledges limitations in terminology and strives for respectful and precise representation throughout. The full preface on terminology can be accessed here.

Objective and Approach

The objective of this report is to review the evidence on the effects of **Clinical CBT** in reducing and/or preventing violence and offending involving children and young people. Cognitive Behavioural Therapy (CBT) is widely utilised to address psychological difficulties and prevent violence (Smeets et al. 2015) by targeting the negative or impulsive thought patterns that may precipitate aggressive behaviour (Youth Endowment Fund, 2021.). Such interventions help children and young people to recognise and address emotions and unhelpful thoughts, develop a more accurate understanding of others' motivations , strengthen problem-solving skills in response to challenging situations and address emotion dysregulation(Hoogsteder et al., 2018). A fundamental component that distinguishes CBT from general social skills training is cognitive restructuring, which involves structured exercises designed to identify, evaluate and modify thinking patterns and deep-seated cognitive distortions, such as hostile attribution biases (the tendency to interpret ambiguous social cues as intentionally hostile) or the tendency to displace blame (Lipsey et al., 2007).

This systematic review and meta-analysis focuses specifically on the specialist-led, clinical strand of CBT interventions. These clinical programmes are designed for children and young people aged 0 to 17 who are at a high risk of reoffending or who possess formal clinical diagnoses linked to aggression, such as conduct disorder, oppositional defiant disorder, or attention-deficit/hyperactivity disorder(ADHD). In contrast to preventative, curriculum-based programmes delivered by lay practitioners in community settings (which is the defining feature of our CBT Skills strand), Clinical CBT interventions are characterised by their high intensity, often exceeding six months or 30 contact hours. Furthermore, they are delivered by licensed mental health professionals or specialised forensic practitioners within clinical environments, such as hospitals and outpatient clinics, or within justice settings, including prisons, secure children's homes, and Youth Offending Teams.

A core theoretical underpinning of these targeted interventions is the risk-need-responsivity (RNR) model, which posits that treatments must correspond to a

young person's risk level, target modifiable criminogenic needs, and be tailored to the individual's specific capabilities and social environment (Koehler et al., 2013). To maintain a strict focus on violence prevention among high-risk and offending populations, this review explicitly excludes low-intensity universal programmes, generic psychodynamic therapies without CBT elements, and trauma-focused therapies that are primarily designed for victimisation recovery. Ultimately, this review evaluates studies with credible counterfactuals to synthesise the evidence on the effectiveness of high-intensity, specialist-led CBT in reducing crime and violence among high-risk children and young people.

This technical report draws on a comprehensive systematic review methodology and includes:

- **Fourteen effectiveness studies** providing outcomes data for meta-analysis, with one being based in the UK: 28 violence outcomes across nine studies; 35 crime and offending outcomes across six studies; and 128 outcomes across the whole of the YEF outcomes framework.
- **Ten studies** providing implementation insights. Findings are reported according to Proctor's (2011) framework.

The remainder of this report is structured as follows: First, the **Description of the Intervention** outlines the key components of Clinical CBT and its intended implementation. Second, **How Effective is the Intervention?** presents findings from our meta-analysis on violence and crime reduction and broader social outcomes. Third, **Who Does it Work For?** examines evidence on the populations that benefit most from Clinical CBT. Fourth, **What Factors Affect Implementation?** explores key factors which influence the implementation of Clinical CBT interventions. Fifth, **How Much Does It Cost?** reviews available cost data. Finally, the **Conclusion and Takeaway Messages** summarise key findings and recommendations, followed by **Appendices** detailing the systematic review methodology and characteristics of included research.

Description of the Intervention

In the following section details are provided on the interventions which inform this report, noting their key components, any equipment, materials, supplies or training required, the duration and intensity of interventions, who delivered the interventions, and where and how the interventions were delivered.

Features of the approach

Cognitive Behavioural Therapy (CBT) is a type of talking therapy that has been used to address a range of psychological difficulties. When it is applied to violence prevention, it is based on the idea that negative or impulsive thoughts and behaviours might make someone more likely to lash out or act aggressively.

CBT aims to help children and young people become more aware of these negative thoughts and learn to change or manage them. The therapist or facilitator might work with a child to explore how their assumptions relate to reality, better understand their own and other people's behaviour and motivations, and use problem-solving skills to cope with difficult situations.

Clinical CBT interventions often focus on cognitive restructuring. They target specific criminogenic needs or behaviours associated with violence and offending in high risk populations with diagnosed disorders. These interventions are likely to be high-intensity, collaborative and are "responsive," meaning they are tailored to the individual's needs and specific risk profile. They are delivered by clinicians or specialised forensic practitioners in settings that manage high risk, including hospitals, CAMHS, prisons, and Youth Offending Teams (YOTs).

Key components of Clinical CBT

Programmes were typically structured with manuals or lesson plans but could also be tailored to individual needs. For example, CBT Relapse Prevention (CBT-RP; Goldston, 2021) and Responsive Aggression Regulation Therapy (Re ART; (Hoogsteder et al., 2018) included both mandatory and optional modules. Hoogsteder et al. (2018) described the intervention as rooted in Risk, Need and Responsivity (RNR) principles, enabling practitioners to identify high risk young

people, target individual needs, and tailor content to criminogenic risks, motivations and learning styles.

While interventions were typically designed to address aggression, several targeted specific behaviours, such as fire setting (Kolko et al., 2006), substance use (Barrett et al., 2001), or co-occurring substance use and suicidality (Esposito-Smythers et al., 2011; Goldston et al., 2021). Others addressed specific needs, including autism (Clifford et al., 2022; Townsend et al., 2024), Oppositional Defiant Disorder (Helander et al., 2023), conduct problems (Koegl et al., 2008) and disruptive mood dysregulation disorder (Soleimani-Rad et al., 2025). Treatment was adapted accordingly; for example, Townsend et al. (2024) incorporated behavioural management strategies and social skills training that used rewards and young people's special interests.

Although Clinical CBT offered more intensive, targeted approaches than CBT Skills Programmes, many interventions also aimed to build broader skills. These included:

- **Social skills** and communication through modelling, role-play and structured modules (Barrett et al., 2001; Curran et al., 1977; Erickson, 2013; Helander et al., 2023; Hoogsteder et al., 2018; Koegl et al., 2008; Raine et al., 2016; Rohde et al., 2004; Soleimani-Rad et al., 2025; Townsend et al., 2024).
- **Problem-solving**, such as generating solutions for anger-provoking situations, conflict resolution and evaluating alternative responses (Barrett et al., 2001; Clifford et al., 2022; Curran et al., 1977; Ducharme et al., 2021; Erickson, 2013; Goldston et al., 2021; Helander et al., 2023; Kolko et al., 2006; Raine et al., 2016; Rohde et al., 2004; Soleimani-Rad et al., 2025).
- **Emotional awareness and regulation**, including managing negative mood (Barrett et al., 2001; Curran et al., 1977; Ducharme et al., 2021; Helander et al., 2023; Kolko et al., 2006; Raine et al., 2016; Reardon & Tosi, 1977; Soleimani-Rad et al., 2025).
    - **Anger management and behavioural regulation**, involving the identification of emotions and the use of relaxation or meditation (Clifford et al., 2022; Ducharme et al., 2021; Erickson, 2013; Helander et

al., 2023; Raine et al., 2016; Reardon & Tosi, 1977; Rohde et al., 2004; Soleimani-Rad et al., 2025).

- **Cognitive restructuring**, supporting young people to recognise, evaluate and replace unhelpful thoughts or hostile attributions (Erickson, 2013; Esposito-Smythers et al., 2011; Helander et al., 2023; Hoogsteder et al., 2018; Reardon & Tosi, 1977; Rohde et al., 2004; Soleimani-Rad et al., 2025; Townsend et al., 2024).

Some studies described mechanisms for practising these skills. For example, in Ducharme et al. (2021), young people applied strategies in a videogame that monitored heart rate and prompted relaxation when arousal increased, followed by facilitator-led reflection.

Interventions also used behaviour management techniques such as point systems or behaviour contracts (Esposito-Smythers et al., 2011; Rohde et al., 2004) and parent-supported plans for consequences of behaviour (Kolko et al., 2006).

Motivational interviewing featured across several studies, helping young people generate self-motivational statements, manage resistance, and plan for change (Curran et al., 1977; Esposito-Smythers et al., 2011; Goldston et al., 2021). Barrett et al. (2001) similarly used non-confrontational strategies to enhance motivation, prioritise goals and develop treatment plans.

Many interventions were multi component. For example, Stop Now And Plan (SNAP; Koegl et al., 2008) offered additional supports such as family counselling, befriending and tutoring, while Re-ART (Hoogsteder et al., 2018) provided access to a mentor. Other supplementary components included family therapy (Esposito-Smythers et al., 2011) or school related skills (Barrett et al., 2001).

Although interventions primarily targeted children and young people, several incorporated parental/caregiver involvement. This included parent sessions focused on understanding and managing emotions and aggression (Clifford et al., 2022; Esposito-Smythers et al., 2011; Helander et al., 2023; Kolko et al., 2006; Raine et al., 2016; van de Wiel et al., 2007) and parents attending or being informed about child sessions to support home practice (Clifford et al., 2022; Ducharme et al., 2021; Kolko et al., 2006; Raine et al., 2016; van de Wiel et al., 2007).

Equipment, materials or supplies

While not always stated, all interventions appeared to include a degree of standardisation or structured delivery. In practice, this suggests practitioners were likely guided by a manual, intervention guide or session plan to support consistency in how the intervention was delivered..

A variety of learning materials supported skill development, including facilitated discussions and role plays (Erickson, 2013; Hoogsteder et al., 2018; Koegl et al., 2008), problem solving situations and group based moral reasoning tasks (Erickson, 2013), worksheets (Esposito-Smythers et al., 2011), and quizzes and homework (Rohde et al., 2004). Erickson (2013) also described a "hassle log" for recording skills practice, available in both written and pictorial formats to enhance accessibility.

Additional engagement materials in Clifford et al. (2022) included stories and pictures to help with externalisation and perspective taking. These featured a character with anger regulation difficulties, alongside creative resources such as meditation tools, art supplies, Lego, modelling clay, books and comics to support relaxation. Children received a completion certificate summarising their triggers and coping strategies.

Several interventions also incorporated a range of materials for children, young people and parents. For instance, Raine et al. (2016) provided a manual and notebook for both young people and parents, while Clifford et al. (2022) employed tools such as an anger thermometer to help families identify emotions and triggers. Reardon & Tosi (1977) described the use of a self-directed behaviour change instrument.

Digital tools were also used in some programmes. In Townsend et al. (2024), an app provided training, guidance and an algorithm that adapted content to individual needs. The RAGE Control videogame (Ducharme et al., 2021) paired a non-violent, space themed game with a pulse oximeter; increases in heart rate led the spaceship to fire "blanks", prompting the use of relaxation techniques to reduce arousal.

Some interventions also required screening or fidelity materials. Curran et al. (1977) and Erickson (2013) used assessment and screening tools to confirm eligibility, and four studies monitored implementation fidelity using audio or video recordings (Curran et al., 1977; Erickson, 2013; Esposito-Smythers et al., 2011; van de Wiel et al., 2007). Finally, Helander et al. (2023) supported families by phone to complete questionnaires.

Who delivers Clinical CBT

In most cases, interventions were delivered by therapists, who were typically master's or doctoral students or were educated to at least master's level (Barrett et al., 2001; Goldston et al., 2021; Reardon & Tosi, 1977; Townsend et al., 2024). Although, the level and nature of CBT-specific training therapists had received varied and was not always clearly reported. In Van De Wiel et al. (2007), therapists held master's degrees in psychology but had limited clinical experience, while in Raine et al. (2016) the intervention was delivered by a licensed therapist with experience in CBT.

Six studies reported that interventions were delivered by psychologists or trainee psychologists (Clifford et al., 2022; Esposito-Smythers et al., 2011; Koegl et al., 2008; Kolko et al., 2006; Rohde et al., 2004; Soleimani-Rad et al., 2025). In Clifford et al. (2022), psychologists also visited children's schools to introduce time out procedures and support teachers in helping children identify when to take a break and implement calming techniques.

Other professionals were involved in three studies. Re-ART (Hoogsteder et al., 2018) was delivered by therapists and mentors, while Erickson (2013) used curriculum trainers. Ducharme et al. (2021) reported that the intervention was delivered by clinical social workers.

How was the intervention delivered

Most interventions were delivered face to face to individual children and young people (Barrett et al., 2001; Clifford et al., 2022; Ducharme et al., 2021; Esposito-Smythers et al., 2011; Goldston et al., 2021; Raine et al., 2016; Reardon & Tosi, 1977; Soleimani-Rad et al., 2025; Townsend et al., 2024). In Kolko et al. (2006), the

intervention was delivered either individually or jointly with children, young people and their parents.

A smaller number of interventions were delivered face to face in groups (Curran et al., 1977; Erickson, 2013; Helander et al., 2023; Koegl et al., 2008; Rohde et al., 2004; van de Wiel et al., 2007). In Koegl et al. (2008), however, family counselling, befriending and tutoring were delivered individually. Hoogsteder et al. (2018) offered a combination of individual and group sessions, all delivered in person.

Where separate parent sessions were included, these were usually delivered in groups (Clifford et al., 2022; Esposito-Smythers et al., 2011; Helander et al., 2023; Koegl et al., 2008; van de Wiel et al., 2007), though in two studies (Kolko et al., 2006; Raine et al., 2016) they were delivered to individual parents.

Where Clinical CBT is delivered

Most of the interventions were delivered in outpatient settings, typically within child mental health centres or psychiatric hospitals (Barrett et al., 2001; Clifford et al., 2022; Ducharme et al., 2021; Goldston et al., 2021; Koegl et al., 2008). Esposito-Smythers et al. (2011) described sessions as being delivered in the community following hospital discharge.

Other interventions were held in juvenile justice/rehabilitation facilities (Erickson, 2013; Hoogsteder et al., 2018; Reardon & Tosi, 1977; Rohde et al., 2004), while the Ross Programme (Curran et al., 1977) worked with young people attending an Education and Care Centre in Scotland who were on a residential or day placement. The CBT intervention in Raine et al. (2016) was delivered within the family home.

Training for the providers of Clinical CBT

Training, sometimes described as “extensive” or “considerable”, was reported in many studies (Barrett et al., 2001; Clifford et al., 2022; Ducharme et al., 2021; Erickson, 2013; Esposito-Smythers et al., 2011; Kolko et al., 2006; Raine et al., 2016; Reardon & Tosi, 1977; Townsend et al., 2024; van de Wiel et al., 2007). The most extensive training was described in Van De Wiel et al. (2007), where therapists completed six months of preparation.

Training typically involved familiarisation with manuals and intervention materials. In Van De Wiel et al. (2007), this included background reading, reviewing videotapes, and conducting pilot sessions. In Esposito-Smythers et al. (2011), training also incorporated role-plays.

Two studies reported minimal educational or experiential requirements but did not describe specific training (Goldston et al., 2021; Hoogsteder et al., 2018). In Soleimani-Rad et al. (2025), therapists were described as "expert," but no details of training were provided.

Many studies also described how supervision, typically weekly, was provided to ensure quality, consistency and fidelity (Barrett et al., 2001; Clifford et al., 2022; Curran et al., 1977; Ducharme et al., 2021; Esposito-Smythers et al., 2011; Rohde et al., 2004; Townsend et al., 2024; van de Wiel et al., 2007). In Hoogsteder et al. (2018), trainers received support through 'intervision' sessions, described as a form of peer supervision.

Four studies also referenced the use of audio or video recordings to monitor fidelity, with tapes reviewed and feedback provided to trainers (Curran et al., 1977; Erickson, 2013; Esposito-Smythers et al., 2011; van de Wiel et al., 2007).

Duration and intensity of Clinical CBT

Where intensity was reported, sessions were mostly delivered on a weekly basis (Clifford et al., 2022; Ducharme et al., 2021; Esposito-Smythers et al., 2011; Koegl et al., 2008; Kolko et al., 2006; Raine et al., 2016; Reardon & Tosi, 1977; Rohde et al., 2004; Soleimani-Rad et al., 2025; Townsend et al., 2024).

Some interventions followed different schedules. In Barrett et al. (2001) sessions were delivered once a week or once a fortnight, while Re-ART (Hoogsteder et al., 2018) was delivered once or twice a week. Sessions in Aggression Replacement Training (Erickson, 2013) occurred three times a week. CBT-RP (Goldston et al., 2021) was delivered at least weekly for the first 12 weeks, with high risk young people receiving twice weekly sessions. After the initial 12-week period, sessions occurred weekly or every other week, depending on need.

In respect of duration, eight interventions lasted between one and three months (Clifford et al., 2022; Ducharme et al., 2021; Erickson, 2013; Koegl et al., 2008; Kolko et al., 2006; Raine et al., 2016; Reardon & Tosi, 1977; Rohde et al., 2004) while four interventions lasted between three and six months (Barrett et al., 2001; Goldston et al., 2021; Soleimani-Rad et al., 2025; Townsend et al., 2024). Goldston et al. (2021) also offered the option of an additional six sessions if needed.

Longer interventions were also reported: the intervention in Van De Wiel et al. (2007) lasted between six months and one year; Re-ART (Hoogsteder et al., 2018) lasted six months to two years depending on individual need; and Integrated-CBT (Esposito-Smythers et al., 2011) lasted between one and two years.

How Effective is the Intervention?

This section examines the effectiveness of Clinical CBT in reducing violence, crime and offending, and other related outcomes through a systematic review and meta-analysis, to provide a robust and objective summary of existing evidence, incorporating advanced statistical techniques, including robust variance estimators (Pustejovsky & Tipton, 2022), for improved accuracy.

Quantitative data from **14 effectiveness studies** provided information across a variety of **128 outcomes** related to the impact of Clinical CBT for children and young people.

These studies, included in the meta-analysis, assessed the effectiveness of Clinical CBT across a total of **1,042 children and young people.** These 14 effectiveness studies employed a range of study designs, including:

- Quasi-Experimental Designs (QED): (n = 1, 7%)
- Randomised Controlled Trials (RCT): (n = 13, 93%)

The 14 effectiveness studies varied considerably in their methodological design and reporting characteristics (see Appendix 3 for individual study details).

Most studies used a prospective design and assigned participants at the start of the study (n = 12, 86%), while two studies used pre-existing (natural) differences to create comparison groups (n = 2, 14%).

Most studies allocated at the level of individual participants (n=13, 93%), while Hoogsteder et al., (2018) used a quasi-experimental design rather than true randomised assignment, citing organisational barriers to randomising their participants. Participants in this study were allocated to groups based on practical availability.

The effectiveness studies have spanned the past three decades, with the earliest conducted by Barrett et al. (2001) and the most recent by Soleimani-Rad et al. (2025).

The studies were conducted in six different countries including:

- USA: (n = 7, 50%)

- Netherlands: (n = 3, 21%)
- Canada: (n = 1, 7%)
- Scotland: (n = 1, 7%)
- Sweden: (n = 1, 7%)
- Iran: (n = 1, 7%)

Fuller descriptions of the included studies are available in Appendix 3. Studies were assessed for methodological quality using the YEF-EQA critical appraisal tool and were rated as follows:

- High: (n = 2, 14%)
- Moderate: (n = 10, 71%)
- Low: (n = 2, 14%)

Four studies did not report their funding source (29%), while Soleimani-Rad et al. (2025) and Hoogsteder et al. (2018) explicitly state that the authors received no specific financial support or funding for their research (14%). Where reported across the remaining eight studies, funding mainly came from Government and National Institutes (n=6; 43%) while two were funded by Foundations, Councils, and Academic Institutions (n=2; 14%). Two of the included studies explicitly declared a conflict of interest; Ducharme et al. (2021) state that authors have equity in the company that owns the game technology tested in the intervention; while in Townsend et al. (2024) several authors disclosed financial interests, gained through sales of publications and materials about the treatment of anxiety disorders in youth, though largely outside the present work.

The intensity of interventions varied from 2-3 times per week (n = 3, 21%) to weekly (n = 9, 64%). Two studies did not state how frequently they were delivered (14%). Interventions also differed in their duration, with half of the included studies lasting 1-3 months (n = 7, 50%); three lasting 3-6 months (21%); two lasting 6-12 months (14%); and one lasting up to two years (7%). One study did not state how long the intervention lasted (7%).

Given the significant heterogeneity in the time commitment, frequency, and duration of the interventions, treatment intensity was classified by a senior

researcher into three distinct tiers of intensity to facilitate the moderator analysis. Very intensive programmes involved substantial time commitments (often ≥6 months, multiple weekly sessions, and multi-component delivery), while moderately intensive programmes typically consisted of weekly or bi-weekly sessions over 8–16 weeks. Low-intensity interventions were brief, usually involving 10 or fewer sessions. Further details on programme intensity are provided in Appendix 6.

Interventions were predominantly delivered in clinical, outpatient, or mental health centre settings (n = 8, 57%) or in residential, inpatient, or juvenile justice institutions (n = 4, 29%). One study was conducted in a school setting (7%), and one used a combination of clinic and home visits (7%). Interventions were delivered by a range of professionals, with the vast majority delivered by a counsellor, therapist, psychologist, or other health professional (n = 9, 64%). Two studies used a combination of therapists alongside other professionals such as firefighters or correctional facility staff (14%), and two were delivered by dedicated programme facilitators, trainers, or service providers (14%). One intervention was delivered primarily by clinical social workers (7%).

In terms of demographic representation, most studies had mixed-sex samples (n = 9, 64%). Of these mixed-sex samples eight were predominantly male (57%), and one was majority female (7%).

In terms of developmental stages, the studies were evenly split; half focused primarily on children and pre-adolescents typically under the age of 13 (n = 7, 50%), while the other half targeted older adolescents and young adults ranging from 13 to 25 years of age (n = 7, 50%)[1].

Regarding primary clinical presentations and inclusion criteria, participants most frequently presented with disruptive behaviour disorders, conduct problems, or clinically significant aggression (n = 5, 36%). Other primary participant groups included young people who have been adjudicated or incarcerated (n = 3, 21%), young people with a primary diagnosis of autism spectrum disorder (ASD) (n = 2,

[1] Where studies include wider age bands, data must be extractable for participants aged 0–17 (or the sample mean age must be ≤17 and the vast majority of participants under 18) for inclusion.

14%), and young people with primary substance abuse disorders (n = 2, 14%). The remaining studies focused on highly specific clinical presentations, including disruptive mood dysregulation disorder (DMDD) (n = 1, 7%) and clinical fire setting behaviour (n = 1, 7%). Information on the concurrent use of psychiatric medication varied across the literature. Five studies explicitly reported the baseline medication status of their participants (36%). One study explicitly excluded participants who were taking psychiatric medications at the time of the trial (7%), while the remaining eight studies did not provide explicit data on participants medication use during the study (57%).

Reporting on the ethnic and racial composition of participants varied considerably across the literature. Of the 14 unique studies, eight provided specific statistical breakdowns of participants' race or ethnicity (57%), two reported more broadly on nationality or immigration status rather than specific racial or ethnic groups (14%), and four did not report any demographic information related to ethnicity (29%). Four studies featured highly diverse samples in which participants from Black and Global Majority backgrounds made up the clear majority of the sample (29%). A further four studies reported mixed samples that were majority white but still included notable representation of children and young people from Black and Global Majority backgrounds, ranging from approximately 30% to 60% of the sample (29%). Two studies reported highly homogeneous, predominantly white samples: one study included a sample that was nearly 90% white, and another explicitly noted an under-representation of immigrant families, resulting in a predominantly native Dutch, white sample.

