

# The FRIENDS preventive programme for reducing anxiety symptoms in children and adolescents: A systematic review and meta-analysis

Trine Filges<sup>1</sup> | Geir Smedslund<sup>2</sup> | Tine Eriksen<sup>3</sup> | Kirsten Birkefoss<sup>4</sup> |  
 Malene Wallach Kildemoes<sup>1</sup>

<sup>1</sup>VIVE Campbell, VIVE – The Danish Center for Social Science Research, Copenhagen, Denmark

<sup>2</sup>Norwegian Medical Products Agency, Oslo, Norway

<sup>3</sup>VIVE – The Danish Center for Social Science Research, Aarhus, Denmark

<sup>4</sup>VIVE – The Danish Center for Social Science Research, Copenhagen, Denmark

## Correspondence

Trine Filges, VIVE Campbell, VIVE – The Danish Center for Social Science Research, Central Denmark Region, Copenhagen, Denmark.

Email: [tif@vive.dk](mailto:tif@vive.dk)

## Funding information

VIVE Campbell, Denmark

## Abstract

**Background:** Anxiety and stress responses are often considered normative experiences, and children and adolescents may benefit from anxiety prevention programmes. One such programme is FRIENDS which is based on a firm theoretical model which addresses cognitive, physiological and behavioural processes. FRIENDS is manualised and can, thus, easily be integrated into school curriculums.

**Objectives:** What are the effects of the FRIENDS preventive programme on anxiety symptoms in children and adolescents? Do the effects differ between participant age groups, participant socio-economic status, type of prevention, type of provider, country of implementation and/or implementation issues in relation to the booster sessions and parent sessions?

**Search Methods:** The database searches were carried out in September 2023, and other sources were searched in October 2023. We searched to identify both published and unpublished literature. A date restriction from 1998 and onwards was applied.

**Selection Criteria:** The intervention was three age-appropriate preventive anxiety programmes: Fun FRIENDS, FRIENDS for Life, and My FRIENDS Youth. Primary outcome was anxiety symptoms and secondary outcome was self-esteem. Studies that used a control group were eligible, whereas qualitative approaches were not.

**Data Collection and Analysis:** The number of potentially relevant studies was 2865. Forty-two studies met the inclusion criteria. Twenty-eight studies were used in the data synthesis. Four studies had a critical risk of bias. Six studies did not report data that enabled calculation of effect sizes and standard errors. Two studies had partial overlap of data to other studies used, and two were written in Persian. Meta-analyses were conducted on each outcome separately. All analyses were inverse variance weighted using random effects statistical models.

**Main Results:** Studies came from 15 different countries. Intervention start varied from 2001 to 2016. The average number of participants analysed was 240, and the average number of controls was 212. Twenty-five comparisons reported on anxiety

symptoms post-intervention. The weighted average standardised mean difference (SMD) was 0.13 (95% CI 0.04 to 0.22). There was some heterogeneity. Twelve comparisons reported on anxiety symptoms at 12 months follow-up. The weighted average SMD was 0.31 (95% CI 0.13 to 0.49). There was a large amount of heterogeneity. Five comparisons reported on self-esteem post-intervention with a weighted average SMD of 0.20 (95% CI -0.20 to 0.61) and a large amount of heterogeneity. At follow-up, we found evidence that programmes implemented by mental health providers appears to perform better than programmes implemented by teachers. The evidence was inconclusive beyond 12 months follow-up.

**Authors' Conclusions:** Our results indicate that the FRIENDS intervention may reduce anxiety symptoms in children and adolescents when reported by children and adolescents themselves. The majority of trials employed a wait-list design, implying only a few studies reported on the long-term effects of the FRIENDS intervention. Our findings suggest that the FRIENDS intervention may increase the reduction in anxiety symptoms 12 months after the intervention. This emphasises the need for future research that apply designs that allows for long-term follow-up. We are uncertain about the effects on self-esteem. The overall certainty of evidence varied from low to very low. There is a need for more rigorously conducted studies.