In terms of socio-economic status, only one study reported that the majority of families in the sample were on low incomes or receiving financial assistance (7%), while three reported that some families were on low incomes or had mixed income distributions (21%), and one explicitly noted a predominantly middle- to high-income sample (7%). The remaining nine studies (64%) did not provide explicit information on the socio-economic status or income of participants.

Measured outcomes

Across the 14 effectiveness studies, 128 outcomes were identified and extracted, spanning **11 outcome categories** within the YEF Outcomes Framework:[2]

- Crime and offending (k=35, n=6[3])
- General mental health (k=27, n=5)
- Behavioural difficulties (k=26, n=10)
- Regulating and managing emotions (k=11, n=5)
- Drug and alcohol use (k=8, n=2)
- Parenting practices (k=7, n=2)
- Building and maintaining relationships (k=3, n=2)
- Ability to resolve conflict (k=2, n=2)
- Helping others (prosocial behaviour) (k=6, n=1)
- Family relationships and support (k=2, n=1)
- Self-esteem (k=1, n=1)

To inform our violence estimate, outcomes must be a measure of violent crime or violent offending, or a measure of overt or relational aggression that was directed at another person. Measures of bullying can inform the violence estimate where the study reports bullying as repeated physical violence or repeated verbal harassment.

Across the 9 studies that met this criteria, we identified 28 violence outcomes. Thirteen of which were related to behavioural difficulties and 15 related to crime and offending.

Collectively, these measures assess aggression, including overt hostility, maladaptive anger coping strategies, and general behavioural difficulties, as well as severe behavioural proxies, such as structured risk assessments, specific destructive interests, and multi-year recidivism rates for violent and property

[2] The YEF Outcomes Framework identifies specific outcomes linked to reducing the risk of children and young people becoming involved in crime and violence, providing a structured approach for measuring the impact of interventions.

[3] In this context, k = number of outcomes measured, and n = number of studies.

offenses to capture both the immediate form, clinical severity, and longitudinal risk of aggressive and offending behaviour across youth populations.

Outcomes were derived in several ways, most predominantly through a multi-informant approach utilising standardised survey measures (k=87; 68%), clinician assessment/observation (k=10; 8%), and administrative records (k=31; 24%).

Survey measures were reported by parents (k=46; 36%) or teachers (k=7; 5%) and self-reported survey measures completed by the young people themselves (k=34; 27%).

We present summary results from two separate multivariate meta-analyses on violence, and crime and offending below (Table 1).

Table 1: *Summary of findings on violence and crime outcomes*

Outcome	SMD	CI (95%)	*P*	% reduction	Impact rating	Number of studies	Evidence rating
Violence only	-0.711	-1.55 to 0.13	0.096	69%	High	9	3
Crime and Offending	-1.29	-2.16, -0.41	0.004 **	88%	High	6	2

Meta-analysis of violence outcomes in Clinical CBT interventions

Clinical CBT is estimated to have a high impact on violence, corresponding with a 69% reduction in violence, based on 28 measured outcomes across nine studies.

The primary focus in the initial analysis is the reduction and prevention of violence, as defined by YEF. Violence is a broad construct that incorporates incidents/behaviours as well as convictable offences. Violence may be of a physical, verbal, psychological, or sexual nature (YEF, 2023: p.12).

The team identified 28 outcomes measuring **violence** specifically across nine studies: 1 Type A, 7 Type C, and 1 Type D, giving an overall security rating of **Level 3**.

Most violence outcomes measured forms of aggression e.g. 'overt aggression' with several outcomes measuring violent offending.

A total of $k = 28$ outcomes were included in the analysis. The estimated average outcome based on the random-effects model was $\hat{\mu} = -0.71$, (95% CI: -1.55, 0.13); p<0.1.

This estimate was statistically significantly different from zero $t = -1.67$, $p < 0.01$, and although the effect remained the same (g=-0.71) when adjusting for clustering across studies using robust variance estimation, the p-value increased, and became non-significant at the 0.1 level (t = -1.67, df = 7.12, p = .139).

Table 2: *RVE Output for meta-analysis on violence outcomes*

	Estimate	SE	t-stat	d.f (Satt)	p-val (Satt)	Sig	YEF impact rating	YEF evidence security rating
Intercept	-0.71	0.43	-1.67	7.12	0.139	NS	High	3

In practical terms, this indicates that the intervention had a meaningful impact on violence-related outcomes across nine studies. Based on YEF impact categorisation, the effect size (g = -0.71) corresponds to a "high impact" rating. However, as the p-value is above .05, the reader cannot rule out the possibility that the true average effect is actually zero. The prediction interval is very wide (-3.17 to 1.74) and gives us the range of effects a new (but similar) study might plausibly find.

According to the Q-test, high heterogeneity was present; I^2 = 81.5%, Q(df = 27) = 145.895, p <.001. This level of variability suggests that the studies are inconsistent with each other, beyond what would be expected by chance.

The SMD of -0.71 corresponds to a relative risk reduction of 68.9% in the risk of re-committing violence, with an absolute risk reduction of 11.02%. Among those who participate in Clinical CBT, 5% go on to be involved in violence, compared to 16% of those who do not receive Clinical CBT (Figure 1).

YOUTH ENDOWMENT FUND

NATIONAL CHILDREN'S BUREAU

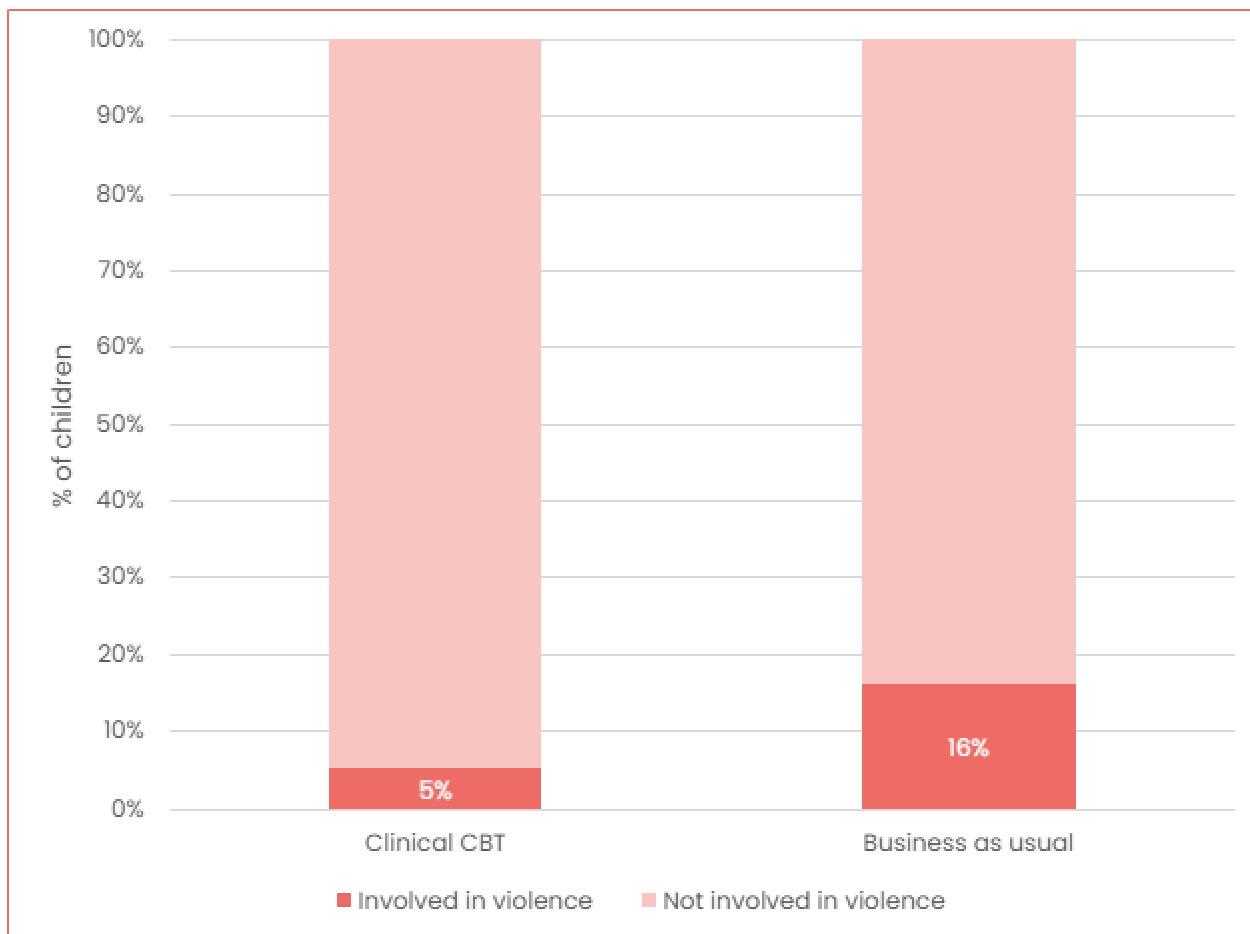

Figure 1: *Absolute risk reduction in violence from Clinical CBT interventions compared to business as usual*

Meta-analysis of crime and offending outcomes in Clinical CBT interventions

Clinical CBT may reduce crime and offending, corresponding with an 88% decrease (a high impact), based on 35 measured outcomes across six studies.

The team identified 35 outcomes measuring **crime and offending** specifically across six studies: 2 Type A, 3 Type C, and 1 Type D giving an overall security rating of **Level 2**. Most crime and offending outcomes measured official criminal involvement and recidivism, e.g. 'general recidivism' and 'total convictions' with several outcomes measuring specific adverse legal consequences such as 'arrests'.

A total of $k = 35$ outcomes were included in the analysis. The estimated average outcome based on the multilevel random-effects model was $\hat{\mu} = -1.29$, (95% CI: --2.16, -0.41). This estimate was statistically significantly different from zero in the initial model *t* = -2.89, *p* = .0039. When adjusting for clustering across studies using robust variance estimation, the effect remained the same (g = -1.29), but the p-value increased slightly and became borderline at the 0.05 level (t = -2.83, df = 3.8, p = .0502), although it remained statistically significant at the 0.1 level.

Table 3: *RVE Output for meta-analysis on crime and offending outcomes*

	Estimate	SE	t-stat	d.f (Satt)	p-val (Satt)	Sig	YEF impact rating	YEF evidence security rating
Intercept	-1.29	0.45	-2.83	3.8	0.0502	.	High	2

In practical terms, this indicates that the intervention had a large and meaningful impact on crime-related outcomes. Based on YEF impact categorisation, the effect size (g = -1.29) corresponds to a high impact rating. However, the prediction interval is wide (-3.21 to 0.64), giving the range of effects that a new but similar study might plausibly find. This suggests that although the average effect is strongly beneficial, future studies may still find smaller effects or no clear effect.

According to the Q-test, moderate heterogeneity was present; I^2 = 32.1%, Q(df = 34) = 50.10, p = .037. This level of variability suggests some inconsistency between studies, although heterogeneity was substantially lower than in the violence analysis.

The SMD of -1.29 corresponds to a relative risk reduction of 87.55% in the risk of being involved in crime and offending, with an absolute risk reduction of 21.89%. In simple terms, among those who participate in Clinical CBT, 3% go on to be involved in crime or offending, compared to 25% of those who do not receive Clinical CBT (Figure 2).

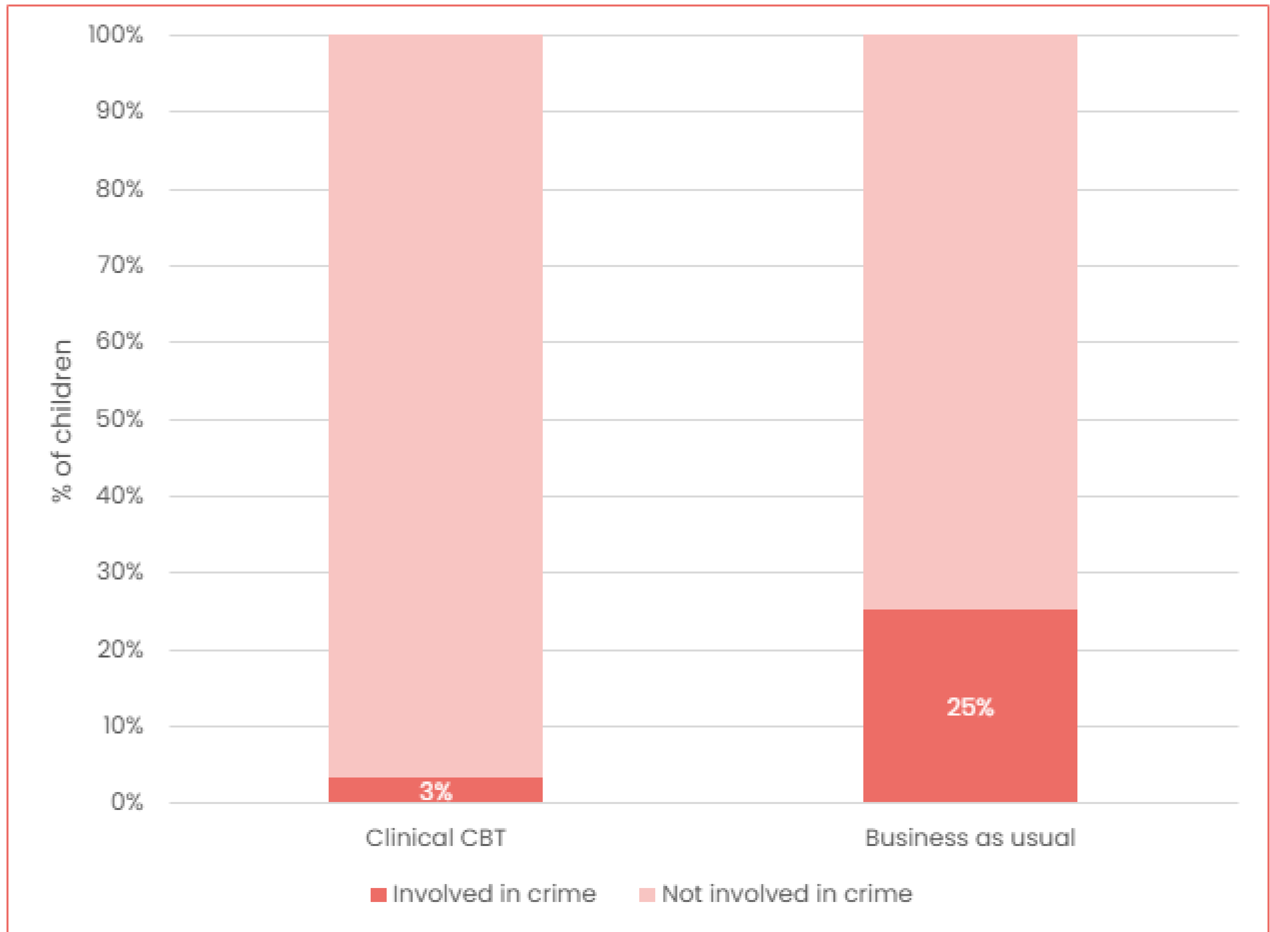

Figure 2: *Absolute risk reduction in crime and offending from Clinical CBT interventions compared to business as usual*

Meta-analysis of all related outcomes in Clinical CBT interventions

Including the violence, crime, and offending outcomes described above, the team extracted a total of 128 measured outcomes from 14 effectiveness studies that align with the YEF outcomes framework. These studies included 2 Type A, 10 Type C, 2 Type D, giving an overall security rating of **Level 3**.

A total of $k = 128$ outcomes were included in the analysis. The estimated average outcome based on the random-effects model was $\hat{\mu} = -0.48$, (95% CI: -0.88, -0.08).

This estimate was statistically significantly different from zero, $t = -2.38$, $p < 0.05$, and remained large and statistically significant when adjusting for clustering

across studies using robust variance estimation ($t = -2.37$, $df = 11.6$, $p < 0.05$). The estimate remained the same ($g = -0.48$) suggesting the original meta-analysis was stable.

Table 4: *RVE Output for meta-analysis of all related outcomes*

	Estimate	SE	t-stat	d.f (Satt)	p-val (Satt)	Sig	YEF impact rating	YEF evidence security rating
Intercept	-0.48	0.20	-2.37	11.6	<0.05	*	High	3

In practical terms, this indicates that Clinical CBT had a meaningful overall impact across the full set of outcomes examined. The pooled effect size suggests that, on average, participants receiving Clinical CBT had better outcomes than those in the comparison conditions. However, the prediction interval was wide (-1.92 to 0.95), giving the range of effects that a new but similar study might plausibly find. This suggests that although the average effect is beneficial, future studies may still find smaller effects or no clear effect.

According to the Q-test, high heterogeneity was present; $I^2 = 80.0\%$, $Q(df = 127) = 634.92$, $p < .001$. This level of variability suggests substantial inconsistency across studies, beyond what would be expected by chance alone.

Such significant variability suggests that the differences in effect sizes are not solely due to random chance but may be influenced by specific study characteristics or contexts. To explore potential sources of this heterogeneity, conducting moderator analyses is recommended.

Subgroup analysis

Details on the moderator categories are provided in Appendix 6.

Intensity of the programme

To explore whether programme effectiveness varied by intensity of the programme, subgroup meta-analyses were conducted (Table 5).

Table 5: *Subgroup analysis on programme intensity*

Intensity subgroup	k	n	SMD (robust SE)	p-value	YEF impact rating
Brief / Low Intensity	14	3	−0.53 (0.33)	0.266	High
Moderate / Standard Clinical Intensity	75	6	−0.12 (0.09)	0.307	Moderate
High Intensity / Long-Term Interventions	39	5	−0.91 (0.78)	0.341	High
Between-group heterogeneity (Q_n test); 3 subgroups, Q_n = 1.19, p = .303					

Note. Estimates computed using robust variance estimation (RVE) with CR2 adjustment for dependent effect sizes (Tipton, 2015). Subgroup-specific p-values are based on robust t-tests with Satterthwaite small-sample degrees of freedom. The omnibus test of differences across outcome domains was conducted using a CR2-robust Wald chi-square test. Significance levels: $p < .05$ (*), $p < .01$ (**), $p < .001$ (***).

Clinical CBT treatment interventions showed negative effect size estimates across all three intensity subgroups, suggesting effects in the beneficial direction. The largest estimated effect was observed for High Intensity / Long-Term interventions (SMD = −0.91, k = 39 effect sizes, 5 studies), followed by Brief / Low Intensity interventions (SMD = −0.53, k = 14, 3 studies), while High Intensity / Long-Term Interventions showed a smaller estimated effect (SMD = −0.12, k = 75, 6 studies). However, none of these subgroup-specific effects reached statistical significance, with p-values ranging from .266 to .341.

There was also no evidence that the effectiveness of Clinical CBT treatment interventions differed significantly by intensity subgroup. The robust Wald chi-square test for between-group heterogeneity was non-significant, $\chi^2(2) = 1.19$, $p = .303$, indicating that variation in effect sizes across brief, moderate, and high-intensity interventions was not greater than would be expected by chance. In practical terms, this suggests that intervention intensity was not a clear

moderator of treatment effects in this analysis, although the small number of studies in each subgroup means these findings should be interpreted cautiously.

Inclusion criteria of the participants

To explore whether programme effectiveness varied based on the inclusion criteria of the trial, subgroup meta-analyses were conducted (Table 6).

Table 6: *Subgroup analysis on inclusion criteria*

Inclusion criteria	k	n	SMD (robust SE)	p-value	YEF impact rating
CYP with an offending history	50	6	−0.21 (0.19)	0.332	Moderate
Externalising problems or difficulties (i.e. conduct disorder)	26	3	−0.09 (0.13)	0.596	Small
Neurodiversity	19	2	−0.16 (0.14)	0.46	Moderate
Other psychiatric diagnosis	33	3	−1.05 (0.95)	0.43	High
Between-group heterogeneity (Q_n test); 4 groups; Q_n = 0.42; p = .742					

Across inclusion-criteria subgroups, none of the estimated effects were statistically significant. Young people included on the basis of an offending history showed a small, non-significant effect favouring the intervention (SMD = −0.21, k = 50, n = 6, p = .332). Likewise, effects were small and non-significant for samples defined by externalising problems or difficulties (SMD = −0.09, k = 26, n = 3, p = .596) and neurodiversity (SMD = −0.16, k = 19, n = 2, p = .460). The subgroup for other psychiatric diagnoses showed the largest estimated effect (SMD = −1.05), but this estimate was highly imprecise and remained non-significant (k = 33, n = 3, p = .430).

The test for between-group heterogeneity was not statistically significant. A CR2-robust Wald chi-square test indicated no evidence that intervention effects differed by inclusion criteria, $\chi^2(3) = 0.42$, p = .742. In practical terms, this suggests that the observed effects were broadly comparable across the four inclusion-criteria categories, although the very small number of studies in some subgroups means this finding should be interpreted cautiously.

Intervention Setting

To explore whether programme effectiveness varied by intervention setting, subgroup meta-analyses were conducted (Table 7).

***Table 7*: Subgroup analysis on intervention setting**

Intervention setting	k	n	SMD (robust SE)	p-value	YEF impact rating
Community-based	46	6	−0.29 (0.06)	.014 *	High
Justice settings	33	2	+0.07 (0.02)	0.146	Harmful
Residential / care-based	29	4	−0.05 (0.08)	0.604	Small
Unclear	20	2	−2.42 (0.36)	0.093	High
Between-group heterogeneity (Q_n test); 4 settings; Q_n = 29.1; p < .001 ***					

Clinical CBT demonstrated variation in effectiveness across intervention settings. For community-based interventions (k = 46 effect sizes, 6 studies), which referred to outpatient therapy delivered in the community after adolescents had been discharge from hospital, there was a statistically significant beneficial effect favouring CBT treatment (SMD = −0.29, p = .014). In contrast, effects were not statistically significant for justice settings (SMD = +0.07, k = 33, p = .146) or residential/care-based settings (SMD = −0.05, k = 29, p = .604). The justice setting

estimate was very small and in the unfavourable direction, but should not be interpreted as reliable evidence of harm given the non-significant p-value and small number of studies. The estimate for interventions delivered in settings classified as unclear was large and negative (SMD = −2.42, k = 20), but this effect did not reach conventional levels of statistical significance (p = .093) and should be interpreted cautiously given the very small number of studies.

Finally, a Q-test for between-group heterogeneity indicated that the impact of interventions varies significantly by intervention setting. A robust Wald chi-square test indicated significant between-group heterogeneity across settings ($\chi^2(3)$ = 29.1, p < .001), providing strong evidence that effect sizes differ depending on where the intervention was delivered. In practical terms, this suggests that Clinical CBT interventions may be more beneficial in some settings than others, with the clearest evidence of benefit observed in community-based contexts. However, this finding should be interpreted cautiously, as differences by setting may partly reflect variation in participant risk, need or complexity rather than the effect of setting alone. The especially large estimate for the "unclear" category should be treated with caution, as it is based on only two studies and may reflect instability rather than a true underlying effect.

Intervention Modality

To explore whether programme effectiveness varied by mode of delivery, subgroup meta-analyses were conducted (Table 8).