#### KEY WORDS

anxiety, FRIENDS, meta-analysis

## 1 | PLAIN LANGUAGE SUMMARY

There is evidence to indicate that the FRIENDS intervention may reduce anxiety symptoms in children and adolescents when measured by children and adolescents themselves.

### 1.1 | The review in brief

In this review, we aimed to find evidence of the effectiveness of the FRIENDS intervention, a cognitive behavioural therapy programme, on reduction of anxiety symptoms in children and adolescents. The evidence indicates that the FRIENDS intervention may reduce anxiety symptoms in children and adolescents when reported by children and adolescents themselves. There is also evidence to suggest that the FRIENDS intervention may increase the reduction in anxiety symptoms up to 12 months after the intervention.

### 1.2 | What is this review about?

The intervention is three age-appropriate preventive anxiety programmes: Fun FRIENDS, FRIENDS for Life, and My FRIENDS Youth. Children and young people between 4 and 16 years of age who did not meet diagnostic criteria for an anxiety disorder diagnosis were eligible. The outcome of primary interest is anxiety symptoms.

### 1.3 | What is the aim of this review?

We examine the effects of the FRIENDS preventive programme on anxiety symptoms in children and adolescents.

### 1.4 | What studies are included?

Forty-two studies were identified. Only 28 were assessed to be of sufficient methodological quality and reported usable outcome data to be included in the data synthesis. Studies came from 15 different countries. The timespan of studies was 15 years, from 2001 to 2016. The average number of participants analysed was 240, the average number of controls was 212. The majority of studies were randomised controlled trials.

### 1.5 | What is the effect of the FRIENDS preventive programme on anxiety symptoms in children and adolescents?

Our results indicate that the FRIENDS intervention may reduce anxiety symptoms in children and adolescents when reported by children and adolescents themselves. On average, we found a positive, statistically significant effect in the short-term and at 12 months follow-up.

At follow-up, we found evidence that programmes implemented by mental health providers may perform better than programmes implemented by teachers.

## 1.6 | What do the findings of this review mean?

We believe these positive average effect sizes are of a meaningful magnitude as the intervention was primarily given as a universal intervention, that is, to children and adolescents with no immediate anxiety problems.

The majority of trials employed a wait-list design, implying that only a few studies reported on the long-term effects of the FRIENDS intervention. Half the number of studies reporting post-intervention

The overall certainty of evidence varied from low to very low. Only one trial was rated low risk of bias on almost all domains. There is a need for more rigorously conducted studies.

## 1.7 | How up-to-date is this review?

We searched for studies up to October 2023.

## 2 | SUMMARY OF FINDINGS

### Summary of findings 1

#### Summary of findings

Quality assessment							Summary of findings		
# Effect sizes	Design	Limitations	Directness	Consistency	Precision	Publication bias	# Participants	Effect (95% CI)	Certainty
<i>Anxiety child reported post-intervention</i>									
25	17 CRCTs, 3 RCTs, 5 NRSs	14 high risk of bias, 2 serious risk of bias; very serious (-2) limitation	None	None	None	None	Treated: 6592, Control: 5787	0.13 [0.04, 0.22]	Low
<i>Anxiety parent reported post-intervention</i>									
5	4 CRCTs, 1 RCT	4 high risk of bias; very serious (-2) limitation	None	None	None	None	Treated: 670, Control: 690	-0.08 [-0.20, 0.04]	Low
<i>Anxiety child reported 12 months follow-up</i>									
12	10 CRCTs, 2 NRSs	7 high risk of bias, 1 serious risk of bias; very serious (-2) limitation	None	None	None	None	Treated: 3382, Control: 2493	0.31 [0.13, 0.49]	Low
<i>Anxiety parent reported 12 months follow-up</i>									
3	3 CRCTs	3 high risk of bias; very serious (-2) limitation	None	None	None	None	Treated: 545, Control: 466	-0.10 [-0.26, 0.06]	Low
<i>Self-esteem post-intervention</i>									
4	2 RCTs, 2 NRSs	2 high risk of bias, 1 serious risk of bias; very serious (-2) limitation	None	Important inconsistency (-1)	None	None	Treated: 385, Control: 294	0.20 [-0.20, 0.61]	Very low

effects reported effects at 12 months follow-up, and even fewer reported on follow-up effects beyond 12 months. Our findings suggest that the FRIENDS intervention increases the reduction in anxiety symptoms up to 12 months after the intervention. This may have implications for practice, as there is a need to continuously evaluate the sustainability of treatment results and emphasises the need for future research not using wait-list design, so more follow-up effects are available.