***Table 8*: Subgroup analysis on mode of delivery**

Method of delivery	k	n	SMD (robust SE)	p-value	YEF impact rating
Groups, face-to-face	56	6	−0.14 (0.10)	0.27	Moderate
Individuals and groups, face-to-face	29	2	−0.44 (0.00)	.002 **	High

Individuals only, face-to-face	43	6	−0.94 (0.67)	0.248	High
Between-group heterogeneity (Q_n test); 3 methods; Q_n = 4.51; p = .011*					

Subgroup analyses by method of delivery suggested some variation in effects across delivery formats. Interventions delivered in groups only, face-to-face showed a small non-significant beneficial effect (SMD = −0.14, k = 56 effect sizes, 6 studies, p = .270). Interventions delivered to individuals only, face-to-face showed a larger beneficial effect in magnitude (SMD = −0.94, k = 43, 6 studies), although this estimate was also not statistically significant (p = .248), reflecting substantial uncertainty around the effect size. By contrast, interventions delivered through a combination of individuals and groups, face-to-face were associated with a statistically significant beneficial effect (SMD = −0.44, k = 29, 2 studies, p = .002). However, this estimate should be interpreted cautiously because it was based on only two studies and accompanied by an extremely small robust standard error, which, in robust variance estimation meta-analysis, can signal unstable variance estimates when study numbers are limited and may overstate precision.Finally, a Q-test for between-group heterogeneity indicated significant differences across delivery methods. A robust Wald chi-square test showed significant between-group heterogeneity ($\chi^2(2)$ = 4.51, p = .011), indicating that intervention effects differed according to how the intervention was delivered. In practical terms, this suggests that delivery format may matter for intervention effectiveness, although the very small number of studies in some categories means these findings should be treated with caution.

Family/parent Involvement

To explore whether programme effectiveness varied due to the additive effects of combining child CBT with family therapy or Parent Management Training, subgroup meta-analyses were conducted (Table 9).

***Table 9*: Subgroup analysis on family involvement**

Family involvement	k	n	SMD (robust SE)	p-value	YEF impact rating
No family component mentioned	14	3	−0.20 (0.34)	.615	Moderate
Parent check-ins / optional component only	49	4	−0.26 (0.18)	.358	High
Parent training or family sessions included	65	7	−0.55 (0.42)	.271	High
Between-group heterogeneity (Q_n test); 3 groups; Q_n = 0.23; p = .795					

Clinical CBT interventions did not show statistically significant effects within any family-involvement subgroup. Programmes with no family component mentioned showed a small, non-significant effect (SMD = −0.20, k = 14, p = .615). Interventions with parent check-ins or an optional family component also showed a small, non-significant effect (SMD = −0.26, k = 49, p = .358). The largest estimated effect was observed for interventions that included parent training or family sessions as part of the CBT protocol (SMD = −0.55, k = 65, p = .271), although this effect was also not statistically significant.

The test for between-group heterogeneity was not significant. A robust Wald chi-square test indicated no evidence that intervention effects differed across family-involvement subgroups, $\chi^2(2) = 0.23$, p = .795. In practical terms, this suggests that the presence or intensity of a family component was not associated with meaningful differences in intervention effectiveness in this analysis.

Outcome type

To explore whether programme effectiveness varied due to the outcome measured, subgroup meta-analyses were conducted (Table 9).

***Table 10*: Subgroup analysis on outcome domain**

Outcome domain	k	n	SMD (robust SE)	p-value	YEF impact rating
Crime and offending	35	6	−1.94 (0.72)	0.111	High
General Mental Health	27	5	−0.59 (0.42)	0.289	High
Behavioural Difficulties	26	10	−0.12 (0.07)	0.165	Moderate
Regulating and Managing Emotions	11	5	−0.61 (0.53)	0.339	High
Drug and Alcohol Use	8	2	−0.33 (0.12)	0.223	High
Parenting practices	7	2	−0.07 (0.11)	0.642	Small
Helping others (prosocial behaviours)	6	1	+0.03 (0.00)	.043 *	Harmful
Building and maintaining relationships	3	2	+0.06 (0.06)	0.504	Harmful

Ability to resolve conflict	2	2	−1.23 (0.04)	.023 *	High
Family relationship and support	2	1	−0.02 (0.00)	.005 **	No effect
Self-esteem	1	1	n/a †		

The subgroup results indicate statistically significant effects for three outcome domains for Clinical CBT treatment interventions. Statistically significant beneficial effects were observed for ability to resolve conflict (SMD = −1.23, k = 2, n = 2, p = .023) and family relationship and support (SMD = −0.02, k = 2, n = 1, p = .005), as negative effect sizes indicate better outcomes in this meta-analysis. By contrast, a statistically significant effect was also observed for helping others/prosocial behaviours (SMD = +0.03, k = 6, n = 1, p = .043), but here the positive effect size indicates a worse outcome for participants receiving Clinical CBT treatment interventions relative to the comparison group. However, this effect was very small and based on a single study, so it should be interpreted cautiously and should not be taken as strong evidence of harm. No statistically significant effects were observed for behavioural difficulties, building and maintaining relationships, crime and offending, drug and alcohol use, general mental health, parenting practices, or regulating and managing emotions.

In substantive terms, the pattern suggests that Clinical CBT treatment interventions may have more evident effects for some interpersonal and family-related outcomes than for other domains, although the direction of effects is mixed. The largest estimated beneficial effect was for ability to resolve conflict, although this estimate is based on only two effect sizes from two studies and very limited effective degrees of freedom, so it should be interpreted cautiously. Likewise, the statistically significant beneficial result for family relationship and support is based on only one study. In contrast, the statistically significant result for helping others/prosocial behaviours is in the adverse direction, indicating poorer prosocial outcomes for the intervention group, although this finding is also based on only one study and is accompanied by an extremely small standard

error. These features may reflect sparse data or model instability. The self-esteem estimate could not be formally tested because the standard error was reported as zero and the associated inferential statistics were undefined. The Q-test for between-group heterogeneity was not estimable because several outcome domains were supported by too few independent studies, resulting in a non-positive-definite robust contrast covariance matrix.

Overall, these findings suggest selective rather than uniform effects of Clinical CBT treatment interventions across outcome domains. Under the coding used in this meta-analysis, negative effect sizes indicate beneficial effects and positive effect sizes indicate worse outcomes. On that basis, the evidence points to some beneficial effects for conflict resolution and family relationship/support outcomes, alongside a potentially adverse effect for prosocial behaviour, while the remaining domains showed no statistically significant differences between groups.

Country

To investigate potential differences in the effectiveness of Clinical CBT programmes, subgroup analyses were conducted by country (Table 11).

Table 11: *Subgroup analysis on country*

Country	k	n	SMD (robust SE)	p-value	YEF impact rating
Netherlands	40	3	−0.12 (0.02)	0.104	Moderate
Scotland	3	1	−1.45 (0.01)	.003 **	High
Sweden	18	1	−0.00 (0.00)	0.492	No effects
USA	53	7	−0.19 (0.10)	0.133	Moderate
Canada	3	1	−0.44 (0.00)	.002 **	High
Iran	11	1	−2.46 (0.18)	.047 *	High

YOUTH ENDOWMENT FUND

NATIONAL CHILDREN'S BUREAU

Between-group heterogeneity (Q_n test); 6 countries; $\chi^2(5) = 6933$; $p < .001$ ***

Country subgroup analyses indicated significant beneficial effects of Clinical CBT interventions in some national contexts, although the evidence base was highly uneven across countries. Significant effects favouring Clinical CBT interventions were observed in Scotland (k = 3, n = 1, SMD = −1.45, p = .003), Canada (k = 3, n = 1, SMD = −0.44, p = .002), and Iran (k = 11, n = 1, SMD = −2.46, p = .047).

By contrast, effects were not statistically significant in the Netherlands (SMD = −0.12, k = 40, p = .104), Sweden (SMD = −0.00, k = 18, p = .492), or the USA (SMD = −0.19, k = 53, p = .133). Although the USA and Netherlands contributed the largest numbers of effect sizes, neither subgroup reached conventional levels of statistical significance.

Finally, the omnibus test indicated significant between-group heterogeneity across countries. A robust Wald chi-square test showed that effect sizes differed significantly by country, $\chi^2(5) = 6933$, $p < .001$. In practical terms, this suggests that the apparent effectiveness of Clinical CBT interventions varies across national contexts. However, this finding should be interpreted cautiously because several country subgroups were based on only one study, limiting confidence in the precision and generalisability of those subgroup-specific estimates.

Publication year

Subgroup meta-analyses were conducted by decade to explore variation in effect sizes over time, given the wide range of publication years across included studies (Table 12).

Table 12: **Subgroup analysis on publication year**

Publication period	K	n	SMD (robust SE)	p-value	YEF impact rating
2000-2009	26	6	−0.28 (0.18)	0.197	High
2010-2019	50	3	−0.19 (0.12)	0.367	Moderate

2020-2029	52	5	−0.68 (0.65)	0.392	High
Between-group heterogeneity (Q_n test); 3 time periods; $\chi^2(2) = 0.33$; p = .717					

Across publication decades, Clinical CBT interventions did not show statistically significant effects in any subgroup. The estimated effect was largest in studies published in the 2020s (SMD = −0.68, k = 52 effect sizes, 5 studies), but this was not statistically significant (p = .392) and was accompanied by substantial uncertainty. Effects were also non-significant for studies published in the 2010s (SMD = −0.19, k = 50, 3 studies, p = .367) and the 2000s (SMD = −0.28, k = 26, 6 studies, p = .197).

A Q-test for between-group heterogeneity indicated that the effectiveness of Clinical CBT interventions did not vary significantly by publication decade. A robust Wald chi-square test showed no evidence of between-group heterogeneity across decades ($\chi^2(2) = 0.33$, p = .717), suggesting that any apparent differences in effect sizes between studies published in the 2000s, 2010s, and 2020s are consistent with chance variation rather than meaningful subgroup differences.

Age band

Subgroup meta-analysis was conducted to examine whether the effectiveness of Clinical CBT interventions varied by the age range of participants. Studies were grouped into the age categories available in the dataset: Children (<13) and Adolescents (13+) (Table 13).

***Table 13*: Subgroup analysis on developmental age band**

Age range	k	n	SMD (robust SE)	p-value	YEF impact rating
Adolescents (13+)	66	7	−0.28 (0.13)	0.117	High
Children (<13)	62	7	−0.57 (0.50)	0.327	High
Between-group heterogeneity (Q_n test); 2 age bands; χ^2 (1) = 0.31; p =.578					

Subgroup analyses by age range did not identify any statistically significant effects for either children or adolescents. For adolescents aged 13 and over (k = 66 effect sizes, n = 7 studies), Clinical CBT interventions were associated with a small beneficial effect (SMD = −0.28, p = .117), although this did not reach statistical significance. For children under 13 (k = 62, n = 7), the estimated effect was somewhat larger in magnitude (SMD = −0.57), but this estimate was also not statistically significant (p = .327).

The omnibus test showed no significant between-group heterogeneity by age range. A robust Wald chi-square test indicated that effect sizes did not differ significantly between children and adolescents, $\chi^2(1) = 0.31$, p = .578. In practical terms, this suggests there is no clear evidence from the present dataset that Clinical CBT interventions are more effective for one of these age groups than the other. Although the number of effect sizes was similar across the two subgroups, the wide standard error for the children subgroup suggests considerable uncertainty around that estimate.

Ethnicity

To explore whether programme effectiveness varied by ethnicity, studies were grouped into three categories: Majority white (>85% white participants), Some diversity (15–49% Black and Global Majority), and Diverse/balanced (≥50% Black and Global Majority) (Table 14).

***Table 14*: Subgroup analysis on ethnicity**

Ethnicity	k	n	SMD (robust SE)	p-value	YEF impact rating
Diverse/balanced	46	4	−1.43 (1.01)	0.322	High
Majority white	30	2	−0.22 (0.06)	0.178	Moderate

Not stated	28	4	−0.18 (0.24)	0.556	Moderate
Some diversity	24	4	−0.08 (0.16)	0.651	Small
Between-group heterogeneity (Q_n test); 4 subgroups; χ^2 (3) = 0.71; p = .545					

Across ethnicity subgroups, Clinical CBT interventions did not demonstrate statistically significant effects within any category. The largest estimated effect was observed in samples classified as diverse/balanced (SMD = −1.43, k = 46 effect sizes, 4 studies), but this estimate was imprecise and not statistically significant (p = .322). Effects were also non-significant in majority white samples (SMD = −0.22, k = 30, p = .178), not stated samples (SMD = −0.18, k = 28, p = .556), and samples with some diversity (SMD = −0.08, k = 24, p = .651).

A Q-test for between-group heterogeneity indicated that the effectiveness of Clinical CBT interventions did not vary significantly across ethnicity subgroups. Consistent with this, a robust Wald chi-square test showed no evidence of between-group heterogeneity (χ^2(3) = 0.71, p = .545), indicating that the analysis did not detect systematic variation in effect size by ethnicity.

Gender

To investigate whether the effectiveness of Clinical CBT programmes varied according to the gender composition of the study samples, subgroup analyses were conducted across three categories: studies with only male participants, studies with only female participants, and mixed samples (Table 15).

***Table 15*: Subgroup analysis on gender**

Sample gender composition	k	n	SMD (robust SE)	p-value	YEF impact rating
Mixed sex	96	9	−0.18 (0.08)	0.099	Moderate
Single sex - boys	28	4	−1.10 (0.83)	0.293	High

Single sex - girls	4	1	+0.07 (0.00)	< .001 ***	Harmful
Between-group heterogeneity (Q_n test); 3 gender groupings; χ Q_n = 5.94; p = .003 **					

Across gender subgroups, Clinical CBT interventions did not show statistically significant effects for mixed-sex samples (SMD = −0.18, k = 96 effect sizes, 9 studies, p = .099) or single-sex boys' samples (SMD = −1.10, k = 28, 4 studies, p = .293). In single-sex girls' samples, the estimated effect was small and positive (SMD = +0.07, k = 4, 1 study, p < .001), indicating worse outcomes for the intervention group. However, this subgroup is based on only one study, and the magnitude of effect is very small, so that estimate should be interpreted very cautiously.

A Q-test for between-group heterogeneity indicated that the impact of Clinical CBT interventions varied significantly across gender subgroups. A robust Wald chi-square test showed significant between-group heterogeneity ($\chi^2(2)$ = 5.94, p = .003), suggesting that effect sizes differed by gender composition of the sample. In practical terms, this means the effectiveness of Clinical CBT interventions may not be uniform across mixed-sex, boys-only, and girls-only samples, although the girls-only result is especially uncertain because it is based on a single study.

Multiple meta-regressions (models 1-5)

Following these ten subgroup analyses, we ran five separate random-effects meta-regression models with robust variance estimation to examine pre-specified moderators: (1) quality, (2) setting, (3) intervention features, (4) population characteristics, and (5) outcome characteristics. The subgroup analyses provided unadjusted descriptive estimates within categories, while the meta-regression models formally tested pre-specified moderators and accounted for clustering of multiple effect sizes within studies. See Appendix 1 for an overview of the methods used in this section and Appendix 3 for a list of the studies that provided data for these analyses.

YOUTH ENDOWMENT FUND

NATIONAL CHILDREN'S BUREAU

Moderator Analysis 1 – Quality Moderators

The characteristics in model 1 include:

- Study design
- Study timing
- Unit of allocation into group[4]
- Method of allocation into group
- Quality appraisal as assessed by the YEF-EQA tool

***Table 16*: Meta-regression (with robust inference) on quality moderators**

Moderator	β (estimate)	SE	t	df	p	95% CI
Intercept	−1.44	1.108	−1.30	2.25	0.312	−5.73, 2.86
Design = RCT (vs reference)	0.58	0.469	1.23	1.04	0.429	−4.90, 6.05
Timing = 25 month–3 year follow-up (vs reference)	−0.14	0.036	−3.80	1.68	0.082	−0.32, 0.05
Timing = 3 month–1 year follow-up (vs reference)	−0.35	0.329	−1.08	1.27	0.445	−2.92, 2.21
Timing = Post-intervention (vs reference)	−0.21	0.188	−1.12	1.15	0.445	−1.98, 1.56
Method of allocation = Random (vs reference)	−0.34	0.085	−3.96	6.54	0.006	−0.54, −0.13
Study quality (YEF EQA) = Low	0.71	1.214	0.59	2.01	0.616	−4.49, 5.92

[4] This was dropped from the meta-regression model as only one level, i.e., all studies allocated at the individual level

Study quality (YEF EQA) = Moderate	1.14	1.031	1.11	1.66	0.404	−4.30, 6.58

Note. Values are from a multivariate random-effects meta-regression estimated using REML ($k = 135$), clustering on study ID to account for non-independence of multiple effect sizes within studies. Statistical inference is based on CR2-robust standard errors with Satterthwaite small-sample t-tests.

The estimated between-study variance component was $\tau^2 = 0.38$ (SD = 0.61), indicating substantial between-study heterogeneity. A test for residual heterogeneity remained highly significant, QE(df = 120) = 404.22, $p < .0001$, showing that the moderators included in the model did not fully explain variation in effect sizes. The conventional omnibus test of moderators from the REML model was statistically significant, QM(df = 7) = 15.79, $p = .0271$; however, the cluster-robust coefficient tests suggest that only method of allocation being random was a statistically significant moderator after robust adjustment.

Studies using random allocation had effect sizes that were significantly more negative than studies using quasi-random allocation ($\beta = -0.34$, SE = 0.08, $t = -3.96$, df = 6.54, $p = .006$, 95% CI [-0.54, -0.13]). Given that negative SMD values favour the intervention group in this meta-analysis, this indicates that random allocation was associated with more favourable intervention effects than quasi-random allocation. All other moderators, including study design, study timing, and study quality, were not statistically significant under the robust inference approach, although the 25 month–3 year follow-up category was marginal ($p = .082$). In practical terms, this suggests that, in this model, method of allocation is the only moderator showing reliable evidence of explaining heterogeneity in effect sizes once dependence among effects is taken into account.

Moderator Analysis 2 – Setting-level Moderators

The characteristics in model 2 include:

- Country of intervention
- Setting of intervention

***Table 17*: Meta-regression (with robust inference) on setting-level moderators**

Moderator	β (estimate)	SE	t	df	p	95% CI
Intercept	-0.37	0.24	-1.58	1.98	0.26	-1.40, 0.65
Country = Scotland (vs reference country)	-1.41	0.13	-10.52	1.40	0.03	-2.30, -0.52
Country = Sweden (vs reference country)	0.03	0.19	0.17	2.16	0.88	-0.74, 0.81
Country = USA (vs reference country)	0.08	0.22	0.37	2.34	0.75	-0.74, 0.90
Country = Canada (vs reference country)	-0.07	0.24	-0.29	1.99	0.80	-1.09, 0.95
Country = Iran (vs reference country)	-1.46	0.85	-1.71	1.31	0.29	-7.77, 4.84
Intervention setting = Justice settings (vs reference setting)	0.16	0.28	0.56	1.81	0.64	-1.18, 1.50
Intervention setting = Residential / care-based (vs reference setting)	0.33	0.14	2.31	1.79	0.16	-0.36, 1.03

Note. Netherlands was the reference category for country, and community-based was the reference category for intervention setting. All coefficients for country and setting represent differences from these reference categories

The estimated between-study variance component was $\tau^2 = 0.05$ (SD = 0.21), indicating some remaining variation across studies. A test for residual heterogeneity remained highly significant ($QE(df = 119) = 214.35$, $p < .0001$), showing that the moderators included did not fully account for heterogeneity across effects. Although the conventional omnibus test of moderators was

significant (QM(df = 8) = 88.26, $p < .0001$), the robust results suggest that only the Scotland contrast was statistically significant. In substantive terms, the study from Scotland showed significantly more favourable effects, while the other country and intervention-setting contrasts were not statistically significant.

Moderator Analysis 3 – Intervention-level Moderators

The characteristics in model 3 include:

- Intervention intensity
- Intervention duration
- Intervention provider
- Training for intervention providers
- Delivery method
- Family sessions or components in addition to the CBT protocol

***Table 18*: Meta-regression (with robust inference) on intervention-level moderators**

Moderator	β (estimate)	SE	t	df	p	95% CI
Intercept	0.94	3.22	0.29	3.12	0.79	-9.10, 10.98
Duration = 1–3 months	-0.30	1.19	-0.25	1.87	0.83	-5.79, 5.19
Duration = 3–6 months	-1.63	2.55	-0.64	2.30	0.58	-11.34, 8.09
Duration = 6–12 months	-0.46	1.13	-0.41	2.75	0.71	-4.25, 3.33
Duration = 6–24 months	0.14	2.36	0.06	4.75	0.95	-6.01, 6.29
Duration = Unclear	-1.39	2.14	-0.65	3.90	0.55	-7.38, 4.60
Intensity = High intensity / long-term interventions	0.44	1.52	0.29	2.22	0.80	-5.54, 6.42

Intensity = Moderate / standard clinical intensity	0.03	2.47	0.01	0.27	0.99	-6.01, 6.29
Implementer training = Yes	-1.01	1.52	-0.66	2.20	0.57	-7.02, 5.00
Method = Individuals and groups, face-to-face	-0.92	1.53	-0.60	2.22	0.60	-6.90, 5.07
Method = Individuals only, face-to-face	0.81	1.52	0.53	2.23	0.64	-5.14, 6.76
Provider = Psychologist	-0.61	0.98	-0.62	1.76	0.61	-5.44, 4.22
Provider = Social workers	0.37	2.47	0.15	1.07	0.91	-26.63, 27.36

The estimated between-study variance component was τ^2 = 1.73 (SD = 1.32), indicating substantial variation across studies. A test for residual heterogeneity remained highly significant (QE(df = 115) = 373.72, $p < .0001$), showing that the intervention moderators included did not explain the remaining heterogeneity in effect sizes. In addition, the overall test of moderators was not significant (QM(df = 12) = 2.41, $p = .9985$), indicating that, in this model, intervention duration, intensity, implementer training, delivery method, and provider type did not meaningfully explain variation in effect sizes.

Moderator Analysis 4 – Population-level Moderators

The characteristics in model 4 include:

- Age band
- Ethnicity
- Gender
- Socio-economic status

***Table 19*: Meta-regression (with robust inference) on population-level moderators**

Moderator	β (estimate)	SE	t	df	p	95% CI
Intercept	0.46	1.94	0.24	2.38	0.83	-6.75, 7.66
Age range = Children (<13)	-0.42	1.43	-0.30	1.78	0.80	-7.38, 6.53
Ethnicity = Majority white	0.32	0.70	0.46	1.84	0.69	-2.96, 3.61
Ethnicity = Not stated	0.57	1.42	0.40	1.79	0.73	-6.28, 7.43
Ethnicity = Some diversity	0.44	0.85	0.52	2.88	0.64	-2.33, 3.21
Gender = Single sex – boys	-0.70	0.98	-0.72	2.46	0.54	-4.25, 2.85
Gender = Single sex – girls	0.19	0.73	0.26	1.86	0.82	-3.18, 3.57
SES = Middle to high SES	-0.48	1.48	-0.33	2.62	0.77	-5.59, 4.63
SES = Not stated	-1.02	1.94	-0.52	1.53	0.67	-12.40, 10.37

The estimated between-study variance component was $\tau^2 = 0.58$ (SD = 0.76), indicating substantial variation across studies. A test for residual heterogeneity remained highly significant (QE(df = 119) = 273.16, $p < .0001$), showing that the participant moderators included did not fully account for heterogeneity across effects. The overall test of moderators was not significant (QM(df = 8) = 4.46, $p = .81$), indicating that, in this model, age range, ethnicity, gender composition, and socioeconomic status did not meaningfully explain variation in effect sizes.