### GRADE Working Group grades of evidence (Guyatt et al., 2013)

- High certainty: further research is very unlikely to change our confidence in the effect estimate.
- Moderate certainty: further research is likely to have an important impact on our confidence in the effect estimate and may change the estimate.

- Low certainty: further research is very likely to have an important impact on our confidence in the effect estimate and may change the estimate.
- Very low certainty: we are very uncertain about the effect estimate.

### 3 | BACKGROUND

#### 3.1 | The problem, condition or issue

An estimated one in eight (12.7%) children and youth aged 4–18 years from high-income countries (as classified by the World Bank Group [2021]) have mental disorders at any given time, causing symptoms and impairment which requires treatment (Barican et al., 2022). Anxiety disorders are amongst the most common psychiatric disorders, occurring in 5.2% of all children and youth aged 4–18 years from high-income countries (Barican et al., 2022).

Every child and adolescent faces normal, developmentally appropriate worries, fears, and shyness. For example, primary school-age children commonly have worries about injury and natural events, whereas older children and adolescents typically have worries and fears related to school performance, social competence, and health issues (Beesdo et al., 2009). Pathological anxiety significantly, however, interferes with a child's ability to handle a wide variety of everyday activities, such as interpersonal relationships, social competence, peer relations and school adjustment. If left untreated, childhood anxiety may develop over the years into a chronic adult anxiety disorder or, in some cases, clinical depression (Barrett & May, 2007). Studies show that most youths who experience psychological distress do not seek professional help (Biddle et al., 2004), and youths with a mental disorder are less likely to use mental health services than adults (Mack et al., 2014). It is estimated that less than 25% of all children and youths with an anxiety disorder receive professional help (Merikangas et al., 2011; Wang et al., 2007). In addition, there is often a delay of 9–23 years from onset to first treatment for the disorder (Wang et al., 2007). For these reasons, it is important to prevent that children and youth with elevated anxiety symptoms move on to fully develop anxiety disorders.

As anxiety, fear, and stress responses are often considered normative experiences, children and adolescents may benefit from anxiety prevention programmes regardless of risk status.

An anxiety prevention programme that is manualised, well-structured, and can be easily integrated into school curriculums is the FRIENDS programme. FRIENDS is based on a firm theoretical model which addresses cognitive, physiological and behavioural processes that are seen to interact in the development, maintenance and experience of anxiety (Barrett & May, 2007). FRIENDS is an acronym for the skills taught throughout the programme: · Feelings. · Remember to Relax. Have quiet time. · I can do it! I can try (Inner helpful thoughts) · Explore Solutions and Coping Step Plans. · Now reward yourself! You've done your best! · Don't forget to practice. · Smile! Stay calm, Stay Strong and talk to your support networks!

In a meta-analysis of anxiety prevention programmes, Fisak et al. (2011) found FRIENDS to be more effective at reducing anxiety

symptoms than other prevention programmes. However, the vast majority of the studies included in the review were performed in Australia; 8 of the 10 evaluations on FRIENDS were performed in Australia. As noted by the review authors, more research is needed to determine the degree to which the effectiveness of the programme is generalisable to nations other than Australia. Since the review of Fisak et al. (2011) was carried out, several trials on the effectiveness on FRIENDS have been carried out in countries other than Australia, and it may now be possible to answer the question on effectiveness more generally.

#### 3.2 | The intervention

The FRIENDS programme is a 10-session manualised cognitive behavioural therapy (CBT) programme which can be used as both prevention and treatment of child and youth anxiety (Barrett et al., 2014).

The FRIENDS protocol has been adapted into three developmentally sensitive programmes:

- Fun FRIENDS (4–7 years).
- FRIENDS for life (8–11 years).
- My FRIENDS Youth (12–16 years).

However, note that these three age-appropriate programmes and their titles are the current versions of the FRIENDS programme. The FRIENDS programme was developed by Dr. Barrett in 1998 (a refinement of the programme 'Coping Koala' to reflect a user-friendly early intervention and prevention format), and was expanded into two parallel age groups – FRIENDS for Children 7–11 years, and FRIENDS for Youth 12–16 years.

A new general title for the programme, 'FRIENDS for Life' was introduced in 2005 (Barrett & May, 2007) and a developmentally tailored, downward extension of the two pre-existing FRIENDS for Life programmes was added (the Fun FRIENDS programme, Pahl & Barrett, 2007). Today the FRIENDS programme is broken up into the three age groups shown above (in addition there is a version aimed at adults aged 16+ which is not included in this review). Each of these developmentally tailored programmes is structured and implemented in the same way (Higgins & O'Sullivan, 2015). FRIENDS is a manual based programme which consists of 10 1-h lessons (although the programme allows for flexible roll-out as long as the sequence of the sessions are maintained) plus two follow-up booster sessions, during which the key cognitions and behaviours associated with anxiety are targeted and addressed. The programmes also involve a parent component which consists of parent psychoeducational sessions where parents are helped to understand anxiety, develop appropriate strategies to deal with their own anxiety, if necessary, and improve their child management and problem-solving skills. Certification is required for all professionals who want to use the FRIENDS programme. Certified professionals further have to be re-certified every third year to ensure that they are updated with the latest developments of the programme. FRIENDS can be run by teachers or mental

health care professionals, and it can be run as a whole class programme, or as a small group intervention.

The intervention of interest is the anxiety preventive programme FRIENDS (the three age-appropriate programmes). The comparison population are children and adolescent who do not participate in FRIENDS programmes.

As recommended by the Institute of Medicine Report (Mrazek & Haggerty, 1994), and the updated report (O'Connell et al., 2009) we will define prevention as those interventions that occur before the onset of a clinically diagnosed disorder. Although the programme has been designed to be effective as both a treatment and a prevention programme (mostly school-based), we will only include preventive programmes.

The type of prevention may be universal or targeted. The Institute of Medicine report published in 1994, categorised prevention programmes based on the population targeted (Mrazek & Haggerty, 1994).

Specifically, universal prevention programmes are applied to the general population, without focusing on the risk status. Targeted programmes can either be selective or indicated. Selective programmes target those (individuals or groups) who are identified as exhibiting an elevated risk for developing a disorder based on established risk factors (e.g., socio-economic status [SES], Barrett et al., 2017), and indicated programmes target those (individuals) who exhibit problematic behaviours predicting a high level of risk or initial symptoms of a disorder but who do not yet meet criteria for the disorder. In general, the type of prevention programme utilised (i.e., universal, selective or indicated) may be a crucial methodological factor associated with programme effectiveness (Donovan & Spence, 2000). All types of preventive programmes are eligible whereas treatment programmes (for those with a diagnosis and in need of treatment) are not eligible.

All types of providers (e.g., teachers, mental health providers) and all types of settings (e.g., school based, community based) will be eligible.

### 3.3 | How the intervention might work

The FRIENDS programmes are a suite of programmes (including Fun FRIENDS, FRIENDS for Life and FRIENDS for Youth), which aim to improve resilience (or coping) skills in children and youth and reduce anxiety and improve mental health and wellbeing.

The programme is based on cognitive behavioural therapy, and positive psychology, and uses a play-based, and experiential learning approach to provide cognitive behavioural skills in a developmentally appropriate manner. During each session children and youth are taught skills, aimed at helping them to increase their coping skills through stories, games, videos, and activities.

The theoretical model for the prevention and early intervention of anxiety, specifically in relation to FRIENDS for Life, is shown in Figure 1 (based on Barrett, 1999; Wigelsworth et al., 2018 and Barrett, 2020). The programme has gone through various updates

since 1998. The figure therefore shows the original content as reported in Barrett (1999) and the additions in *cursive* as reported in Wigelsworth et al. (2018) and Barrett (2020).