Moderator Analysis 5 – Outcome Moderators

The characteristics in model 5 include:

- YEF's outcome category
- Methods used to collect the outcome data

***Table 20*: Meta-regression (with robust inference) on outcome-level moderators**

Moderator	β (estimate)	SE	t	df	p	95% CI
Intercept	-1.27	0.53	-2.41	1.69	0.16	-3.98, 1.45
Outcome = Behavioural difficulties	0.89	0.53	1.67	1.18	0.31	-3.87, 5.66
Outcome = Building and maintaining relationships	1.09	0.56	1.94	2.37	0.17	-1.00, 3.17
Outcome = Crime and offending	0.60	0.57	1.05	1.65	0.42	-2.45, 3.65
Outcome = Drug and alcohol use	1.17	0.59	2.00	1.64	0.21	-1.95, 4.30
Outcome = Family relationship and support	1.08	0.48	2.24	1.35	0.21	-2.30, 4.45
Outcome = General mental health	1.12	0.59	1.90	1.50	0.24	-2.42, 4.66
Outcome = Helping others (prosocial behaviours)	0.59	0.53	1.13	1.14	0.44	-4.44, 5.63
Outcome = Parenting practices	0.81	0.51	1.58	1.38	0.30	-2.67, 4.29
Outcome = Regulating and managing emotions	0.73	0.53	1.38	1.27	0.36	-3.38, 4.84
Outcome = Self-esteem	1.20	0.54	2.21	1.24	0.23	-3.22, 5.62

Data source = Clinician assessment/observation	-0.08	0.24	-0.35	4.02	0.74	-0.75, 0.58
Data source = Official administrative records	0.40	0.15	2.62	1.13	0.21	-1.08, 1.89
Data source = Parent reported	-0.26	0.09	-2.77	3.37	0.06	-0.55, 0.02
Data source = Teacher reported	-0.02	0.12	-0.18	1.49	0.88	-0.75, 0.71

The estimated between-study variance component was $\tau^2 = 0.46$ (SD = 0.68), indicating substantial variation across studies. A test for residual heterogeneity remained highly significant ($QE(df = 113) = 371.63$, $p < .0001$), showing that the moderators included in this model did not fully account for heterogeneity across effects. The conventional omnibus test of moderators was statistically significant ($QM(df = 14) = 35.31$, $p = .0013$), suggesting that outcome domain and/or data source were associated with variation in effect sizes. However, when using robust clustered standard errors and small-sample inference, none of the individual moderator contrasts reached conventional levels of statistical significance,

Publication bias

Publication-bias analyses were conducted on 14 study-level aggregated effect sizes. Egger's regression test did not indicate significant funnel plot asymmetry ($z = -0.93$, $p = .353$), and Begg's rank-correlation test was also non-significant (Kendall's $\tau = -0.27$, $p = .193$), providing no clear evidence of small-study effects. Trim-and-fill estimated that no studies were missing on the right side of the funnel plot, and the adjusted pooled effect remained statistically significant and in the favourable direction ($\beta = -0.48$, SE = 0.19, $z = -2.53$, $p = .011$, 95% CI [-0.84, -0.11]).

A step-function[5] selection model was fitted to examine possible publication bias and/or selective reporting, using two-sided p-value intervals of ≤ .025, .025–.05, .05–.10, and >.10. The model was estimated using a random-effects model with ML estimation[6] based on 14 studies. The adjusted pooled effect remained negative and statistically significant ($\beta = -0.86$, SE = 0.29, $z = -2.95$, $p = .003$, 95% CI [-1.43, -0.29]), indicating that the intervention effect remained in the favourable direction after allowing for possible selection.

The test of the selection model parameters was statistically significant (LRT(3) = 9.07, $p = .028$), suggesting that the probability of studies being observed may have varied across p-value intervals and therefore that selective reporting cannot be ruled out. However, the individual selection weights were very imprecisely estimated, with wide confidence intervals, and none of the individual weight parameters was statistically significant. As illustrated in Figure 3, the estimated pattern of selection did not show a clear trend in which studies with smaller p-values were consistently more likely to be observed. Accordingly, this result should be interpreted cautiously, particularly given the small number of studies ($k = 14$).

[5] A step-function selection model is a publication-bias sensitivity analysis that divides studies into bands according to their p-values and estimates whether studies in some bands are more likely to be observed than others.
[6] ML refers to maximum likelihood estimation, a method used to estimate the model parameters by identifying the values that make the observed data most probable under the assumed model.

Residual heterogeneity remained substantial in the selection model (τ^2 = 0.50, τ = 0.71), and the heterogeneity test remained highly significant (LRT(1) = 454.44, $p < .0001$), indicating considerable between-study variation even after accounting for possible selection.

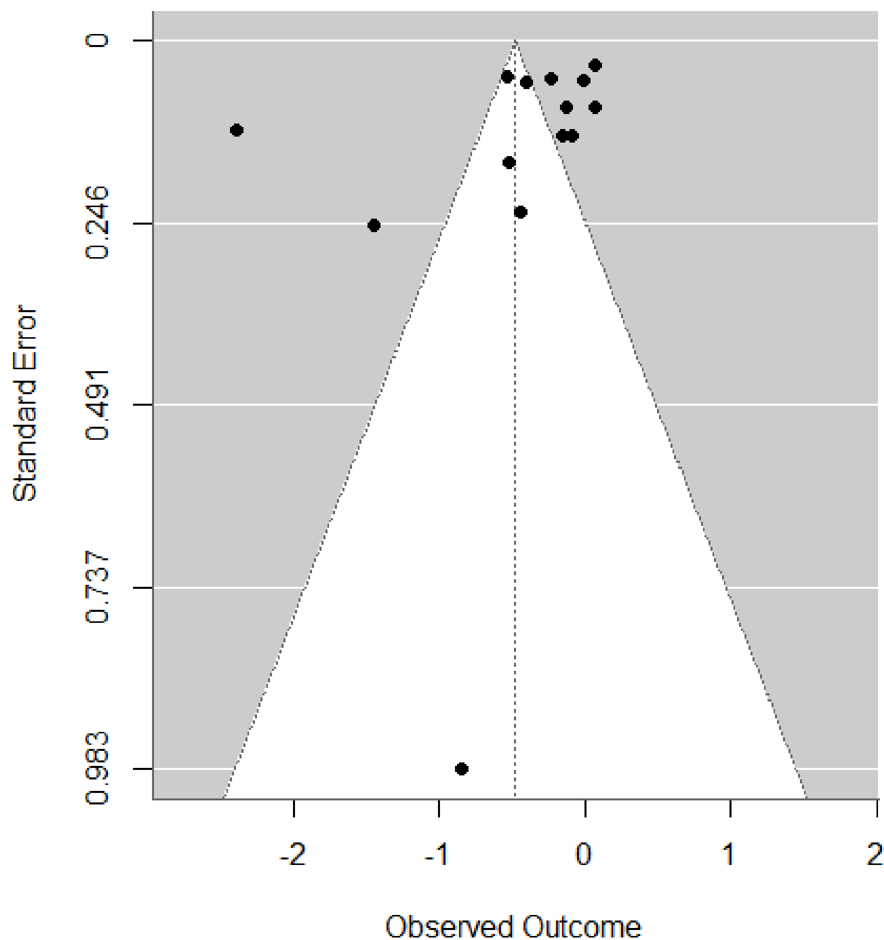

Figure 3**: Funnel plot of individual studies**

Sensitivity analyses

An examination of the studentized residuals revealed that several studies had values larger than ±3.5603 and may be potential outliers in the context of this model. Similarly, according to the Cook's distances, several estimates could be considered to be overly influential (Figure 4). These study-level influence diagnostics indicated that Soleimani-Rad (2025) was notably influential, with the

largest Cook's distance (1.04) and the largest DFBETA[7] for the intercept (-2.01). In addition, several effect sizes from this one study showed large studentised residuals, some exceeding ±3, suggesting potential outlying observations.

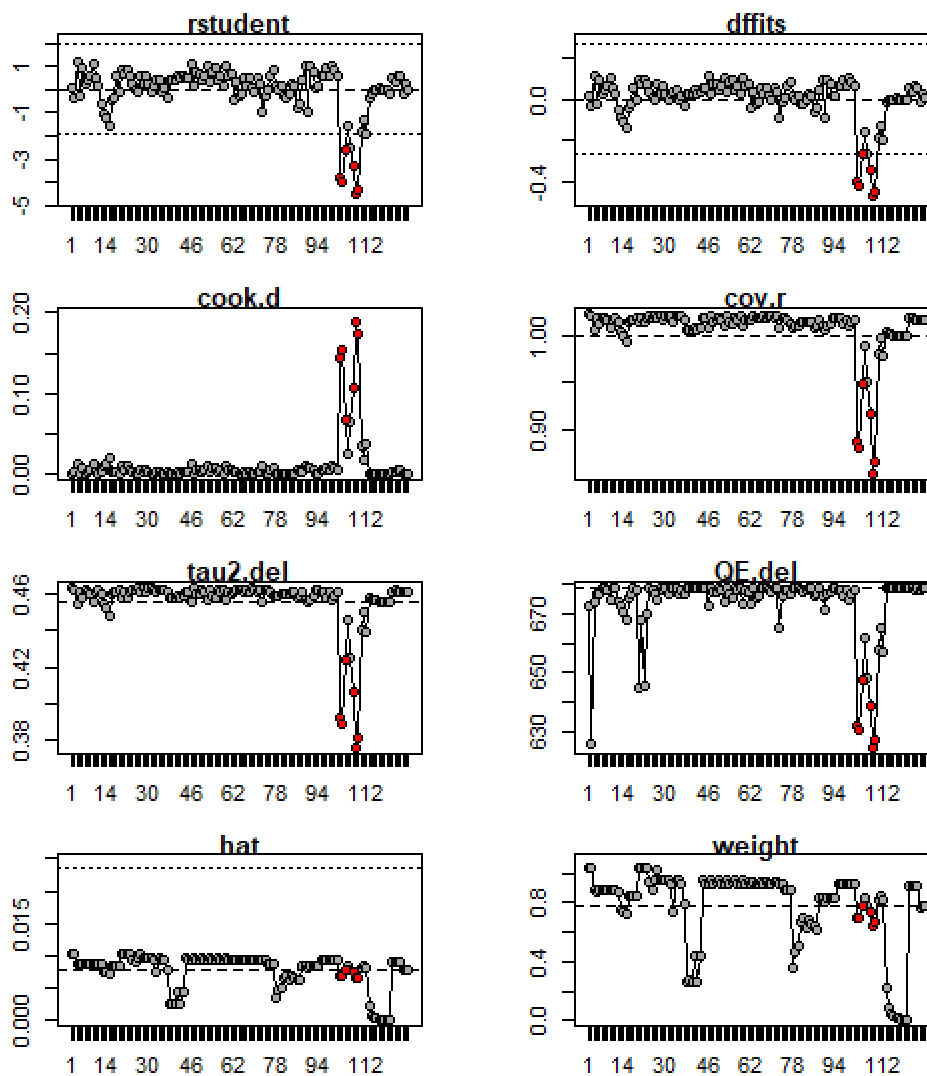

***Figure 4*: Influence diagnostics for individual studies**

First, the team ensured that the outcomes from Soleimani-Rad (2025) were checked to ensure that a mistake had not been made when extracting outcomes,

[7] In metafor, DFBETAS stands for "difference in betas" with beta meaning a model coefficient value. DFBETAS tells the reader how many standard deviations the estimated coefficient changes after excluding that study from the model fit.

when satisfied that was not the case, a leave-one-study-out analysis was performed to evaluate the influence of individual studies on the pooled effect (Figure 5). Removal of each study in turn produced pooled effect estimates that ranged from approximately 0.28 to -0.52 meaning no single study reversed the direction of the effect, indicating that the beneficial findings and conclusions of this meta-analysis were not unduly driven by any individual study. Omitting Soleimani-Rad (2025) reduced the pooled effect to SMD = -0.28 (95% CI [-0.50, -0.07], p = .015) and substantially reduced residual heterogeneity (τ^2 = 0.10), indicating that this study contributed disproportionately to both the effect magnitude and the remaining heterogeneity in the model. Omitting Curran (2009) yielded the only iteration in which the confidence interval included zero (SMD = -0.40, 95% CI [-0.81, 0.01], p = .054), although the estimated effect remained in the same favourable direction.

To assess the robustness of the overall effect (SMD = -0.48, SE = 0.19) based on quality decisions, a series of sensitivity analyses were conducted. First, analyses were repeated after excluding studies rated as 'low' quality. The overall pooled effect remained statistically significant and almost identical to the main analysis (SMD = -0.48, SE = 0.19, p =.0261), indicating that the primary findings were not driven by lower-quality studies.

Second, the analysis was restricted to RCTs only, excluding quasi-experimental designs. The pooled effect estimate in this subset was again almost identical to the main result and remained statistically significant (SMD = -0.48, SE = 0.20, p = .037), suggesting that study design did not materially influence the overall conclusion.

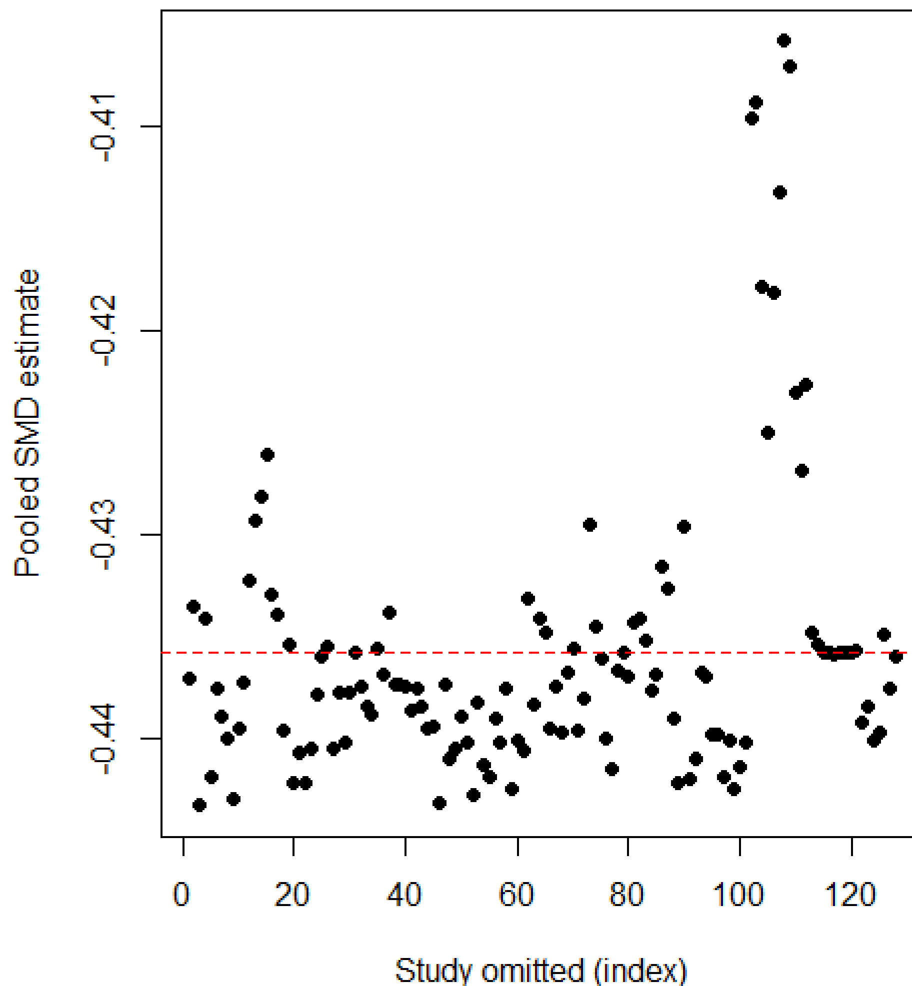

***Figure 5*: Leave-one-out sensitivity analysis**

Overall, the main conclusion appears reasonably robust: the pooled effect remains beneficial across all sensitivity analyses. However, the magnitude of the effect and the amount of unexplained heterogeneity are influenced substantially by effects from Soleimani-Rad (2025), and statistical significance becomes borderline when Curran (2009) is omitted.

How Secure is the Evidence?

Violence outcomes

Our confidence in the findings on the prevention and reduction of violence is **Moderate.** The evidence for violence achieved an initial security rating of **Level 3.** This is based on nine impact evaluations assessing the impact of Clinical CBT treatment on children and young people

The studies included:

- Eight RCTs (one high quality Type A study; six moderate quality Type C studies and one low quality Type D)
- One QED (a Moderate quality Type C study)

Heterogeneity was high (I^2 = 81.5%) and moderator analyses were employed to investigate differences.

Crime and offending outcomes

Our confidence in the findings on crime and offending is **Low**. The meta-analysis included 35 crime and offending related outcomes drawn from six studies that assessed the impact of Clinical CBT treatment on children and young people.

The studies included:

- Five RCT's, of these, two were rated as high quality (Type A); two were rated as moderate quality (Type C) and one as low quality (Type D)
- One QED rated as moderate quality (Type C)

Heterogeneity was moderate (I^2 = 32.1%), meaning the **Level 2** evidence security rating was maintained.

Overall Clinical CBT Treatment

Our confidence in the overall findings for Clinical CBT treatment is **Moderate.** The full effectiveness analysis achieved an overall security rating of **Level 3.** The combined evidence base consists of 14 impact evaluations, including 13 RCTs and

one QED. Of these studies, two were rated as high quality (Type A), 10 as moderate quality (Type C), and one as low quality (Type D).

Heterogeneity was high (I^2 = 80%). Across the moderator analyses, residual heterogeneity remained significant in all models, indicating that the included moderators did not fully explain between-study variation in effect sizes. Only the model including outcome domain and data source yielded a significant omnibus test of moderators, suggesting that these variables may explain some heterogeneity; however, substantial unexplained heterogeneity remained.

The results were generally robust across sensitivity analyses and after applying robust variance estimation to account for dependent effect sizes. Although the magnitude of the pooled effect was influenced by one notable study, the overall pattern of findings remained consistent and in the favourable direction. These findings therefore support continued confidence in the statistically significant beneficial effect of Clinical CBT treatment programmes, and the **Level 3** evidence security rating was retained.

Who does it work for?

This narrative summary of equity-related outcomes supplements the meta-analysis by providing additional insights into how Clinical CBT programmes may affect different demographic groups.

Two studies provided detail in relation to personal characteristics that help to explain for whom Clinical CBT programmes work (Helander et al., 2023; Hoogsteder et al., 2018). The studies covered gender, ethnicity, neurodiversity, SEND and age. Studies took place in Sweden (Helander et al., 2023) and the Netherlands (Hoogsteder et al., 2018). Using the YEF-EQA tool, one study was rated as moderate quality (Hoogsteder et al., 2018) and one as low quality (Helander et al., 2023).

Studies where personal characteristics of the sample were described (e.g., gender) but not specifically related to outcomes of interest or do not contribute to the understanding of who Clinical CBT programmes work for, have not been included in this section. No studies explored socioeconomic status, education, care-experience, intersectionality or place of residence.

Gender

One low quality study explored whether gender moderated the effects of a Clinical CBT-based intervention for children with oppositional defiant disorder (ODD) (Helander et al., 2023). The study examined the two-year follow-up outcomes of children taking part in the treatment-as-usual (TAU) group Parent Management Training (PMT) compared with PMT combined with the CBT-based Coping Power Program (CPP). Gender representation in the sample was uneven, with girls comprising approximately one quarter of participants at baseline and an even smaller proportion were retained in the TAU group at follow-up (5 relative to 26 in the CBT condition). Gender moderated outcomes on one measure of social skills (prosocial competence; SSRS), with girls in the CBT condition demonstrating significantly greater improvement compared with girls receiving TAU. No other gender moderation effects were observed across measures of disruptive behaviour or additional social skills outcomes. These findings suggest that the addition of a CBT-based group component may provide specific social

skills benefits for girls. However, the limited representation and retention of girls within the study restricts the strength of conclusions that can be drawn and highlights an equity consideration for behavioural intervention research.

Ethnicity

One moderate quality study explored whether ethnicity moderated the effectiveness of a Clinical CBT programme in a juvenile justice setting (Hoogsteder et al., 2018). The study evaluated Responsive Aggression Regulation Therapy (Re-ART) compared with treatment as usual for adolescents in a Dutch juvenile justice institution, examining outcomes related to violence and general recidivism. Participants were categorised as 'native' and 'non-native', with the Centraal Bureau voor Statistiek (Statistics Netherlands) defining an individual as 'non-native' when at least one parent was born in a country other than the Netherlands. Analyses found no evidence that ethnicity moderated treatment effects indicating that the intervention had comparable effects on recidivism outcomes for native and non-native young people. However, the categorisation of participants into broad 'native' and 'non-native' groups limits the ability to understand how the intervention may perform across more diverse ethnic or cultural backgrounds. As such, while the evidence indicates no differential impact by ethnicity in this context, the findings should be interpreted cautiously given the limited granularity of the ethnicity measure and the relatively small sample size.

Neurodiversity

One low quality study explored how neurodiversity-related characteristics influenced outcomes in a Clinical CBT-based intervention for children with ODD (Helander et al., 2023). Many participants (66.1%) met diagnostic criteria for attention-deficit/hyperactivity disorder (ADHD) at baseline, reflecting the high prevalence of co-occurring neurodevelopmental conditions within the sample. However, ADHD diagnosis did not moderate intervention effects across outcomes measuring disruptive behaviour (DBD-ODD), social communication (P-COMP), social skills (SRSS), or prosocial behaviour (SDQ). Similarly, prescribed medication status at baseline was not associated with differential treatment effects across these outcomes.

Clinician rated ODD severity showed partial moderation effects when categorised into high versus low/moderate severity groups. Specifically, a significant three-way interaction between time, treatment condition, and baseline ODD severity indicated that treatment effects over time differed for children with high baseline ODD symptoms, both for ODD outcome trajectories and for selected social skills measures (P-COMP and SDQ-prosocial). This suggests that children with more severe initial presentations may respond differently to treatment across follow-up. . However, these effects did not hold when severity was analysed as a continuous variable, underscoring the complexity of behavioural symptoms beyond simple categorical thresholds. Overall, the findings indicate limited evidence that CBT intervention (CPP) produces differential effects based on ADHD diagnosis or medication status, though variation in symptom severity may influence responses to treatment.

SEND

One moderate quality study examined whether special educational needs and disabilities moderated intervention effects of a Clinical CBT programme when compared to treatment as usual (Hoogsteder et al., 2018). Specifically, the authors assessed whether the effectiveness of Re-ART on recidivism differed for adolescents with and without a mild intellectual disability (MID), defined as an IQ score between 70 and 85 combined with reduced adaptive functioning. Approximately one quarter of participants were identified as having MID. Analyses found no evidence that the presence of MID moderated treatment effects on either violence or general recidivism, indicating that the intervention produced comparable outcomes for adolescents with and without mild intellectual disabilities. These findings suggest that the programme may be accessible to young people with mild cognitive impairments within this setting. However, MID was identified through institutional records and broad diagnostic criteria, and the sample of young people with MID was relatively small. As such, while the evidence indicates similar effects across groups, further research is needed to understand how interventions can best support young people with a wider range of SEND profiles.

Age

One moderate quality study examined whether the effects of Re-ART differed by age (Hoogsteder et al., 2018). Participants were grouped into younger (15–17 years) and older (18–21 years) groups to assess whether age moderated outcomes.[8] Analyses found no evidence that age influenced the effectiveness of the intervention on either violent or general recidivism. The results indicated that treatment outcomes were comparable across the two age groups, suggesting that the programme produced similar reductions in recidivism for younger and older participants. These findings suggest that Re-ART may be responsive to a range of developmental stages within adolescence and early adulthood. However, age was examined using broad categorical groupings and within a relatively small institutional sample.

What factors affect implementation?