The programme addresses child and youth anxiety by focusing on (a) physiological or body reactions: the physical reactions our bodies experience when we are feeling worried, nervous, or afraid; (b) cognitive or 'mind' processes: the inner thoughts we have about ourselves, others, and situations; (c) learning or behavioural skills: the acquisition of new skills to cope with and manage anxiety; (d) attachment: stable, unconditional loving relationships; and (e) values: identify personal qualities important to you. The programme aims to teach coping skills such as understanding and managing emotions to assist children and youth in responding to uncomfortable emotions in appropriate and helpful ways. The coping skills increase the children's (and youth's) resilience and protects them from developing anxiety.

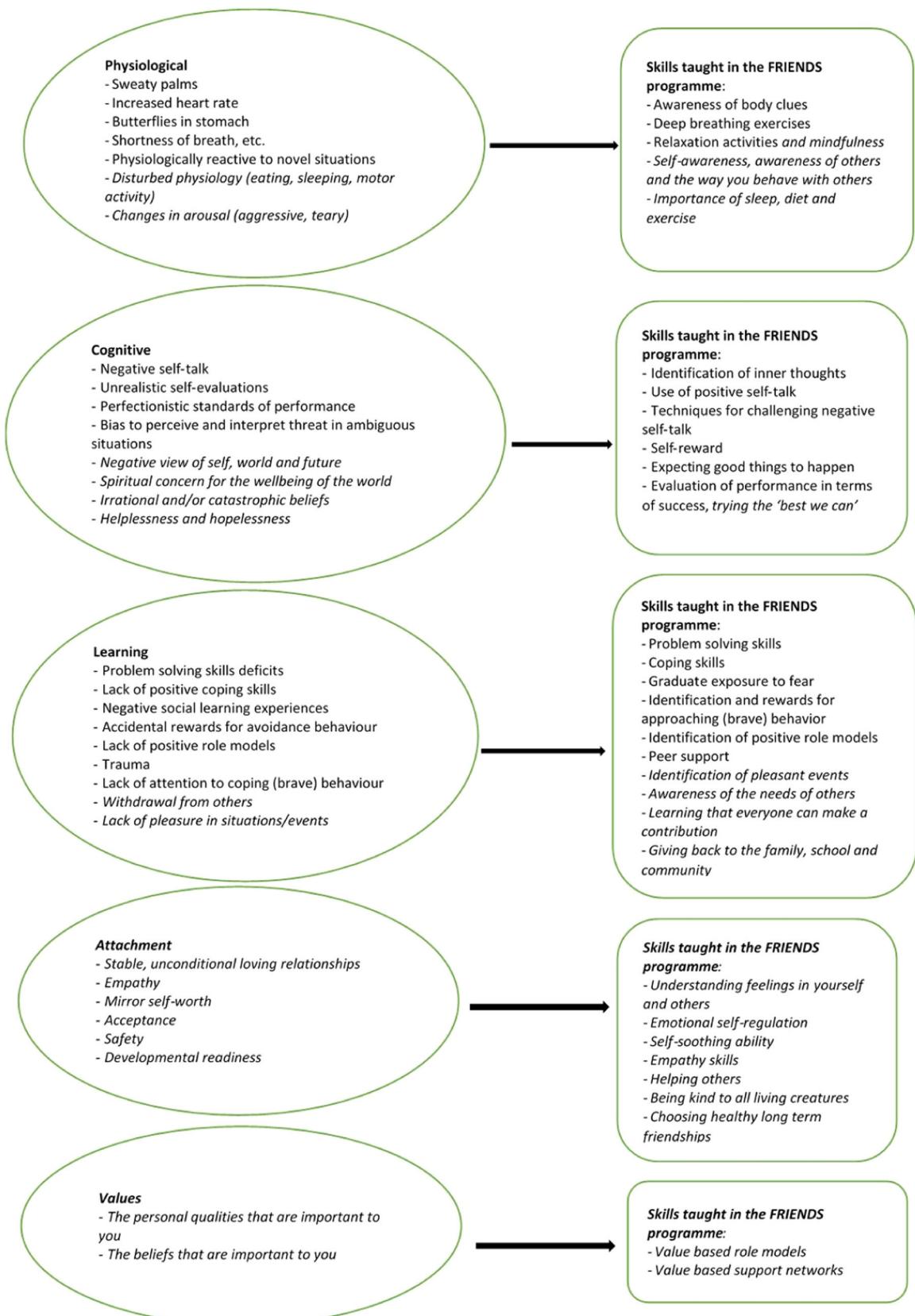
### 3.4 | Why it is important to do this review