Ten studies provided evidence related to implementation (Barrett et al., 2001; Ducharme et al., 2021; Erickson, 2013; Esposito-Smythers et al., 2011; Goldston et al., 2021; Hoogsteder et al., 2018; Kolko et al., 2006; Soleimani-Rad et al., 2025; Townsend et al., 2024; van de Wiel et al., 2007). Using the YEF-EQA, two of these papers were classified as high quality (Esposito-Smythers et al., 2011; Soleimani-Rad et al., 2025), seven as moderate quality (Barrett et al., 2001; Ducharme et al., 2021; Erickson, 2013; Hoogsteder et al., 2018; Kolko et al., 2006; Townsend et al., 2024; van de Wiel et al., 2007) and one as low quality (Goldston et al., 2021). Individual study details are available in Appendix 4.

Factors that influenced the implementation of Clinical CBT programmes are organised using Proctor et al.'s (2011) Implementation Outcomes Framework. Appendix 5 highlights the availability of evidence according to each of Proctor's implementation outcomes. Some features of Proctor's framework were not evident in studies included in this review (adoption, reach and penetration, and sustainability).

[8] Although this study included some participants over 18, it remains eligible for inclusion in this review because a majority of the sample was under 18.

Acceptability

Acceptability explores aspects of the intervention or change that children and young people find agreeable or satisfactory. One high quality study (Esposito-Smythers et al., 2011), one moderate quality study (Ducharme et al., 2021) and one low quality study (Goldston et al., 2021) examined the acceptability of Clinical CBT programmes. Across the three studies, acceptability was assessed through a combination of participant feedback, engagement indicators, and treatment completion rates, providing insight into how young people and families experienced these CBT programmes.

Goldston et al. (2021) assessed the acceptability of a Cognitive Behavioural Therapy–Relapse Prevention (CBT-RP) programme for adolescents with depression, suicidality, and substance use difficulties. Acceptability was evaluated using parent and youth ratings of the intervention, alongside opportunities to provide qualitative feedback. Among the six parents who completed ratings, four reported that the intervention helped reduce their child's substance use, and five indicated that it contributed to reductions in depression and suicidal thoughts. Five parents also stated that they would recommend the programme to others. One parent did not endorse this item in the survey despite providing positive comments, however the study authors suggested may have reflected misunderstanding of the response scale rather than dissatisfaction with the intervention. Young people's responses were broadly positive but with greater variation than the parents. Five of the six adolescents reported that the intervention helped with depression and suicidal thoughts, and three reported improvements in substance use; however, only half indicated that they would recommend the programme to others. Qualitative feedback from both parents and youth highlighted the non-judgmental approach of therapists and the opportunity for adolescents to learn coping skills, while parents also valued the regular support and accessibility of therapists. Although these findings suggest generally positive perceptions of the programme, the very small sample limits the strength of conclusions about overall acceptability.

Ducharme et al., (2021) examined the acceptability of incorporating the RAGE-Control biofeedback video game into Anger Control Therapy (ACT) for youth with

clinically significant anger dysregulation. Participant engagement with the intervention was evidenced by lower attrition in the active intervention group, where only one individual withdrew compared to five in the sham gameplay condition. Although this difference did not reach statistical significance, the pattern suggests that the biofeedback component of the intervention may have supported stronger engagement, as reflected in the lower attrition from the intervention group. The authors proposed that the use of a video game format may increase willingness among young people to participate in therapy, particularly for those who might otherwise find traditional treatment approaches less appealing. The intervention also incorporated brief parent check-ins at the end of each session to maintain transparency about the child's progress and ensured parents had a clear understanding of what was being practiced in therapy. In later sessions, children and parents played the game together, with children demonstrating the coping skills they had learned and practiced. This joint activity not only reinforced skill-use but also created a more collaborative therapeutic environment. The authors interpreted these transparency-building and involvement-focused features as contributing to the intervention's high acceptability as an adjunct to traditional CBT.

Similarly, Esposito-Smythers et al., (2011) evaluated the acceptability of an Integrated Cognitive Behavioural Therapy (I-CBT) programme for adolescents with co-occurring suicidality and alcohol or other drug use disorders. Acceptability was reflected primarily through treatment engagement and completion rates over the course of the intensive 12-month intervention. Approximately 74% of adolescents and families in the I-CBT group were classified as treatment completers, meaning they attended at least 24 adolescent sessions and 12 parent sessions. Parent participation was particularly high, with 90% meeting the completion threshold. In addition, adolescents receiving I-CBT attended significantly more outpatient treatment sessions over the 12-month period than those receiving enhanced treatment as usual, suggesting greater engagement with the intervention. These findings indicate that despite the demanding structure of the programme, adolescents and families were willing to participate in and continue with treatment.

Appropriateness

Appropriateness refers to the perceived fit or relevance of an intervention to the given context or problem. It can include discussion of adaptations that are made to improve the intervention's fit with the context and the perceived usefulness of the intervention. One high quality study (Esposito-Smythers et al., 2011) and two moderate quality studies (Townsend et al., 2024; van de Wiel et al., 2007) examined the appropriateness of Clinical CBT programmes. Across the three studies, each intervention was adapted from an established cognitive behavioural framework to better align with the developmental, behavioural, and contextual needs of the target populations. In simple terms, rather than implementing standard protocols, the studies demonstrate how evidence-based approaches were tailored to meet the specific needs of participants or adjusted flexibly to accommodate routine service delivery constraints.

In the Integrated Cognitive Behavioural Therapy (I-CBT) programme, the intervention was designed for adolescents presenting with co-occurring suicidality and alcohol and other drug (AOD) use (Esposito-Smythers et al., 2011). The structure of the intervention reflected the recognition that these difficulties frequently share underlying mechanisms such as cognitive distortions, poor coping skills, and ineffective communication. To increase relevance for this high-risk group, the programme incorporated individual adolescent sessions focused on cognitive restructuring, affect regulation, and refusal skills, alongside family sessions and parent training aimed at improving monitoring and behavioural contracting. The inclusion of motivational interviewing and case management[9] further expanded the scope of the intervention to address contextual stressors such as school or social service involvement. These elements indicate an attempt to ensure that treatment could respond to both the psychological and environmental factors influencing adolescents' risk behaviours, while integrated

[9] Motivational interviewing is a collaborative, person-centred counselling approach that helps individuals explore and strengthen their own motivation for change. Case management is a coordinated support process in which a practitioner assesses needs, connects individuals to appropriate services, and provides ongoing monitoring and guidance.

safety procedures ensured that the programme remained clinically appropriate for individuals experiencing acute suicidality.

In van de Wiel et al. (2007) adaptations to the Utrecht Coping Power Program (UCPP) were primarily directed toward improving the programme's suitability for children with disruptive behaviour disorders receiving routine outpatient care. The original Coping Power Program had been designed as a school-based preventive intervention, and therefore several structural changes were made to reflect the needs of a more clinically complex population and the practical constraints of clinical services. Sessions were made more activity-based and less discussion-focused to accommodate children's developmental capacities and attention span. The overall duration of treatment was also shortened, reducing both child and parent sessions to align the intervention with the time limitations typical of outpatient settings. In addition, parent and child components were fully integrated and delivered by the same therapists. This approach aimed to strengthen continuity of care, facilitate communication between therapists and parents, and enable parents to reinforce treatment goals at home through structured feedback and selected video material from child sessions.

Townsend et al. (2024) similarly adapted cognitive-behavioural therapy for anxiety to improve its relevance for adolescents with autism spectrum disorder (ASD) and co-occurring anxiety. The adapted intervention utilised the Behavioural Interventions for Anxiety in Children with Autism (BIACA) model, which preserved core CBT elements such as exposure and cognitive reappraisal while incorporating additional strategies designed to address the developmental and behavioural characteristics associated with ASD. The treatment included extended sessions with equal emphasis on child- and parent-focused components, recognising the role of parents in supporting exposure tasks. A modular structure guided by a treatment plan allowed therapists to adjust the content of sessions according to the child's needs, including the integration of behavioural management strategies or social skills training where necessary. Engagement was further supported by the child's special interests and a reward system implemented across home and school environments.

Across the studies, parent or family involvement played a notable role in enhancing the relevance of the interventions. In I-CBT, parents were trained in

monitoring and contingency management to support adolescents' behavioural change. The UCPP incorporated parent involvement through integrated sessions and feedback mechanisms designed to help them reinforce coping and anger management strategies at home. Likewise, the BIACA model allocated substantial session time to parent participation to facilitate exposure practice and reduce accommodation of anxiety.

Flexibility in treatment delivery was also evident in the design of these programmes. Both I-CBT and BIACA used modular or menu-based approaches that enabled clinicians to tailor intervention components according to individual needs. This structure allowed therapists to emphasise skills, repeat key elements, or incorporate additional supports when necessary. While the UCPP was less explicitly modular, its adaptations such as activity-based sessions and integrated parent–child delivery similarly reflected attempts to align the programme with the developmental and practical realities of the clinical population.

Despite these efforts to enhance contextual fit, several considerations remain. In each study the rationale for adaptations was largely based on theoretical or clinical assumptions about the needs of the target population rather than direct evaluation of participants' perceptions of relevance. Appropriateness was therefore inferred primarily through intervention design and treatment outcomes rather than through explicit measurement of the experiences or perspectives of children, young people and their families. In addition, greater flexibility and individual tailoring may introduce variability in how interventions are delivered which can complicate efforts to maintain fidelity to core treatment components. For example, reductions in session length and duration in the UCPP were intended to improve feasibility within routine services, but such modifications may also influence the overall intensity of treatment, and this potential impact was not examined in the studies.

Taken together, the three studies illustrate how Clinical CBT programmes can be modified to better align with the developmental characteristics, clinical complexity, and environmental contexts of different populations of children and young people. While the specific adaptations differed across studies, each intervention reflects an effort to maintain core therapeutic mechanisms while

adjusting delivery methods, content, and structure to increase relevance for the intended population and service setting.

Adoption

Adoption concerns the decision or action to employ an intervention or implementation target. It also refers to the uptake of an intervention provided by services and communities. No studies examined the adoption of Clinical CBT programmes.

Feasibility

Feasibility refers to the extent to which an intervention can be successfully implemented within a specific setting, with emphasis on its practicality and the ability to deliver it effectively in the target environment. This includes evidence of practicality or utility, and other indicators demonstrating that the intervention can be carried out as intended in real-world conditions. One moderate quality study (Ducharme et al., 2021) and one low quality study (Goldston et al., 2021) explored the feasibility of Clinical CBT programmes. Across the two studies reviewed, feasibility was evaluated through implementation indicators (e.g., participation and retention) and the ability of participants to successfully engage with novel treatment components within an existing CBT intervention.

Goldston et al. (2021) assessed the feasibility of Cognitive Behavioral Therapy–Relapse Prevention (CBT-RP) when delivered alongside enhanced treatment as usual for adolescents presenting with co-occurring depression, suicidality, and substance use difficulties. Feasibility was primarily examined through recruitment and retention outcomes in this pilot randomised trial. Here, retention refers to the extent to which participants remain enrolled and continue taking part in a study or intervention over time. All adolescents and families who met eligibility criteria agreed to participate, suggesting a high level of willingness among the study population to engage with the intervention. Retention throughout the study period was also relatively high. Of the 13 participants randomised to the treatment condition, only one individual in the CBT-RP group withdrew from the study before the midpoint of the 20-week intervention due to court-mandated therapeutic services unrelated to the study. The remaining participants continued receiving

treatment throughout the study period resulting in a retention rate of around 86%. These findings suggest that adolescents and their families were generally able to engage with and remain involved in the integrated CBT-RP intervention, supporting its practical implementation within this high-risk population. However, the small sample size and pilot nature of the trial restrict conclusions about feasibility beyond the immediate study context.

Ducharme et al. (2021) explored the feasibility of integrating RAGE-Control, a biofeedback-based video game designed to reinforce physiological regulation skills, into Anger Control Therapy for youth with clinically significant anger dysregulation. Participants in the active intervention demonstrated the ability to use the biofeedback mechanism as intended, successfully reducing their median heart rate during gameplay across the ten treatment sessions. These results demonstrate that the biofeedback could be used to practice regulation strategies within the therapeutic context. Physiological changes were also associated with decreases in aggression and oppositional behaviours, suggesting potential transfer of these skills beyond the game environment to real-world functioning. As a small proof-of-concept trial (n = 40), the study provides preliminary evidence that biofeedback-enhanced components can be feasibly embedded within CBT, though further research with larger samples is required to confirm feasibility in routine clinical settings.

Fidelity

Fidelity refers to the degree to which an intervention was delivered as intended. Eight studies explored the fidelity of implementing Clinical CBT programmes (Barrett et al., 2001; Ducharme et al., 2021; Erickson, 2013; Esposito-Smythers et al., 2011; Hoogsteder et al., 2018; Kolko et al., 2006; Soleimani-Rad et al., 2025; Townsend et al., 2024). This involved two high quality studies (Esposito-Smythers et al., 2011; Soleimani-Rad et al., 2025) and six moderate quality studies (Barrett et al., 2001; Ducharme et al., 2021; Erickson, 2013; Hoogsteder et al., 2018; Kolko et al., 2006; Townsend et al., 2024). Across the studies, fidelity was generally supported through structured implementation procedures such as manualised protocols, supervision, and systematic monitoring of session delivery. Although the specific

strategies varied, several consistent approaches can be observed in how fidelity was maintained and evaluated.

A common feature across many of the studies was the use of manualised treatment protocols and structured training procedures to guide intervention delivery. In Esposito-Smythers et al. (2011) therapists delivering the Integrated Cognitive Behavioural Therapy (I-CBT) intervention received structured training involving didactic instruction and role-plays, and were required to demonstrate competence before delivering the intervention. Similarly, Ducharme et al. (2021) implemented Anger Control Therapy (ACT) using a detailed manual adapted from an established anger control training programme, with therapists receiving initial training and ongoing weekly supervision. Barrett et al. (2001) also relied on standardised treatment manuals and session checklists across four treatment conditions, while Soleimani-Rad et al. (2025) delivered CBT through a structured 15-session protocol, ensuring consistent treatment exposure for participants. These approaches highlight the reliance on clearly defined intervention structures to guide therapists and reduce variability in treatment delivery.

Supervision and ongoing oversight also played a key role in supporting fidelity across studies. In Esposito-Smythers et al. (2011) therapists participated in weekly group supervision in which clinical worksheets were reviewed, and feedback was provided based on session recordings. Therapist competence was assessed using the Cognitive Therapy Rating Scale (CTRS), with scores of 44 or above indicating acceptable delivery. In this study, 92% of rated sessions met this standard and 85% met the adherence threshold of at least 80% of session components, indicating strong implementation of the intervention model. The authors noted that if therapist drift (i.e., the gradual, unintentional deviation from an intervention's prescribed methods or protocols) was detected during the review process, the team addressed the obstacles and conducted additional role-plays to enhance training and ensure the intervention remained on track. Ducharme et al. (2021) similarly incorporated weekly supervision throughout the intervention period, while Barrett et al. (2001) used supervision sessions to review videotaped therapy sessions and adherence checklists. In the Re-ART programme evaluated by Hoogsteder et al. (2018) therapists participated in biweekly peer supervision sessions to maintain consistency in programme

delivery, alongside requirements that therapists meet minimum educational standards and maintain an active caseload of at least three clients.

Many of the studies also used recordings and independent ratings to formally evaluate fidelity. Audio- or video-recorded sessions allowed researchers to examine whether key therapeutic components were delivered as intended. For instance, Kolko et al. (2006) reviewed approximately 30% of recorded CBT sessions, reporting a high mean fidelity rating of 92% (range = 85–99%) relative to the treatment manual. Ducharme et al., (2021) used fidelity checklists to evaluate session content, with an independent rater reviewing 20% of recorded sessions, resulting in a mean fidelity score of 93% (range = 81–100%). Similarly, Townsend et al. (2024) monitored adherence to CBT adapted for autism (BIACA) through coding of 92 recorded sessions, finding that therapists adhered to 97% of required treatment components, with strong interrater reliability (ICC = 0.85). These findings suggest that structured monitoring procedures can help ensure that therapists consistently deliver core intervention components.

Several studies also considered treatment exposure and completion rates as indicators of fidelity. In the CBT intervention examined by Soleimani-Rad et al. (2025) 85% of participants (23 of 27) completed all 15 therapy sessions, and missed sessions were replaced with compensatory sessions delivered within the same week to ensure full exposure to treatment content. Similarly, Barrett et al. (2001) reported high adherence across multiple treatment modalities, with mean adherence ratings ranging from 8.9 to 9.5 on a 10-point scale, and no significant differences in adherence across treatment conditions. These findings indicate that structured session planning and adherence monitoring helped ensure that participants received the intended dosage of treatment.

Although most studies reported high fidelity, some implementation challenges were also noted. Erickson (2013), who evaluated Aggression Replacement Training® (ART®) in a residential juvenile justice setting, used multiple monitoring procedures including session observations, videotaped recordings, and evaluation forms completed by trainers and supervisors. Notably, only one curriculum trainer delivered the intervention. While this enhanced consistency, it also increased the risk of delivery drift (i.e. gradual deviation from the intervention's intended structure or delivery). Although the trainer had completed

standardised ART® training, she was unable to attend a planned booster training session during the intervention period. The master trainer subsequently assigned a composite fidelity score of 1.8 on a 0–3 scale ("nearly competent"), with written evaluations rated only as "satisfactory," suggesting that protocol adherence was incomplete and some degree of delivery contamination may have occurred. Similarly, Hoogsteder et al. (2018) reported that while Re-ART was largely delivered according to protocol, 23.6% of staff respondents identified weaknesses in communication between therapists and residential mentors, and 16.3% reported inconsistent reinforcement of treatment goals within the living environment. Although the study deviated from formal Re-ART indication criteria for 12 adolescents (seven aged 15 and five with IQ scores between 68 and 69), these decisions were made on clinical grounds, with the determination that participants possessed sufficient cognitive capacity to engage in the intervention. These findings highlight how organisational coordination and staff collaboration can influence the consistency with which interventions are implemented.

Reach and penetration

Reach and penetration refer to the extent to which the intervention has been integrated or reached eligible recipients. No studies examined the reach and penetration of Clinical CBT programmes.

Sustainability

Sustainability refers to the ability of an implementer to maintain or institutionalise the intervention over time. No studies examined the sustainability of Clinical CBT programmes.

Experiences of Children and Young People

One low quality study examined the experiences of adolescents participating in a Cognitive-Behavioural Therapy – Relapse Prevention (CBT-RP) programme, designed for youth with depression, suicidal behaviours, and substance use problems (Goldston et al., 2021). Thirteen participants were enrolled in the study, with six completing the intervention and providing feedback on their experiences.

Participants reported that the programme helped them manage depression and suicidal thoughts, with five out of six indicating improvements. Three participants also noted reductions in alcohol and drug use. Adolescents highlighted the "open minds" of the therapists and the "optimistic advice" they received as particularly valuable. They described the opportunity to talk about problems, learn coping strategies for stress, engage in problem-solving, and take part in drug testing as helpful aspects of the intervention. When asked what they learned from the programme, adolescents most frequently mentioned coping skills and strategies for managing emotions. They also described learning about the negative effects of substance use. While experiences were largely positive, only three of six participants said they would recommend the programme to others, and two reported mixed feelings. This variation suggests that while the programme was beneficial for many, individual experiences and engagement differed.

These findings offer useful insights, more research is needed to explore children and young people's experiences of Clinical CBT programmes across diverse populations, contexts, and therapeutic approaches to better understand which elements of treatment are most helpful from their perspective.

How much does it cost?

Across the studies implementing Clinical CBT programmes for children and young people, none provided data on the financial or resource costs associated with delivering the interventions. Most studies reported implementation details such as session numbers, therapist training, fidelity monitoring, and, in some cases, adaptations to enhance engagement or appropriateness of the intervention for specific participant groups. However, none provided information on staff time, supervision costs, materials, or other programmatic expenses associated with the delivery of these programmes. As a result, the economic feasibility and cost-effectiveness of these Clinical CBT programmes remain unclear, limiting the ability of practitioners and policy makers to evaluate the resources required for implementation.

Conclusion and Takeaway Messages

Evidence from this systematic review and meta-analysis indicates that Clinical CBT is a promising, targeted approach for reducing violence, crime and offending among high-risk children and young people. Using robust techniques to pool the evidence base, the direction of effect was consistently beneficial to young people who received the intervention. However, heterogeneity was substantial and some estimates, particularly for violence and crime/offending after robust adjustment, remained statistically uncertain, so the findings should be interpreted with appropriate caution.

Violence

The evidence suggests a possible 69% relative risk reduction in violence from Clinical CBT interventions, equivalent to an absolute risk reduction of 11 percentage points meaning around 5% of children and young people receiving Clinical CBT went on to be involved in violence compared with 16% of those in comparison conditions. These findings are based on a meta-analysis of 28 violence outcomes across nine studies. Although the pooled effect was large and in the beneficial direction, it became non-significant after adjusting for clustering across studies using robust variance estimation ($p = .139$), and the overall evidence security rating is Level 3.

These findings are also supported by wider evidence on CBT for aggression and externalising behaviour. For example, Riise et al. (2021) reported large effects for CBT delivered in routine clinical care for children and young people with externalising disorders, including Attention Deficit Hyperactivity Disorder (ADHD), Conduct Disorder (CD), and Oppositional Defiant Disorder (ODD). They found large and significant within-group effect sizes of $g = 0.91$ at post-treatment, and when broken down by type of externalising disorder, the effect sizes remained large ($g = 0.80$ for ADHD and $g = 0.98$ for CD and ODD). Similarly, de Ribera et al. (2024) conducted a systematic review of the Reasoning & Rehabilitation (R&R) CBT programme and found that it was effective in reducing violence/aggression ($g = 0.38$), anger/hostility ($g = 0.25$), and impulsivity/inhibition ($g = 0.27$). Direct comparison remains limited, however, because these reviews covered broader

behavioural outcomes, wider age ranges, and programme variants, rather than the narrower and more specific clinical strand examined here.

Crime and Offending

The estimate points to an 88% relative risk reduction in crime and offending outcomes resulting from Clinical CBT (corresponding to an absolute risk reduction of 22 percentage points), but confidence is limited. Meaning around 3% of participants receiving Clinical CBT went on to be involved in crime and offending compared with 25% of those receiving business as usual. These findings are based on 35 outcomes across six studies. After adjusting for clustering across studies, this high-impact estimate was borderline statistically significant ($p = .0502$), and the overall evidence security rating is Level 2.

These findings are consistent with two earlier reviews of CBT delivered to young people at a high risk of reoffending. Lipsey et al. (2007) reviewed the effects of CBT programmes for criminal offenders and found a 25% overall reduction in recidivism. Importantly, the meta-analysis examined whether treatment effects differed by age group. Around 29% of the included studies focused on young people in justice settings, while 71% focused on adult offenders. The authors found no relationship between treatment effect size and age group, and concluded that CBT appears to be a useful intervention in both juvenile and adult criminal justice settings. Similarly, Koehler et al. (2012) focused specifically on rehabilitation programmes for young offenders in Europe. Their meta-analysis found that cognitive-behavioural treatments performed above average compared with other interventions, resulting in a 27% relative risk reduction in recidivism compared with control groups. It is possible that our observed effects are higher than those found in these previous reviews. One potential explanation is that those earlier reviews focused mainly on CBT delivered in criminal justice settings to participants explicitly identified as offenders, whereas our review considered a narrower clinical CBT strand within a broader high-risk population.

All outcomes

Beyond violence, crime, and offending outcomes, we conducted a meta-analysis across all 128 identified YEF outcomes (14 effectiveness studies, 1,042 children and

young people). Clinical CBT demonstrated a statistically significant overall beneficial effect across this broad spectrum of outcomes (SMD = -0.48, $p < 0.05$), securing a Level 3 evidence security rating.