Concerns and worries are a normal part of the everyday life of children and adolescents (Weems & Stickle, 2005). Childhood fears such as worries about loss or separation from parents, or worries about personal injury, death, and natural disasters are often a normal part of childhood development (Warren & Sroufe, 2004). However, as Warren and Sroufe (2004) argue, these worries can become problematic if they become persistent, frequent and severe enough to interfere with or limit the child's everyday life and functioning. There are therefore good reasons for increased focus on anxiety prevention over treatment.

Evidence shows that a high proportion of children do not grow out of their anxiety disorders during adolescence and adulthood (Majcher & Pollack, 1996). Anxiety disorders are amongst the most prevalent psychiatric disorders, with a prevalence rate of 10.4% in Western European, North American and Australasian populations (Baxter et al., 2012). According to research by Thompson et al. (2004, 2008) most sufferers of anxiety do not access treatment until well into adulthood and even those who do access appropriate help typically suffer for many years before receiving that help. It is therefore important to prevent that children and youth with elevated anxiety symptoms move on to fully develop anxiety disorders.

## 4 | OBJECTIVES

The main objective of this review is to answer the following research question: What are the effects of the FRIENDS preventive programme on anxiety symptoms in children and adolescents? Further, the review will attempt to answer if the effects differ between participant age groups, participant SES, type of prevention (universal, selective or indicated), type of provider (lay or mental health provider), country of implementation (Australia or other countries) and implementation issues in relation to the booster sessions and parent sessions (implemented, partly implemented or not at all).

**FIGURE 1** How the intervention works.

## 5 | METHODS

### 5.1 | Criteria for considering studies for this review

The project followed standard procedures for conducting systematic reviews using meta-analysis techniques.

#### 5.1.1 | Types of studies

Randomised (and non-randomised) controlled trials were eligible. To summarise what is known about the possible causal effects of programme participation, we included all study designs that used a control group, that is, a group of children/youth not participating in the intervention. The control group may be offered no treatment or treatment as usual.

The study designs we included in the review were:

1. Controlled trials (where all parts of the study are prospective, such as identification of participants, assessment of baseline characteristics, allocation to intervention, which may be randomised or non-randomised, assessment of outcomes and generation of hypotheses [Higgins & Green, 2011]).
2. Non-randomised studies (attendance in programmes has occurred in the course of usual decisions, the allocation to programmes and no programme is not controlled by the researcher, and there is a comparison of two or more groups of participants, that is, at least a treated group and a control group).

Studies using single group pre-post comparisons were not eligible. Non-randomised studies using an instrumental variable approach were not eligible—see Supporting Information S5: Appendix 1 (Justification of exclusion of studies using an instrumental variable (IV) approach) for our rationale for excluding studies of these designs. A further requirement to all types of studies (randomised as well as non-randomised) was that they were able to identify an intervention effect. Studies where, for example, the treatment was offered to children in one unit (e.g., school or class) only and the comparison group was children at another unit (school/class or more schools/classes for that matter) cannot separate the treatment effect from the unit of school/class effect.

#### 5.1.2 | Types of participants

The review included children and adolescents aged 4–16 years who did not meet diagnostic criteria for an anxiety disorder diagnosis. We anticipated that the three age-appropriate programmes were provided to children and adolescents in the corresponding age groups; that is, 'Fun FRIENDS' to 4–7 year olds, 'FRIENDS for Life' to 8–11 year olds and 'My FRIENDS Youth' to 12–16 year olds. Studies including participants out-of-age range were eligible if at least 70% of participants were within the age range corresponding

to the particular programme or results for a discrete age group within the eligible range was provided. If studies included a mix of children and adolescents with and without a clinically diagnosed anxiety disorder, they were included if at least 70% of participants were not diagnosed or results of the eligible subgroup (not diagnosed) was provided.

#### 5.1.3 | Types of interventions

The intervention of interest is the preventive anxiety programme FRIENDS. Prevention is defined as those interventions that occur before the onset of a clinically diagnosed disorder. Treatment programmes (for those with a diagnosis) were not eligible. The type of prevention may be universal (applied to the general population, without focusing on the risk status), selective (target those (individuals or groups) who are identified as exhibiting an elevated risk for developing a disorder based on established risk factors) or indicated (target those (individuals) who exhibit problematic behaviours predicting a high level of risk or initial symptoms of a disorder but who do not yet meet criteria for the disorder).

The three age-appropriate preventive anxiety programmes: Fun FRIENDS, FRIENDS for Life (titled FRIENDS for Children before 2005), and My FRIENDS Youth (titled FRIENDS for Youth before 2005) were eligible.

The comparison population were children and adolescent who did not participate in any of the FRIENDS programmes.

#### 5.1.4 | Types of outcome measures

The intervention is an anxiety prevention programme, and although some studies reported depression outcomes we limited the analysis to anxiety outcomes. As stated in the protocol (Filges, 2023), an analysis on depression outcomes may be biased as not all studies report depression outcomes, and it cannot be ruled out that those who do have results biased towards a positive effect on depression outcomes. Thus, depression outcomes were excluded because they fall outside the scope of both the intervention and the systematic review.

##### *Primary outcomes*

The primary focus was on reduction in anxiety symptoms at all points in time, measured using psychometrically robust measures of anxiety symptoms that yield symptom scores on continuous scales (Myers & Winters, 2002). The measures used in the included studies were:

- Multidimensional Anxiety Scale for Children (MASC) (March et al., 1997).
- Revised Children's Anxiety and Depression Scale (RCADS) –Anxiety Scale (Chorpita et al., 2000).
- Revised Children's Manifest Anxiety Scale (RCMAS) (Reynolds & Richmond, 1985).

- Screen for Child Anxiety Related Emotional Disorders (SCARED) (Birmaher, 1999).
- Spence Children's Anxiety Scale (SCAS) (Spence, 1997).
- The AN-UD Anxiety Scale (Kozina, 2012).
- The Preschool Anxiety Scale, Parent Report (PAS) (Spence, 2001).

These scales were mostly self-reported but multiple reporters (child and parent) were used in a few studies and the studies with preschooler participants used only parent reported measures (PAS).

We analysed reduction in anxiety symptoms separately for (1) self-reported and (2) parent-reported measures. Effect sizes based on all measures reported in included studies were reported for transparency, but only one measure per reporter was included in a particular meta-analysis as, double counting (of participants) increases precision spuriously, which is inappropriate and unnecessary.

We planned to analyse the prevalence of anxiety diagnosis at medium-term follow-up (i.e., between 6 and 12 months) or later but this outcome was not reported in any studies.

#### *Secondary outcomes*

As research shows there is a clear relationship between self-esteem and anxiety (Sowislo 2013), a secondary focus was on improvement in self-esteem, measured using psychometrically robust measures of self-esteem. The measures used in included studies were:

- Culture-Free Self-Esteem Questionnaire (CFSEQ) (Battle, 1992).
- Rosenberg Self Esteem Scale (RSES) (Rosenberg, 1965).
- Self Esteem Inventory (SEI) (Coopersmith, 1989).
- The Coopersmith Self-Esteem Inventory Revised Version (CSEI) (Hills, 2011).

Any adverse effects of interventions were included as an outcome, including a worsening of outcome on any of the included measures.