When looking at specific outcome domains, Clinical CBT had statistically significant beneficial effects on young people's ability to resolve conflict and improving family relationships and support. These positive findings are strongly supported by existing literature. For instance, de Ribera et al. (2024) found that CBT programmes are effective at significantly increasing social problem-solving skills, empathy, and social-perspective taking. Lipsey et al. (2007) similarly concluded that CBT interventions specifically incorporating interpersonal problem-solving and anger control components are associated with larger, more successful treatment effects.

No statistically significant effects were observed for general mental health, behavioural difficulties, regulating emotions, or drug and alcohol use. This finding highlights a data gap noted in other similar reviews where secondary outcomes like general mental health and substance use are historically under-measured or yield inconclusive data in CBT offending evaluations; de Ribera et al. (2024) reported that there were simply not enough studies or data available to conduct meta-analyses on secondary outcomes like drug/alcohol consumption, job attainment and life satisfaction. Riise et al. (2021) echoed this, highlighting that a major limitation in the field is how rarely CBT effectiveness studies report on broad functioning, quality of life, or comorbid mental health conditions.

What Works?

Evidence from ten studies suggests that Clinical CBT programmes are most successfully implemented when they are acceptable to participants, appropriately adapted to clinical and developmental needs, feasible within service constraints, and delivered with high fidelity. Manualised protocols, structured therapist training and supervision, family involvement, flexible programme design, and systematic monitoring of adherence all appeared to support stronger engagement, retention and consistency of delivery This interpretation is supported by wider evidence, with Lipsey et al. (2007) demonstrating that programme effects are stronger when interventions are well

implemented, particularly when supported by high levels of provider training, implementation quality, and fidelity to the treatment model.Riise et al. (2021) also found that CBT retained strong effects when delivered in routine clinical care, suggesting that benefits can be sustained beyond tightly controlled research settings when implementation is robust.

Acceptability appears to be strengthened where programmes emphasise supportive, non-judgmental therapeutic relationships and give children, young people and families clear opportunities to build and practise coping skills. Family involvement and regular communication with trusted adults were also important, helping to reinforce treatment goals and support skill use beyond the therapy setting.

Appropriateness was enhanced where interventions were adapted to respond to co-occurring difficulties, developmental level and the realities of clinical practice. Across the included studies, programmes were often modified through modular content, activity-based delivery, integrated parent-child components, or the use of motivating features such as rewards, games, and young people's special interests. These adaptations appear important for maintaining relevance without abandoning core CBT mechanisms. Feasibility also appeared closely linked to contextual fit, with pilot studies suggesting that structured Clinical CBT can be delivered within clinical settings and that young people can engage with newer components such as biofeedback, although evidence on sustainability and scale-up remains limited.

High fidelity was another consistent feature of successful implementation. Clear manuals, structured training, regular supervision, session recordings, and fidelity checklists helped maintain treatment integrity. At the same time, implementation was influenced by wider organisational factors, including communication between professionals, staffing capacity, and how well treatment goals were reinforced across the service environment.

However, most implementation evidence came from US and European healthcare or justice settings rather than UK services, and no studies reported detailed economic evaluations or resource costs. This limits confidence in how far the

findings transfer to routine UK practice and means firm conclusions about affordability, cost-effectiveness and scalability cannot yet be drawn.

Who Benefits Most?

There was limited but generally encouraging evidence that Clinical CBT can be delivered equitably across diverse groups of children and young people. Broadly comparable outcomes were reported across ethnicity and between children under 13 and older adolescents (13+), though conclusions across developmental stages should be interpreted cautiously given the broad age range and simplified age grouping. Additionally, confidence in these findings is limited by small subgroup samples and the limited representation of underrepresented groups.

Gender

Clinical CBT interventions did not show statistically significant effects for mixed-sex or boys-only samples. While a small effect size indicating worse outcomes was observed for girls-only samples, this estimate relies on a single study and should be interpreted with extreme caution. Narrative evidence was similarly limited.

This inability to detect statistically significant gender differences, and the broader lack of female-specific outcomes, is a pervasive limitation throughout the existing literature. Previous systematic reviews that have attempted to measure gender differences have consistently failed to find significant effects due to the severe under-representation of girls in CBT evaluations. For instance, Lipsey et al. (2007) noted that most of the participants in their included studies were either all male or predominately male, resulting in too little variation in their moderator analysis to understand how well the intervention works specifically for girls. Similarly, de Ribera et al. (2024) initially outlined at protocol to conduct subgroup analyses comparing male, female, and mixed-sex samples, but had to drop this analysis entirely because "there was not much variability across the studies". Ultimately, the inability to draw firm conclusions about gender differences mirrors structural data gaps identified across the field, highlighting a critical need for future research focused on female populations.

Ethnicity

Effectiveness did not vary significantly across ethnicity subgroup analyses, including majority white, diverse/balanced, and some diversity samples. One moderate-quality study in a juvenile justice setting also found no evidence that ethnicity, categorised as 'native' and 'non-native', moderated recidivism outcomes. While this suggests broadly comparable effects across ethnic groups, the evidence is heavily limited by small samples and the use of broad ethnicity categories. This inability to draw firm conclusions about ethnic and cultural differences mirrors the data gaps seen with gender and restricts our understanding of how Clinical CBT performs across different ethnic and cultural groups, highlighting a critical need for more granular demographic research.

Age

The effectiveness of Clinical CBT did not differ significantly between children (under 13) and older adolescents (13+), suggesting no clear evidence that effects varied by broad age group. Similarly, one moderate-quality study using a within-study comparison in a juvenile justice setting found no evidence that outcomes differed between younger (15–17) and older (18–21) adolescents which is supported by a similar finding in Lipsey (2007) which showed CBT was equally effective when comparing child with adult populations. However, underreporting of these basic demographic details remains an issue; de Ribera et al. (2024) initially planned to conduct specific subgroup analyses comparing adults, adolescents, and children, but had to abandon the analysis entirely because there was not enough variability or detailed age reporting across the primary studies. This review relied on broad age brackets to allow subgroup analysis, but unfortunately this restricts a more nuanced understanding of exactly how Clinical CBT performs at specific, granular developmental stages.

Limitations

Challenges with the studies measuring effectiveness

This review highlights substantial between-study heterogeneity (I^2 = 80.0%), meaning that effect sizes were inconsistent across studies beyond what would be expected by chance. A key limitation was the presence of an overly influential

outlier. Sensitivity analyses demonstrated that a single study (Soleimani-Rad, 2025) contributed disproportionately to both the effect magnitude and the residual heterogeneity; removing it reduced the pooled overall effect from -0.48 to -0.28. Additionally, the prediction intervals across our primary outcomes were very wide, suggesting that while average effects are strongly beneficial, future studies might still find smaller effects or no clear effect.

Confidence in the findings is further limited by the small number of studies contributing to some outcomes, the fact that multiple outcomes were often drawn from the same studies, and the underreporting of basic demographic information. The evidence base was also largely drawn from non-UK settings and from varied clinical and high-risk populations, which limits certainty about how far the findings generalise to routine UK services or to more specific groups of children and young people.

Challenges with the studies measuring implementation

Ten studies were eligible for inclusion; however, coverage of implementation outcomes was uneven. While evidence was available for acceptability, appropriateness, feasibility, and fidelity, no studies examined adoption, reach and penetration, or sustainability, limiting understanding of how Clinical CBT programmes are implemented at scale and maintained over time.

All included studies were quantitative, primarily randomised controlled trials, with a focus on treatment effectiveness rather than implementation. Consequently, implementation outcomes were often reported only as secondary information within trials. For example, fidelity monitoring was typically included as part of the methodological procedures used to ensure internal validity within RCTs, rather than as a dedicated examination of how interventions were implemented in routine practice. This means that while fidelity procedures were often well documented, other implementation outcomes were rarely explored in depth.

In addition, the way implementation outcomes were measured varied considerably across studies. Outcomes such as acceptability and feasibility were frequently inferred through indirect indicators including treatment attendance, retention rates, or therapist reports, rather than through direct measurement of the experiences or perspectives of children, young people, families, or

practitioners. This reliance on proxy indicators limits the extent to which conclusions can be drawn about how acceptable or practical the interventions were in real-world settings.

There was also limited evidence examining implementation within a UK context. Several studies were conducted in international clinical or institutional settings, including specialised outpatient services which might differ from the structure and resource constraints of UK child and adolescent mental health or youth justice services. These contextual differences mean that findings cannot always be directly transferred to UK service systems, highlighting the need for further implementation research within UK settings.

Finally, although studies frequently described the structure and delivery of Clinical CBT interventions, none reported detailed cost or resource data associated with programme implementation. While treatment protocols, therapist training, and supervision procedures were often outlined, the absence of economic information limits the ability to assess the cost-effectiveness or affordability of implementing Clinical CBT programmes in routine practice.

Final Thoughts and Recommendations

Overall, this review suggests that Clinical CBT is a promising targeted approach for high-risk children and young people, with the strongest evidence relating to violence, crime, and offending outcomes. The direction of effect was consistently beneficial, and there were also encouraging findings for conflict resolution and family relationships. However, the evidence remains limited by substantial heterogeneity, statistical uncertainty for some outcomes, small samples, and a lack of UK-based research.

The evidence suggests that Clinical CBT is most likely to be successful when delivered with strong implementation support, including manualised protocols, trained therapists, regular supervision, family involvement, and adaptation to developmental and clinical need. While clinical CBT may be considered a potentially useful option for routine practice in this context, the available evidence remains limited and does not yet permit firm conclusions about its effectiveness for these specific populations and settings. Further UK-based research is needed

to test effectiveness, feasibility, cost-effectiveness, and scalability in the UK, and to strengthen understanding of which children and young people benefit most.

Based on our systematic review and meta-analysis, we make the following key recommendations:

- Consider Clinical CBT as a promising targeted intervention for high-risk children and young people, particularly where there are concerns about violence, crime, or offending.
- Strengthen future evaluation in routine UK services. There is a clear need for larger UK-based studies examining not only effectiveness, but also adoption, reach, sustainability, and cost-effectiveness.
- Improve reporting on who benefits most. Future research should report outcomes more consistently by gender, ethnicity, developmental stage, and neurodevelopmental profile so that conclusions can be drawn more confidently about equity and differential impact.
- Measure a broader range of outcomes. In addition to violence, crime, and offending, future studies should capture mental health, substance use, emotional regulation, family functioning, and longer-term life outcomes to provide a fuller picture of impact.
- Adapt Clinical CBT programmes to the developmental, clinical, and contextual needs of the target population while maintaining core therapeutic components, alongside further research and co-design with young people, families, and practitioners to strengthen guidance on how best to achieve this.
- Strengthen engagement by incorporating structured parent/carer involvement and using flexible or motivating intervention features that support participation.
- Ensure that Clinical CBT programmes provide consistent access to trusted adults, as young people emphasised that supportive, open therapeutic relationships were central to their engagement.
- Ensure consistent delivery through clear manuals, training, supervision, and organisational structures that promote communication and reduce implementation barriers.

- As none of the reviewed studies provided economic evaluations or detailed cost data, implementers and researchers should routinely document the resources required for Clinical CBT delivery such as staffing, training, supervision, and materials. Clear reporting of costs is essential for assessing affordability, cost-effectiveness, and the feasibility of wider implementation.

References

Studies included in effectiveness and implementation sections are marked with an asterisk (*) at the beginning of the reference.

* Barrett, H., Slesnick, N., Brody, J. L., Turner, C. W., & Peterson, T. R. (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

* Clifford, P., Gevers, C., Jonkman, K. M., Boer, F., & Begeer, S. (2022). The effectiveness of an attention-based intervention for school-aged autistic children with anger regulating problems: A randomized controlled trial. Autism Research, 15(10), 1971–1984. https://doi.org/10.1002/aur.2800

* Curran, J., Bonn, R., Johnson, B., Grenchanik, C., Moss, K., & Colitti, M. (1977). Nassau County Probation Department Operation Juvenile Intercept Evaluation. Final Report. New York: John Jay College of Criminal Justice.

* Ducharme, P., Kahn, J., Vaudreuil, C., Gusman, M., Waber, D., Ross, A., Rotenberg, A., Rober, A., Kimball, K., Peechatka, A. L., & Gonzalez-Heydrich, J. (2021). A "Proof of Concept" Randomized Controlled Trial of a Video Game Requiring Emotional Regulation to Augment Anger Control Training. Frontiers in Psychiatry, 12, 591906. https://doi.org/10.3389/fpsyt.2021.591906

* Erickson, J. A. (2013). The Efficacy of Aggression Replacement Training with Female Juvenile Offenders in a Residential Commitment Program. https://scholarcommons.usf.edu/etd/4479/

* Esposito-Smythers, C., Spirito, A., Kahler, C. W., Hunt, J., & Monti, P. (2011). Treatment of co-occurring substance abuse and suicidality among adolescents: A randomized trial. Journal of Consulting and Clinical Psychology, 79(6), 728–739. https://doi.org/10.1037/a0026074

* Goldston, D. B., Curry, J. F., Wells, K. C., Kaminer, Y., Daniel, S. S., Esposito-Smythers, C., Doyle, O., Sapyta, J., Tunno, A. M., Heilbron, N., & Roley-Roberts, M. (2021). Feasibility of an Integrated Treatment Approach for Youth with Depression, Suicide Attempts, and Substance Use Problems. Evidence-Based Practice in Child and

Adolescent Mental Health, 6(2), 155–172. https://doi.org/10.1080/23794925.2021.1888664

* Helander, M., Enebrink, P., Hellner, C., & Ahlen, J. (2023). Parent Management Training Combined with Group-CBT Compared to Parent Management Training Only for Oppositional Defiant Disorder Symptoms: 2-Year Follow-Up of a Randomized Controlled Trial. Child Psychiatry & Human Development, 54(4), 1112–1126. https://doi.org/10.1007/s10578-021-01306-3

* Hoogsteder, L. M., Stams, G.-J. J. M., Schippers, E. E., & Bonnes, D. (2018). Responsive Aggression Regulation Therapy (Re-ART): An Evaluation Study in a Dutch Juvenile Justice Institution in Terms of Recidivism. International Journal of Offender Therapy and Comparative Criminology, 62(14), 4403–4424. https://doi.org/10.1177/0306624X18761267

* Koegl, C. J., Farrington, D. P., Augimeri, L. K., & Day, D. M. (2008). Evaluation of a Targeted Cognitive-Behavioral Program for Children with Conduct Problems — The SNAP® Under 12 Outreach Project: Service Intensity, Age and Gender Effects on Short- and Long-Term Outcomes. Clinical Child Psychology and Psychiatry, 13(3), 419–434. https://doi.org/10.1177/1359104508090606

Koehler, J. A., Lösel, F., Akoensi, T. D., & Humphreys, D. K. (2013). A systematic review and meta-analysis on the effects of young offender treatment programs in Europe. Journal of Experimental Criminology, 9(1), 19–43. https://doi.org/10.1007/s11292-012-9159-7

* Kolko, D. J., Herschell, A. D., & Scharf, D. M. (2006). Education and Treatment for Boys Who Set Fires: Specificity, Moderators, and Predictors of Recidivism. Journal of Emotional and Behavioral Disorders, 14(4), 227–239. https://doi.org/10.1177/10634266060140040601

Lipsey, M. W., Landenberger, N. A., & Wilson, S. J. (2007). Effects of Cognitive-Behavioral Programs for Criminal Offenders. Campbell Systematic Reviews, 3(1), 1–27. https://doi.org/10.4073/csr.2007.6

Parker, K., Nunns, M., Xiao, Z., Ford, T., Stallard, P., Kuyken, W., Axford, N., & Ukoumunne, O. C. (2025). Patterns of intra-cluster correlation coefficients in school-based cluster randomised controlled trials of interventions for improving

social-emotional functioning outcomes in pupils: A secondary data analysis of five UK-based studies. BMC Medical Research Methodology, 25(1), 120. https://doi.org/10.1186/s12874-025-02574-6

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with Robust Variance Estimation: Expanding the Range of Working Models. Prevention Science, 23(3), 425–438. https://doi.org/10.1007/s11121-021-01246-3

Raine, A., Cheney, R. A., Ho, R., Portnoy, J., Liu, J., Soyfer, L., Hibbeln, J., & Richmond, T. S. (2016). Nutritional supplementation to reduce child aggression: A randomized, stratified, single-blind, factorial trial. Journal of Child Psychology and Psychiatry, 57(9), 1038–1046. https://doi.org/10.1111/jcpp.12565

Reardon, J. P., & Tosi, D. J. (1977). The effects of rational stage directed imagery on self-concept and reduction of psychological stress in adolescent delinquent females. Journal of Clinical Psychology, 33(4), 1084–1092. https://doi.org/10.1002/1097-4679(197710)33:4%253C1084::AID-JCLP2270330437%253E3.0.CO;2-A

* Rohde, P., Jorgensen, J. S., Seeley, J. R., & Mace, D. E. (2004). Pilot Evaluation of the Coping Course: A Cognitive-Behavioral Intervention to Enhance Coping Skills in Incarcerated Youth. Journal of the American Academy of Child & Adolescent Psychiatry, 43(6), 669–676. https://doi.org/10.1097/01.chi.0000121068.29744.a5

Rutterford, C., Copas, A., & Eldridge, S. (2015). Methods for sample size determination in cluster randomized trials. International Journal of Epidemiology, 44(3), 1051–1067. https://doi.org/10.1093/ije/dyv113

Smeets, K. C., Leeijen, A. A., van der Molen, M. J., Scheepers, F. E., Buitelaar, J. K., & Rommelse, N. N. (2015). Treatment moderators of cognitive behavior therapy to reduce aggressive behavior: a meta-analysis. *European child & adolescent psychiatry*, *24*(3), 255–264. https://doi.org/10.1007/s00787-014-0592-1

Smokowski, P. R., Fraser, M. W., Day, S., Galinsky, M., & Bacallao, M. L. (2004). School-Based Skills Training to Prevent Aggressive Behavior and Peer Rejection in Childhood: Evaluating the Making Choices Program. The Journal of Primary Prevention, 25(2), 233–251. https://doi.org/10.1023/b:Jopp.0000042392.57611.05

* Soleimani-Rad, H., Bahrami, L., Goodarzi, H., & Ariapooran, S. (2025). Cognitive-Behavioral Therapy for Disruptive Mood Dysregulation Disorder: A Randomized Control Trial. Journal of Clinical Psychology, 81(10), 1014–1024. https://doi.org/10.1002/jclp.70015

* Townsend, A. N., Guzick, A. G., Hertz, A. G., Kerns, C. M., Goodman, W. K., Berry, L. N., Kendall, P. C., Wood, J. J., & Storch, E. A. (2024). Anger Outbursts in Youth with ASD and Anxiety: Phenomenology and Relationship with Family Accommodation. Child Psychiatry & Human Development, 55(5), 1259–1268. https://doi.org/10.1007/s10578-022-01489-3

* van de Wiel, N. M. H., Matthys, W., Cohen-Kettenis, P. T., Maassen, G. H., Lochman, J. E., & Van Engeland, H. (2007). The Effectiveness of an Experimental Treatment When Compared to Care as Usual Depends on the Type of Care as Usual. Behavior Modification, 31(3), 298–312. https://doi.org/10.1177/0145445506292855

YEF. (2023). Children, violence and vulnerability: The second annual Youth Endowment Fund report into young people's experiences of violence. Youth Endowment Fund. https://youthendowmentfund.org.uk/wp-content/uploads/2023/11/YEF_Children_violence_and_vulnerability_2023_FINAL.pdf

Youth Endowment Fund. (n.d.). Cognitive Behavioural Therapy. https://youthendowmentfund.org.uk/toolkit/cognitive-behavioural-therapy/

Appendix 1. Methods of the systematic review

Protocol

Prior to initiating this systematic review, we developed a comprehensive protocol for an Evidence and Gap Map (EGM) outlining the research objectives, eligibility criteria, search strategy, data extraction, quality appraisal, and synthesis methods. This protocol was registered and is available on the Open Science Framework (OSF),[10] ensuring transparency and adherence to predefined methods.

Eligibility Criteria

To define the scope of relevant research, we applied the following criteria from the outset:

Table 21: Eligibility criteria for Clinical CBT studies

Eligibility criteria	Inclusion criteria	Exclusion criteria
Population	Children and young people aged 0–17. 'high-risk' population; have formal diagnoses e.g., Conduct Disorder (CD), Oppositional Defiant Disorder (ODD) linked to aggression	Studies with samples predominantly composed of adults aged 18 years and older Where studies include wider age bands, data must be extractable for participants aged 0–17 (or the sample mean age must be ≤17 and the vast majority of participants under 18

[10] Protocol is available to access here: https://osf.io/vamxy

Intervention	Interventions explicitly targeting criminogenic needs, severe behavioural pathology, or psychological drivers of violence (e.g., trauma, impulsivity) using CBT. Interventions must be responsive/tailored to individual needs rather than purely generic curricula	General counselling or psychodynamic therapies not based on CBT Trauma-focused therapies delivered primarily for victims (e.g., PTSD-focused TF-CBT for victimisation recovery), where the theory of change is clinical trauma treatment rather than violence prevention in at-risk/offending populations.
Comparator	Business as usual / treatment as usual (BAU/TAU), no intervention, waitlist, or minimal-contact control In terms of study design, we only include evaluations with a credible counterfactual (i.e. an RCT or a robust quasi-experimental design such as DiD, synthetic control, regression discontinuity, interrupted time series with control, instrumental variables, or matched comparison).	We will exclude evaluations where the comparator is another structured active psychosocial intervention (e.g., another named therapy and/ or curriculum) unless it clearly constitutes BAU within that service context
Dose	High Intensity. Programmes typically lasting >3-6 months or involving >30 contact hours. High dosage is required to match high risk (RNR principle).	Short-term, low-dosage interventions (<15 hours)

Setting	Clinical AND Justice Settings. Hospitals, outpatient clinics (CAMHS), Prisons, Young Offender Institutions (YOI), Secure Children's Homes, and Probation/YOT offices.	Community settings (schools/youth clubs) where the intervention is low-intensity or universal
Activities	Tailored individual or small therapeutic group sessions.	Programmes with modules mainly on psychoeducation about traumatic stress, triggers, affect regulation, trauma narrative/processing, grief/loss processing, trauma-related cognitions
Participation	Targeted indicated/ clinical referral	n/a
Who delivers the intervention	Licensed mental health professionals (Psychologists, Psychiatrists, Clinical Social Workers, Therapists), Forensic Practitioners (To include assistant forensic or clinical psychologists) with advanced, specific certification in the intervention (e.g., correctional psychologists, specialised probation therapists). *Note: An inclusion criterion on who delivers the intervention in the Riise review (2021) is 'therapists who are practicing clinicians for whom provision of service is a substantial part of the job	Lay health workers, teachers, or general youth workers. *Note: Practitioners delivering Low-Intensity CBT (e.g., Psychological Wellbeing Practitioners, Education Mental Health Practitioners) are excluded from this strand and captured in Strand 1, unless the intervention meets the high-intensity dosage criteria (>30 hours/6 months).

Timing	Any	n/a
Study design	Randomised controlled trials (RCTs) and robust quasi-experimental designs (QEDs) e.g. those using difference-in-difference, synthetic controls, regression discontinuity, propensity score matching, or instrumental variables.	Studies without a credible counterfactual e.g. pre/post designs, QEDs with non-robust matching techniques

Details of searching

Prior to running academic database searches, the review team had already identified a number of potentially relevant studies from existing sources (n=121). These included studies coded within the Evidence and Gap Map by the previous team, studies already included at the EGM stage by the current team, and references identified from previous reviews. In addition, a further set of records was identified through academic database searching.

The academic search strategy was developed iteratively and refined through testing in MEDLINE. A set of pre-identified benchmark studies were identified by YEF at scoping stage, selected for their relevance to the review parameters and their coverage of key populations, interventions, and outcomes, was used to assess the sensitivity of the search. The strategy was refined until all benchmark studies were successfully retrieved, providing confidence that the search was capable of identifying relevant literature. The search was designed to capture both CBT Skills programmes and Clinical CBT.

The final MEDLINE search strategy was subsequently adapted, with appropriate modifications to controlled vocabulary, syntax, and field codes, for use across additional databases and platforms.

Searches were conducted in the following sources:

- Ovid MEDLINE (ALL)

- Child Development & Adolescent Studies (EBSCOhost)
- Open Dissertations (EBSCOhost)
- Criminal Justice Abstracts (EBSCOhost)
- ERIC (EBSCOhost)
- APA PsycInfo (Ovid)
- PsycArticles (Ovid)
- Embase (Ovid)

These databases were selected to ensure comprehensive coverage across psychology, education, criminal justice, health, and grey literature sources. The core conceptual structure of the search was retained across platforms, with database-specific adaptations applied to maximise sensitivity while maintaining relevance.

Details of screening

A total of 11,500 records were identified as potentially relevant for the CBT strands. These records were first deduplicated in EndNote, leaving 9,726 records to be imported into EPPI-Reviewer. Following further deduplication in EPPI-Reviewer, 8,651 records remained for title and abstract screening.

The EPPI-Reviewer classifier conducted initial title and abstract screening. A total of 1,772 records were marked as included and 6,879 as excluded. As a quality assurance check, a random sample of 10% of records excluded by the classifier was reviewed by senior team members.

A team of reviewers then screened the full text for all included studies. Senior reviewers conducted a check of any exclusions, and any discrepancies were discussed as a team and reconciled.

Of the 1,772 records screened at full-text level, 1,002 were excluded (for reasons, see Table 22). For inaccessible PDFs, the team attempted to contact lead authors to request access to the report or further data.

Table 22: *Full-text screening results*

Reason for exclusion	Number of Records Excluded at Full-Text Level

Study design not meeting robust causal threshold	343
Outcomes or intervention not relevant	545
Did not target CYP	82
Better placed in a different YEF Toolkit strand	5
Language	31
Duplicate	1

Following full-text screening, 770 studies were flagged as potentially relevant for inclusion. When combined with the 121 records identified via other sources, this gave a total of 891 records assessed as potentially relevant for inclusion.

***Table 23:* Reasons for exclusion after full text screening**

Reason for exclusion	Number of Records Excluded
Retained as EGM records but not causal or CBT-relevant	863
Study design	6
Intervention not relevant	2
Outcomes not disaggregated	1
Multiple reports of same study	2

Initially the team identified 28 Clinical CBT interventions which all went through our normal process for inclusion based on our scoping note. Following this assessment, 19 records were identified as fitting the scope for Clinical CBT. This number reduced to 17 unique records as three records were multiple reports of the same study (Table 24).

***Table 24:* List of linked studies**

Master study	Linked studies

Townsend et al. (2024)	Wood et al. (2020); Storch et al. (2022)

Of the 17 unique study records, 10 studies provided implementation evidence, and 16 studies provided effectiveness evidence. Of these 16 studies, 2 did not provide usable quantitative data for meta-analysis. This left 14 studies contributing usable quantitative evidence to the meta-analysis.

The reasons for excluding these two records from the meta-analysis were as follows:

- **Raine et al. (2016)** met the inclusion criteria for an impact evaluation; however, the study included four groups, of which only two were relevant to this review (CBT vs control). Although the authors reported an intention-to-treat analysis, the results were presented only as unadjusted means with confidence intervals, which did not provide sufficient information for effect size extraction. The author was contacted for additional data but was unable to provide the information requested. The study was excluded from the meta-analysis but was considered in the narrative description of the evidence base on effectiveness.
- **Reardon and Tosi (1977)** also met the inclusion criteria. However, the study used ANCOVA to evaluate the effectiveness of the intervention relative to a control group and reported only approximate p-values. Attempts to locate contact details for the author and request further information were unsuccessful. This study was therefore included in the narrative description of the intervention but was not included in the meta-analysis or reported in Appendix 3. The study was excluded from the meta-analysis but was considered in the narrative description of the evidence base on effectiveness.

Therefore, this report includes **17 unique studies**:

- **10 studies** which provided implementation evidence,
- **16 studies** which provided effectiveness evidence in the broader review,
- **14 studies** which provided usable quantitative data for the meta-analysis.

All exclusions made following full-text screening were reviewed by a senior reviewer to ensure consistency and accuracy.

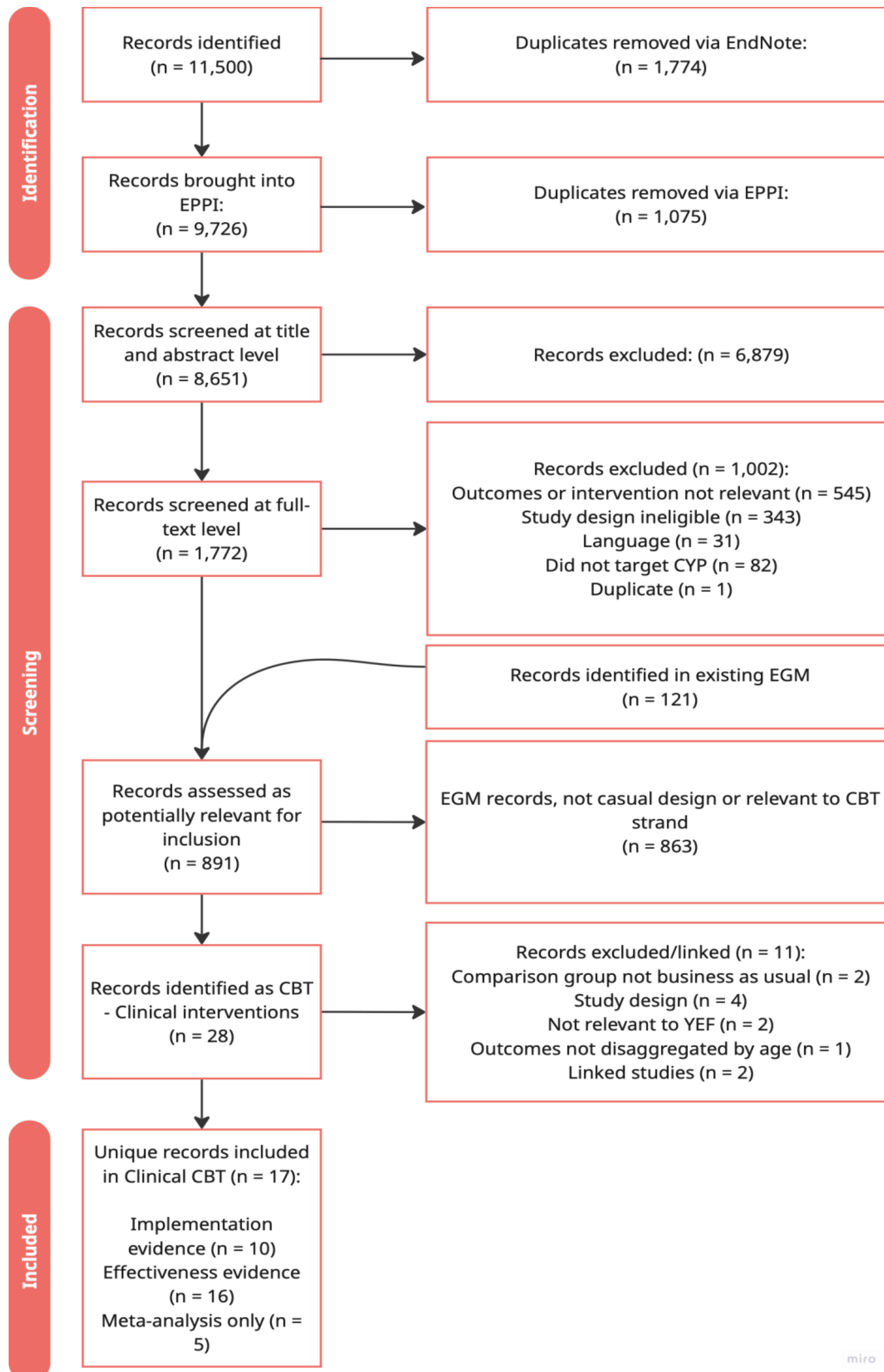

Figure 6: *PRISMA Flow Diagram*

Quality appraisal process for effectiveness studies

The YEF-EQA tool was used across all studies to systematically assess the quality, reliability, and relevance of the research.

Table 25**: Quality appraisal ratings for studies included in the Clinical CBT strand**

Study ID	Overall quality of the study	Study Design
Barrett (2001)	Moderate	RCT and PE
Clifford (2022)	Moderate	RCT
Curran (2009)	Low	RCT
Ducharme (2021)	Moderate	RCT and PE
Erickson (2014)	Moderate	RCT and PE
Esposito-Smythers (2011)	High	RCT and PE
Helander (2023)	Low	RCT
Hoogsteder (2018)	Moderate	QED and PE
Koegl (2008)	Moderate	RCT
Kolko (2006)	Moderate	RCT and PE
Rohde (2004)	Moderate	RCT
Soleimani-Rad (2025)	High	RCT and PE
Townsend (2024)	Moderate	RCT and PE
van de Wiel (2007)	Moderate	RCT and PE

How the findings were analysed and combined

Preparing the data frame for analysis

Included studies comprised a mixture of individually randomised controlled trials, cluster randomised controlled trials, and robust quasi-experimental designs. For all studies, the primary effect size was the standardised mean difference (SMD) expressed as Hedges' g.

As is common in meta-analysis, the team encountered several challenges in harmonising effect sizes across studies. First, there was a need to standardise the direction of SMD values so that negative values consistently indicate a favourable intervention effect. To achieve alignment across all outcomes, we reviewed all effect size direction labels and, where necessary, multiplied positive SMD values by −1 for outcomes where a lower score signified improvement This ensured that all SMDs reflected the same directional meaning, **that negative SMD values always indicate beneficial effects.** This approach supports clear interpretation and comparability across studies within the meta-analysis.

Two outcomes were odds ratios and so were approximately converted into an equivalent SMD by multiplying the natural log of the $\mathrm{OR} \times \frac{\sqrt{3}}{\pi}$ (approximately 0.5513).[11] In practical terms:

$$SMD\ (Cohen's\ d)\ \approx \frac{\sqrt{3}}{\pi}\ln(OR)\ = \frac{\ln(OR)}{1.814}$$

Additionally, eight outcomes across two studies were from cluster randomised controlled trials. In a cluster trial, randomisation occurs at a group-level (e.g. school, classroom). People within the same cluster tend to be more similar to each other than to people in other clusters. There is therefore less information about the effectiveness of an intervention in a trial of 100 people in 10 schools than in a trial of 100 people randomised individually. This non-independence of participants within clusters must be accounted for or the effect size will be over-estimated. For cluster randomised trials which reported the effect sizes from

[11] Because this transformation is a linear scaling, the standard error of the SMD can be obtained by applying the same $\frac{\sqrt{3}}{\pi}$ factor to the standard error of the log OR

cluster-adjusted analyses (e.g. mixed-effects models, generalised estimating equations) along with their standard errors or confidence intervals, we converted those estimates directly to Hedges' g and derived the standard error of g from the reported uncertainty, with no further adjustment required. However, for cluster randomised trials analysed as if they were individually randomised, the effect size precision was adjusted using a design effect (Rutterford et al., 2015). The design effect (DE) was defined as:

$$DE = 1 + (M - 1)\rho$$

Where M is the average cluster size and ρ is the intracluster correlation coefficient (ICC). For these studies we retained the point estimate of g but inflated its standard error by DE (equivalently, we reduced the effective sample size in each arm by dividing the original sample size by DE before computing the sampling variance). When ICCs were not reported, we used values from similar studies in the same substantive area, matching outcome type and reporter (self-report, parent, or teacher) as closely as possible (Parker et al., 2025). The impact of these assumptions was examined in sensitivity analyses by varying the ICC over a plausible range.

Meta-analysis

A random-effects model was fitted to the data. The amount of heterogeneity (i.e., τ^2), was estimated using the restricted maximum-likelihood estimator (Viechtbauer, 2005). In addition to the estimate of τ^2, the Q-test for heterogeneity (Cochran, 1954) and the I^2 statistic (Higgins & Thompson, 2002) are reported. In case any amount of heterogeneity is detected, a prediction interval for the true outcomes is also provided (Riley et al., 2011).

Publication bias

To assess possible publication bias, effect sizes were first aggregated to a single estimate per study using the aggregate() function in the metafor package, to avoid non-independence arising from multiple effect sizes contributed by the same study. A study-level random-effects model was then fitted using rma.uni() with REML estimation. Publication bias / small-study effects were examined using visual inspection of a funnel plot, Egger's regression test for funnel plot

asymmetry, and Begg and Mazumdar's rank-correlation test. Duval and Tweedie's trim-and-fill procedure was used as a sensitivity analysis to estimate the number of potentially missing studies and the corresponding adjusted pooled effect. In addition, a step-function selection model was fitted using selmodel() with two-sided p-value intervals of ≤ .025, .025–.05, .05–.10, and > .10 to examine whether the probability of study inclusion varied according to statistical significance.

Sensitivity analyses

Studentized residuals and Cook's distances are used to examine whether studies may be outliers and/or influential in the context of the model (Viechtbauer & Cheung, 2010). Studies with a studentized residual larger than the $100 \times (1-0.05/(2 \times k))$th percentile of a standard normal distribution are considered potential outliers (i.e., using a Bonferroni correction with two-sided $\alpha=0.05\alpha=0.05$ for k studies included in the meta-analysis). Studies with a Cook's distance larger than the median plus six times the interquartile range of the Cook's distances are considered to be influential. The rank correlation test (Begg & Mazumdar, 1994) and the regression test (Sterne & Egger, 2005), using the standard error of the observed outcomes as predictor, are used to check for funnel plot asymmetry. The analysis was carried out using R (version 4.4.2) (R Core Team, 2020) and the metafor package (version 4.8.0) (Viechtbauer, 2010).

Sub-group analysis

Subgroup-specific effects were estimated using robust variance estimation with CR2 adjustment, and statistical significance was assessed using t-tests with Satterthwaite small-sample degrees of freedom. We used coef_test() from the clubSandwich package to compute cluster-robust standard errors (CR2) with Satterthwaite degrees of freedom for hypothesis testing. Differences in effects across outcome domains were examined using a CR2-robust Wald chi-square test.

Results tables presented in the main report show the robust effect size (SMD) of interventions on different subgroup domains, with robust standard errors (SE) and study counts (k = number of effect sizes; n = number of studies). Negative SMDs

indicate a reduction/improvement for the intervention group compared to BAU control.

All estimates below are calculated using robust variance estimation (RVE) to account for dependence among effect sizes within studies. p-values correspond to two-tailed tests.

Implementation data

Information on factors that influenced, or were perceived to influence, implementation was extracted from studies where this was reported by study authors.

To capture implementation outcomes the toolkit data extraction made use of Proctor et al's (2011) Implementation Outcomes Framework to capture and categorise the barriers and facilitators to achieving good implementation.

The data extraction for the toolkit is an extension of what is already captured in the EGM. For the EGM, the focus was on whether or not implementation outcomes were measured. In other words, does a study report on indicators of how well the programme/intervention was implemented or not. For toolkit data extraction we capture why implementation did or did not go well, what influenced implementation? This is typically thought of as barriers and facilitators to implementation. Information on barriers and facilitators will be presented using Proctor et al's (2011) Implementation Outcomes as headings so that the reader can understand the evidence, and gaps in the evidence, on the following implementation outcomes:

- **Acceptability**: Stakeholders' perceptions that the intervention or change is agreeable, palatable, or satisfactory.
    - Example indicators: Children's views on the intervention, participant engagement, satisfaction with content or delivery.
- **Adoption:** The decision or action to employ an intervention or implementation target.
    - Example indicators: Uptake of the intervention by services, schools, or communities.

- **Appropriateness:** The perceived fit or relevance of the intervention to the given context or problem.
  - Example indicators: Adaptations made to improve the intervention's fit with the context, perceived usefulness.
- **Feasibility**: The extent to which the intervention can be successfully implemented in a specific setting.
  - Example indicators: Evidence of practicality or utility, ability to deliver the intervention in the target environment.
- **Fidelity**: The degree to which the intervention was delivered as intended.
  - Example indicators: Training quality, dosage and intensity of the intervention, adherence to the prescribed approach.
- **Reach/Penetration:** The extent to which the intervention has been integrated into a service setting or reached eligible recipients.
  - Example indicators: Ratio of recipients served to the target population, evidence of saturation or integration.
- **Sustainability:** The ability to maintain or institutionalise the intervention over time.
  - Example indicators: Evidence of routinisation, integration into policies or practices, durability of implementation efforts.

The information extracted on each implementation outcome was narratively summarised. Where implementation barriers/facilitators or influences on an implementation outcome were not measured and/or reported, this is stated.

YOUTH ENDOWMENT FUND

NATIONAL CHILDREN'S BUREAU

Appendix 2. Location Details

	Number of UK Studies	Number (and Location) of International Studies
Contributing to Evidence Quality Rating	1	Netherlands (N=3); Sweden (N=1); USA (N=7); Canada (N=1); Iran (N=1)
Contributing to Estimated Impact on Violence	1	Netherlands (N=3); USA (N=3); Canada (N=1); Iran (N=1)
Contributing to Estimated Impact on Crime and Offending	1	Netherlands (N=1); USA (N=2) ; Canada (N=1) ; Iran (N=1)
Contributing to EDIE Data	0	Netherlands (N=1); Sweden (N=1)
Contributing to Implementation Data	0	Netherlands (N=2); USA (N=7); Iran (N=1)
Contributing to Cost Data	0	0

YOUTH ENDOWMENT FUND

NATIONAL CHILDREN'S BUREAU

Appendix 3. Characteristics of included studies for meta-analysis

Studies contributing to the estimated impact on violence are marked with an asterisk (*).

Authors (Year)	Country	Study Design	Intervention	Population / Place	Delivery Setting	Intervention Provider	Comparison	Outcomes Measured	Quality Level	Findings
Barrett (2001)	USA	RCT and PE	Individual CBT designed to teach self-control and coping skills to prevent substance use.	Adolescents using drugs other than alcohol and tobacco.	Outpatient setting	Psychologist	Treatment as usual	Drug and alcohol use	Moderate	For those in the individual CBT intervention group, there were no significant effects found with regards to the percentage of days adolescents used marijuana, at either post-testing or the 3 month follow-up.

***Clifford (2022)**	Netherlands	RCT	Anger Can Go! is designed to treat anger regulation problems in autistic children aged 8-13.	8-13 year olds with autism.	Outpatient setting	Psychologist	Parental psychoeducation only (active control group)	Behavioural difficulties Regulating and managing emotions	Moderate	The intervention had no significant effect on the Aggressive Behaviour Problems scale, but children in the treatment group increased their use of adaptive anger regulation strategies.
***Curran (2009)**	Scotland	RCT	The Ross Programme is targeted at young people whose antisocial behaviour has led to them being	13-16 year old boys	Education and Care Centre	Unclear	Treatment as usual (waitlist control)	Ability to resolve conflict Behavioural difficulties Crime and	Low	There was a significant difference between the treatment and control groups at endline, suggesting that the programme

			under the supervision of specialised schools, social services or youth justice agencies.					offending		reduced the risk of reoffending in the treatment group.
***Ducharme (2021)**	USA	RCT and PE	Anger Control Training (ACT) augmented with Regulate and Gain Emotional Control (RAGE-Control) is designed to combat high attrition rates and lack of	Young people with clinically significant anger dyscontrol/ aggression	Outpatient setting	Social workers	Treatment as usual	Behavioural difficulties Crime and offending General mental health	Moderate	The results indicate that while both treatment groups improved, the group using the active biofeedback game (ACT-R) saw significantly better behavioural and physiological outcomes than the sham group

YOUTH ENDOWMENT FUND

NATIONAL CHILDREN'S BUREAU

			real world skill translation in traditional programmes targeted at youth emotional dysregulation .								(ACT-S). The most significant finding was that the intervention improved the control of the behavioural expression of anger (e.g., aggression and oppositionality) rather than reducing the internal experience of angry feelings. The findings showed that larger decreases in median heart rate during gameplay were

										significantly associated with larger reductions in aggression and oppositional behaviours in the real world.
***Erickson (2014)**	USA	RCT and PE	The Aggression Replacement Training (ART) intervention is designed to target adolescent aggression by enhancing prosocial skills, managing angry	Girls who were juvenile offenders without physical or mental disabilities	Residential facility	Curriculum trainer	Treatment as usual	Behavioural difficulties Regulating and managing emotions	Moderate	Overall, the results showed that the ART programme did not lead to significant reductions in aggressive or rule breaking behaviour compared with the comparison group, either before or after controlling for traumatic

			feelings and advancing moral reasoning.							distress, and PTSD levels did not moderate the intervention's effects. Positive behaviours increased over the study period for both groups, but ART participation did not produce significantly greater gains.
Esposito-Smythers (2011)	USA	RCT and PE	Integrated Cognitive Behavioural Therapy (I-CBT) designed for adolescents with co-	13-17 year olds with current alcohol or cannabis use disorder and a	Community setting	Psychologist	Enhanced treatment as usual	Behavioural difficulties Crime and	High	I-CBT was associated with significantly fewer heavy drinking days and marijuana use days (but not total drinking

			occurring alcohol/drug use disorders and suicidality.	recent suicide attempt				offending Drug and alcohol use General mental health		days) compared to E-TAU. I-CBT participants had significantly fewer suicide attempts, hospitalisations, ED visits, and arrests.
Helander (2023)	Sweden	RCT	KOMET combined with the child group-based CBT program, Coping Power Program (CPP) trains children in emotion regulation, anger management	8-12 years old, diagnosed with ODD, ODD combined with CD or disruptive behavioural disorder NOS	Not stated	Unclear	Treatment as usual	Behavioural difficulties Helping others (prosocial behaviours)	Low	Both treatments showed overall improvements, but KOMET + CPP did not produce better long term outcomes than KOMET alone. ODD symptoms decreased substantially in both groups, with KOMET alone

			skills, social problem-solving skills, perspective taking, social skills and handling group pressure.					Parenting practices Regulating and managing emotions		continuing to improve after treatment and eventually matching the combined programme. Social skills gains seen during treatment in the KOMET + CPP group were not maintained, with some decline during follow up, while KOMET alone remained stable. Parenting outcomes were similar across groups: harsh parenting

										decreased moderately, parental praise slightly declined over time, and parental stress did not change. Overall, the added CPP component did not yield sustained long term benefits.
***Hoogsteder (2018)**	Netherlands	QED and PE	Responsive Aggression Regulation Therapy (Re-ART) is designed to decrease recidivism in adolescents with severe	Adolescents with aggression problems.	Juvenile justice facility	Therapist	Treatment as usual (e.g. cognitive therapy and art therapy)	Crime and offending	Moderate	The Re-ART group had a significantly lower risk of recidivism than the control group. The results were insensitive to moderators such as ethnicity, age

			aggression problems.							and substance use.
***Koegl (2008)**	Canada	RCT	SNAP (Stop Now And Plan) is a mental health programme for children with identified behaviour problems, designed to interrupt negative behaviour patterns.	Children under 12 are admitted if they have a T-score of 70 or greater on the delinquency subscale of the Child Behaviour Checklist	Not stated	Psychologist	Treatment as usual	Behavioural difficulties Crime and offending	Moderate	There were no statistically significant differences in change scores among the three groups on the delinquency and major and minor aggression subscales.
***Kolko (2006)**	USA	RCT	CBT to tackle fire setting and arson	Each child is involved in a fire in	Not stated	Psychologist	Treatment as usual (firefighter	Ability to resolve conflict	Moderate	There were decreased scores on the Children's

			among children.	the past three months and had damaged property.			home visit)	Behavioural difficulties Family relationship and support Helping others (prosocial behaviours) Parenting practices Regulating and managin		Hostility Inventory, fewer externalising problems on the CBCL scale, and improved general Family Functioning.

								g emotions		
Rohde (2004)	USA	RCT	The Coping Course Intervention is designed to enhance general coping and problem-solving skills among incarcerated youth.	Adolescent boys identified as having extreme concentration or behavioural problems	Correction facility	Psychologist	Treatment as usual	Behavioural difficulties Building and maintaining relationships General mental health Regulating and managing emotions	Moderate	The intervention had a significant effect on reducing externalising problems, reducing suicide proneness, increasing self-esteem, and increasing sharing of feelings with staff.

								Self-esteem		
***Soleimani-Rad (2025)**	Iran	RCT and PE	CBT for disruptive mood dysregulation disorder (DMDD) and child aggression, with an additional focus on implementing CBT in Iran.	Male primary school students, aged 7-12, from Zanjan City, meeting DSM-5-TR criteria for DMDD	Not stated	Psychologist	Treatment as usual	Crime and offending General mental health Regulating and managing emotions	High	The study found that the CBT group showed highly significant improvements compared to the wait-list control group across all measured areas. Specifically, children who received the 15-week CBT intervention demonstrated significant reductions in irritability, anger outbursts, and overt aggressive

										behaviours; decreased internalising problems and externalising problems; improved global functioning and a significant reduction in overall illness severity. The treatment effects were stable, meaning improvements were maintained without significant worsening when evaluated at the 3-month follow-up.

Townsend (2024)	USA	RCT	The Behavioural Interventions for Anxiety in Children with Autism (BIACA) programme is designed to reduce anxiety and improve adaptive functioning for youth with ASD.	7–13-year-olds with a diagnosis of ASD.	Not stated	Therapist	Treatment as usual	Building and maintaining relationship General mental health	Moderate	Results showed that, in comparison to a control group, children who received BIACA displayed significantly improved scores on clinician and parent rated internalising measures and adaptive functioning, at post testing.
***van de Wiel (2007)**	Netherlands	RCT	The Utrecht Coping Power Programme (UCPP) is an adaptation of	8-13 year olds with a diagnosis of disruptive	Not stated	Therapist	Treatment as usual (mental health services	Behavioural difficulties	Moderate	Overall, children in both the UCPP and control groups showed improvements

			the Coping Power Programme, a preventative intervention.	behaviour disorder			typically offered at the facilities)			over time, but UCPP produced a significantly greater reduction in overt aggression than the control condition. For oppositional behaviour and broader externalising problems, both groups improved similarly, with no evidence that UCPP was more effective than the control group on these outcomes.

Appendix 4. Characteristics of included studies for implementation

Authors (Year)	Country	Study Design	Intervention (Name and brief description)	Quality Level	Summary of Findings
Barrett et al. (2001)	USA	Process evaluation	**Individual CBT:** A model designed to teach individual young people self-control and coping skills useful in avoiding substance abuse.	Moderate	**Fidelity:** Adherence was supported through treatment manuals, session checklists, videotaped sessions, and expert supervision; random session rating showed high fidelity for CBT (M = 8.91, SD = 1.04) on a 10-point adherence scale.
Ducharme et al. (2021)	USA	Process evaluation	**Anger Control Therapy (ACT) plus RAGE-Control (Regulate and Gain Emotional Control) video game:** ACT is an empirically supported, manualized CBT for anger control. Augmented with RAGE-Control, a video	Moderate	**Acceptability:** Early dropout rates indicate participant engagement and acceptability. In ACT-R, one participant out of 20 dropped out compared to five out of 20 in the sham group (ACT-S). **Feasibility:** Participants in the ACT-R group successfully achieved decreased median heart rate with gameplay across the 10 sessions, demonstrating that the game could be used to

			game developed to provide a safe but demanding environment to practice skills learned from ACT.		practice heart rate regulation with skills taught in the therapeutic setting. **Fidelity:** There were several structured procedures including delivery of ACT sessions using a detailed manual, a two hour training session with weekly supervision, and monitoring of implementation through audio recording of sessions, resulting in a high mean fidelity score of 93%.
Erickson et al. (2014)	USA	Process evaluation	**Aggression Replacement Training (ART):** A 10-week, evidence-based, group treatment intervention designed to advance moral reasoning, improve social skills, and manage angry feelings.	Moderate	**Fidelity:** Monitored through direct observation, videotaped session review, and standardised evaluations. The single trainer received a "nearly competent" rating (1.8/3), but did not attend booster training indicating partial adherence to ART protocols.

Esposito-Smythers et al. (2011)	USA	Process evaluation	**Integrated outpatient cognitive behavioural intervention (I-CBT)**: a cognitive-behavioural treatment protocol for adolescents with a co-occurring Alcohol or other Drug use disorder (AOD) and suicidality.	High	**Acceptability:** Retention and completion rates indicate participant acceptability. Approximately 74% of families in the I-CBT group were classified as "treatment completers", attending at least 24 adolescent sessions and 12 parent sessions over the intensive 12-month protocol. **Appropriateness:** The I-CBT manual was specifically adapted from earlier protocols to ensure it targeted the functional relationship between substance abuse and suicidal behaviour, making it a more appropriate match for the participants' complex needs. **Fidelity:** A rigorous training and supervision framework ensured implementation with fidelity, in accordance with manualised protocol. All treatment sessions were audiotaped and monitored. 92% of sessions met the competence criteria and 85% met the adherence threshold.

Goldston et al. (2021)	USA	Process evaluation	**Cognitive-Behavioural Therapy – Relapse Prevention (CBT-RP):** targeted at youth with depression, suicide attempts and substance use problems.	Low	**Acceptability:** both parents and young people found the intervention acceptable overall, although more parents than young people said they would recommend the programme. The therapists themselves were viewed as a positive aspect of the programme. **Feasibility:** the authors believed that the intervention had high feasibility due to only one participant not completing the treatment.
Hoogsteder et al. (2018)	Netherlands	Process evaluation	**Responsive Aggression Regulation Therapy (Re-ART**): an accredited behavioural intervention specifically for young people aged 16 to 21 who exhibit severe aggression problems and a high risk of recidivism	Moderate	**Fidelity:** Evaluation forms indicated that the programme was generally delivered as intended, though areas for improvement included therapist-mentor communication, and consistent reinforcement of treatment goals by living group mentors.
Kolko (2006)	USA	Process evaluation	**Cognitive-Behavioural Treatment (CBT)**: CBT for	Moderate	**Fidelity:** Fidelity was assessed through a rating scale, where delivery was rated in comparison

YOUTH ENDOWMENT FUND

NATIONAL CHILDREN'S BUREAU

			boys aged 5-13 who had been referred for fire-setting. The participants were randomised to either CBT or fire safety education, with a third group of preassigned participants receiving a brief home visit from a firefighter.		to the intervention manual. CBT was reported to have high programme fidelity, with a mean rating of 92%.
Soleimani-Rad et al. (2025)	Iran	Process evaluation	**Cognitive behavioural therapy for disruptive mood dysregulation disorder (DMDD):** a structured, individual treatment program specifically designed to address the symptoms of DMDD in children aged 7 to 12.	High	**Fidelity:** Fidelity was supported through standardised delivery of 15 weekly CBT sessions by the same therapist and the use of compensatory make-up sessions for missed appointments.

Townsend et al. (2024)	USA	Process evaluation	**Behavioural Interventions for Anxiety in Children with Autism (BIACA):** a specialised form of adapted cognitive behavioural therapy designed specifically for youth with autism spectrum disorder (ASD) and comorbid anxiety to reduce anxiety and improve adaptive functioning.	Moderate	**Appropriateness:** The BIACA programme modified standard CBT to better fit youth with autism by including 16 extended child- and parent-focused sessions, a flexible modular format, behavioural management for disruptive behaviours, social skills training, integration of the child's special interests, and a comprehensive reward system across settings to enhance engagement and treatment relevance. **Fidelity:** Fidelity to BIACA was high, with therapists adhering to 97% of required components. To further ensure reliability, a second coder independently rated 14.3% of the sampled sessions, yielding strong interrater agreement (intraclass correlation = 0.85).
van de Wiel et al. (2007)	Netherlands	Process evaluation	**Utrecht Coping Power Program (UCPP):** an adaptation of the Coping Power Program (CPP), a school-based preventive	Moderate	**Appropriateness:** UCPP was adapted from a school-based preventive model to better fit outpatient clinical practice for children with disruptive behaviour disorders. Modifications included shortening treatment length,

			intervention program that includes a parent and a child component.		increasing activity-based sessions to match attention spans, and integrating parent and child components under the same therapists.

YOUTH ENDOWMENT FUND

NATIONAL CHILDREN'S BUREAU

Appendix 5. Availability of evidence according to each of Proctor et al.'s (2011) implementation outcomes

Authors (Year)	Acceptability	Adoption	Appropriate-ness	Feasibility	Fidelity	Reach/ penetration	Sustainability	Cost
Barrett et al. (2001)	No	No	No	No	Yes	No	No	No
Ducharme et al. (2021)	Yes	No	No	Yes	Yes	No	No	No
Erickson et al. (2014)	No	No	No	No	Yes	No	No	No
Esposito-Smythers et al. (2011)	Yes	No	Yes	No	Yes	No	No	No
Goldston et al. (2021)	Yes	No	No	Yes	No	No	No	No
Hoogsteder et al. (2018)	No	No	No	No	Yes	No	No	No
Kolko (2006)	No	No	No	No	Yes	No	No	No

Soleimani-Rad et al. (2025)	No	No	No	No	Yes	No	No	No
Townsend et al. (2024)	No	No	Yes	No	Yes	No	No	No
van de Wiel et al. (2007)	No	No	Yes	No	No	No	No	No

YOUTH ENDOWMENT FUND

NATIONAL CHILDREN'S BUREAU

Appendix 6. Moderator categories

Intervention-level moderators

Inclusion criteria

Inclusion category	Examples from text	Authors
Neurodiversity	• Neurodevelopmental condition • Primary ASD diagnosis • Primary ASD diagnosis WITH anxiety	Clifford et al. (2022) Townsend et al. (2024)
Externalising problems or difficulties (ie conduct disorder)	• Clinically significant anger dyscontrol/aggression • Disruptive behaviour disorder (DBD)	Ducharme et al. (2021) Helander et al. (2022) van de Wiel et al. (2007)
CYP with an offending history	• Juvenile offenders • Aggressive offenders • Fire involvement	Curran (2009) Erickson (2013) Hoogsteder et al. (2018) Kolko et al. (2006)

	• Children with a T-score of 70 or greater on the delinquency subscale of the CBCL /or have had recent police contact	Koegl et al. (2008) Rohde (2004)
Other psychiatric diagnosis	• Primary substance abuse disorder • Substance abuse disorder and suicide attempt • Disruptive mood dysregulation disorder (DMDD)	Barrett-Waldron et al. (2001) Esposito-Smythers et al. (2011) Soleimani-Rad et al. (2025)

Intensity

Category	Example	Studies
Very intense programmes	These require substantial time commitments, often lasting six months or more, meeting multiple times per week, or requiring simultaneous participation in	Studies categorised as very intense include: • Esposito-Smythers 2011 – protocol lasts 12 months, divided into a 6-month active phase (weekly

	several different therapy components (e.g., child, parent, and family modules).	sessions), a 3-month continuation phase (bi-weekly), and a 3-month maintenance phase (monthly). Families randomised to the treatment condition received a highly intensive dose of treatment, attending an average of about 45 total sessions combined across individual, family, and parent training formats. • Hoogsteder et al. 2018 – highly individualised with both individual (minimum 1 hour/week) and group modules (12-14 sessions, 1.5 hours each). The total duration of the intervention can vary from half a year to about two years. Participants in the study averaged approximately 46 weeks of treatment.

		• Helander et al. (2022) – combined two full programmes. It included 11 parent group sessions lasting 2.5 hours each (the KOMET program) run simultaneously with the children's CBT group (Coping Power Program). • Erickson 2014 – evaluated a 10-week group curriculum in a residential commitment programme. The intensity is high due to its frequency: young people attended 3 sessions per week, with each session lasting 1 to 1.5 hours. • Koegl et al. (2008) – the enhanced SNAP outreach project included 12 weekly 90-minute child groups and 12 weekly parent groups, further augmented by individual

		befriending, family counselling, and academic tutoring sessions. • van de Wiel et al. (2007) – the Utrecht Coping Power Program (UCPP) is a 9-month, multi-modal program that integrates both child and parent components. The protocol calls for 23 child group sessions and 15 parent group sessions. On average, participants in this condition completed about 33 total sessions.
Moderately intense programmes	They typically involve weekly or bi-weekly sessions lasting 1 to 1.5 hours and are completed over a span of 8 to 16 weeks.	Studies categorised as moderately intense include: • Townsend et al. (2024) – the standard CBT (Coping Cat) was delivered over 16 weekly sessions lasting 60 minutes.

		• Rohde et al. (2004) – the Coping Course consisted of 16 group treatment sessions delivered twice a week over an 8-week period. • Soleimani-Rad et al. (2025) – the CBT protocol for disruptive mood dysregulation disorder consisted of 15 weekly 60-minute sessions. • Barret-Waldron 2011 – Adolescents received 12 hours of therapy in the standalone CBT condition. • Clifford et al. (2022) – the "Anger Can Go!" program consisted of 9 individual 60-minute sessions for the child and 3 psychoeducation group sessions for the parents (12 sessions total).

Low intensity programmes	These interventions are the shortest in duration, requiring 10 or fewer total contact sessions.	Studies categorised as low intensity include: • Ducharme et al. (2021) – the Anger Control Training protocol combined with the RAGE-Control video game was delivered in exactly 10 weekly 1-hour sessions. • Kolko et al. (2006) – all conditions in this study were relatively brief. The Firefighter Home Visit (FHV) was the shortest, with CBT conditions averaging only 5.5 and 7.4 sessions, respectively, over an 8.7-week period. • Curran (2009) – did not report details on the intensity of the intervention and evaluated a specifically abbreviated "Short Version" of the Ross Programme for

		their participating children and young people.

Population-level moderators

Ethnicity

Category	Description
Majority white Sample	>85% of sample identified as white
Some Diversity	Between 15–49% of participants identified as Black and Global Majority (i.e., Black, Hispanic/Latinx, Asian, Indigenous, etc.)
Diverse / Balanced	50% or more Black and Global Majority youth in the sample
Not stated / unclear	Study does not specify the ethnicity of participants

Sex/Gender

Category	Description

All male	Sample is entirely male
All female	Sample is entirely female
Majority male	Sample included >70% male
Majority female	Sample included >70% female
Mixed sex	40–70% male and 30–60% female
Not stated / unclear	Study did not report the gender breakdown of participants

Outcome moderators

Outcome type

Category	Description
Ability to resolve conflicts	Skills to de-escalate conflicts and arguments.
Behavioural difficulties	Disruptive or aggressive behaviours due to distress or needs, excluding hyperactivity and neuro-diverse conditions.
Building and maintaining relationships	Social-emotional skills for listening, cooperating, and understanding emotions.

Crime and offending	Criminal behaviours including non-violent acts (e.g., shoplifting), sexually violent crimes, and violent acts (e.g., assault, robbery).
Drug and alcohol use	Problematic use of substances causing harm to self or others, impacting health, school, or social functioning.
Family relationships and support	Positive and supportive connections with family members.
General mental health	Encompasses psychological mental health, which specifically relates to an individual's cognitive and emotional functioning, including their ability to manage thoughts, behaviours, and cope with life's challenges. Includes mental health conditions like anxiety, depression, PTSD
Helping others (prosocial behaviour)	Engaging in positive, prosocial behaviours such as helping, comforting, or sharing.
Parenting practices	Parenting practices, often learnt from a parent/carer's own experience of being a child, that are not appropriate for a certain situation. It includes harsh or inappropriate discipline, controlling behaviour, inconsistent parenting, or low parental warmth.
Regulating and managing emotions	Also called 'emotion regulation'. Having the skills and techniques to manage feelings and reactions to situations and events, reducing the intensity, duration, and impact of such feelings.

Self-esteem	Viewing yourself positively, including confidence in your abilities, appearance, and self-worth.

Data source

Category	Description
Administrative/school records	From school or criminal justice system records e.g. exam grades, school attendance
Clinical assessment/ Observation	Measured by the researcher directly observing the behaviour
Parent report	Reported by a parent through a survey or interview
Peer report	Reported by a peer through a survey or interview
Child self-report	Reported by the young person through a survey or interview
Teacher report	Reported by a teacher through a survey or interview
Not stated	Study does not report how the outcome was measured

Appendix 7. Measured outcomes across included studies for meta-analysis

YEF outcome framework category *Category description (as described in the framework)*	Measured outcomes *(descriptions by study authors)*	Studies
Crime and offending	"Youth Level of Service (YLS) Total"; "Modified Overt Aggression Scale (MOAS)"; "Arrest"; "SAVRY"; "Mean number of offences - Violence offences"; "Mean number of offences - Property with violence"; "Mean number of offences - Property crimes"; "Mean number of offences - General recidivism"; "Frequency and Severity of General Recidivism after Three Year Follow-up - Number of offenses"; "Frequency and Severity of General Recidivism after Three Year Follow-up - Severity of offenses"; "Frequency and Severity of General Recidivism after Three Year Follow-up - Number of prison sentences"; "Frequency of Recidivism after One year - property with violence"; "Frequency of	**(n = 6, k = 35)** Curran (2009); Ducharme (2021); Esposito-Smythers (2011); Hoogsteder (2018); Koegl (2008); Soleimani-Rad (2025)

	Recidivism after Two years - property with violence"; "Frequency of Recidivism after three years - property with violence"; "Frequency of Recidivism after two years - violent offenses"; "Frequency of Recidivism after three years - violent offenses"; "Frequency of Recidivism after two years - property offenses"; "Frequency of Recidivism after three years - property offenses"; "Frequency of Recidivism after one year - general recidivism"; "Frequency of Recidivism after two years - general recidivism"; "Frequency of Recidivism after three years - general recidivism"; "Modified Overt Aggression Scale - Child (MOAS-C)"; "Modified Overt Aggression Scale - Parent (MOAS-P)"; "Teacher Report Form - Externalizing (TRF-E)"; "Child Behaviour Checklist - Externalizing (CBCL-E)"; "Youth Self-Report - Externalizing (YSR-E)"; "Parent rated CBCL Delinquency"	
Drug and alcohol use	"Percentage of Days of Use for Adolescent Marijuana Use"; "TimeLine FollowBack (TLFB) drinking days"; "TLFB heavy drinking days"; "TLFB days marijuana use"; "TLFB abstinence"; "Rutgers Alcohol Problem Index (RAPI)"; "Rutgers Marijuana Problem Index (RMPI)"	**(n = 2, k = 8)** Barrett (2001); Esposito-Smythers (2011)
Behavioural difficulties	"Aggressive behaviour problems (QSB)"; "aggression"; "Behavioral Anger Response Questionnaire for Children (BARQ) Maladaptive strategies"; "Total Difficulties"; "Disruptive Behaviour Disorders Rating Scale (DBDRS)"; "Aggressive behaviour in the classroom (e.g., argues,	**(n = 10, k = 26)** Clifford (2022); Curran (2009); Ducharme (2021); Erickson (2014);

	fights, attacks, destroys things)"; "Rule-breaking behaviour in the classroom (e.g., lies, cheats, steals, truant)"; "Aggressive behaviour outside of the classroom"; "Behavioural Assessment Scale for Children (BASC) - Conduct Problems"; "Ran away"; "Child behaviour problems - The Parent/Teacher Disruptive Behaviour Disorder rating scale - Oppositional Defiant Disorder subscale"; "Fire Attraction and Interest Scale (FAIS) Aggregate"; "Children's Hostility Inventory"; "CBCL Externalizing"; "Youth Self-Report Externalizing behaviours"; "overt aggression"; "oppositional behaviour"; "externalising behaviour"; Parent rated CBCL "Major aggression"; Parent rated CBCL "Minor aggression"	Esposito-Smythers (2011); Helander (2023); Koegl (2008); Kolko (2006); Rohde (2004); van de Wiel (2007)
Ability to resolve conflict	"Avoidance Style"; " Social Problem Solving (SPS) prosocial solutions"	**(n = 2, k = 2)** Curran (2009); Kolko (2006)
Building and maintaining relationships	"Arguments with staff"; "Sharing with staff"; "Social Responsiveness Scale-2 - Social Communication/Interaction scale"	**(n = 2, k = 3)** Rohde (2004); Townsend (2024)
Family relationship and support	"Parent Questionnaire"; " Family Assessment Device (FAD) General functioning"	**(n = 1, k = 2)** Kolko (2006)

General mental health	“Clinical Global Impression-Severity (CGI-S)”; “Clinical Global Impression-Improvement (CGI-I)”; "Columbia Impairment Scale (CIS) - Teen"; "Columbia Impairment Scale (CIS) - Parent"; "Suicide Ideation Questionnaire (SIQ)"; "Reynolds Adolescent Depression Scale-2 (RADS-2)"; "Screen for Child Anxiety Related Emotional Disorders (SCARED)"; "Behavioural Assessment Scale for Children (BASC) - Depression"; "Behavioural Assessment Scale for Children (BASC) - Anxiety"; "Suicide attempt"; "Inpatient hospitalization"; "Partial hospitalization"; "Emergency dept. visit"; "Youth Self-Report internalizing behaviours"; "Life Attitudes schedule-short form total"; "Child Behavior Checklist - Internalizing (CBCL-I)"; "Children's Global Assessment Scale (CGAS)”; "Youth Self-Report - Internalizing (YSR-I)"; "Teacher Report Form - Internalizing (TRF-I)"; "Paediatric Anxiety Rating Scale - Severity"; "Child Behaviour Checklist - Anxious and Depressed Scale"; "Child Behaviour Checklist - Internalising Scale"; "Social Responsiveness Scale-2 - Restrictive/Repetitive behaviour scale"; "Childhood Anxiety Impact Scale- School"; "Childhood Anxiety Impact Scale- Social"; "Childhood Anxiety Impact Scale- Family"; "Clinical Global Impressions-Improvement - Positive Treatment Response"	**(n = 5, k = 27)** Ducharme (2021); Esposito-Smythers (2011); Rohde (2004); Soleimani-Rad (2025); Townsend (2024)

YOUTH ENDOWMENT FUND

NATIONAL CHILDREN'S BUREAU

Helping others (prosocial behaviours)	"Child social skills - SDQ Prosocial"; "Child social skills - SSRS"	**(n = 1, k = 6)** Helander (2023)
Parenting practices	"Parenting skills and competences - PPI Harsh"; "Parenting skills and competences - PPI Praise"; "Child Rearing Inventory"	**(n = 2, k = 7)** Helander (2023); Kolko (2006)
Regulating and managing emotions	"BARQ Rumination"; "BARQ Adaptive strategies"; "Positive Behaviour in the form of scores (ratio level) derived from the Behaviour Incident Report"; "Child social skills - P-COMP"; "Coping skills"; "Affective Reactivity Index - Child (ARI-C)"; "Affective Reactivity Index - Parent (ARI-P)"	**(n = 5, k = 11)** Clifford (2022); Erickson (2014); Helander (2023); Rohde (2004); Soleimani-Rad (2025)
Self-esteem	"Self-Esteem"	**(n = 1, k = 1)** Rohde (2004)