#### 5.1.5 | Duration of follow-up

Time points analysed were:

- medium-term follow-up (6–12 months);
- long-term follow-up (12–24 months); and
- very long term follow-up (over 24 months).

#### 5.1.6 | Types of settings

All types of settings (e.g., school based, community based) were eligible.

## 5.2 | Search methods for identification of studies

The search was performed by two review authors (KB, TF) of which one (KB) is an information specialist. Relevant studies were identified through electronic searches in bibliographic databases, grey literature repositories and resources, hand search in specific targeted journals, citation tracking, contact to international experts and Internet search engines. As the programme was developed in 1998, a date restriction of 1998 and onwards was applied.

### 5.2.1 | Electronic searches

The following electronic bibliographic databases were searched:

- ERIC (EBSCO) – September 2023.
- Teacher Reference Center – September 2023.
- Academic Search (EBSCO) – September 2023.
- MEDLINE (PubMed) – September 2023.
- Embase (OVID) – September 2023.
- CINAHL (EBSCO) – September 2023.
- Cochrane Library (Cochrane Reviews & Cochrane Central) – September 2023.
- PsycINFO (EBSCO) – September 2023.
- APA PsycNet – September 2023.
- Socindex (EBSCO) – September 2023.
- International Bibliography of the Social Sciences (ProQuest) – September 2023.
- Sociological Abstracts (ProQuest) – September 2023.
- Science Citation Index Expanded (Web Of Science) – September 2023 Social.
- Sciences Citation Index (Web Of Science) – September 2023.

The database searches were performed 29 September 2023.

#### *Description of the search-string*

The search string is based on the PICO(s)-model, and contains two concepts, of which we developed two corresponding search facets: population characteristics and the intervention. The search string includes searches in title and abstract as well as subject terms and/or keywords for each facet. The subject terms in the facets were selected according to the thesaurus or index of each database. Keywords were supplied where the search technique provided additional results. Use of truncation and wildcards were used to address English spelling variants.

#### *Example of a search-string*

Below is an exemplified search string utilised to search ERIC through the EBSCO search interface and exemplifies the search facets as they were searched:

**ERIC. Search performed 9/26/2023. Search modes – Boolean/Phrase. Interface – EBSCOhost Research Databases. Search Screen – Advanced Search.**

#	Query	Limiters/Expanders	Results
S13	S4 AND S5 AND S12	Limiters – Date Published: 19980101-20230931 329 Expanders – Apply equivalent subjects Search modes Boolean/Phrase	
S12	S6 OR S7 OR S8 OR S9 OR S10 OR S11	Expanders – Apply equivalent subjects Search modes Boolean/Phrase	1,005,926
S11	Program* OR 'prevent'* OR intervention OR skill* OR evaluation	Expanders – Apply equivalent subjects Search modes Boolean/Phrase	1,000,864
S10	DE 'Health Promotion' OR DE 'Preventive Medicine' OR DE Intervention	Expanders – Apply equivalent subjects Search modes Boolean/Phrase	62,211
S9	DE 'Prevention' OR DE 'Dropout Prevention'	Expanders – Apply equivalent subjects Search modes Boolean/Phrase	22,007
S8	DE 'Program Effectiveness' OR DE 'Program evaluation' OR DE 'Program Development' OR DE 'Programs'	Expanders – Apply equivalent subjects Search modes Boolean/Phrase	143,743
S7	cognit* N3 therap*	Expanders – Apply equivalent subjects Search modes Boolean/Phrase	1,588
S6	DE 'Behavior Modification' OR DE 'Social Reinforcement'	Expanders – Apply equivalent subjects Search modes Boolean/Phrase	11,816
S5	FRIENDS	Expanders – Apply equivalent subjects Search modes Boolean/Phrase	12,163
S4	S1 OR S2 OR S3	Expanders – Apply equivalent subjects Search modes Boolean/Phrase	65,793
S3	SU (anxi* or panic* or phobia* or phobic* or sociophobi* or socio-phobi* or GAD or 'self esteem' or 'self concept' or 'self-esteem' or 'self-concept' or 'self image' or 'self-image' or 'emotional health')	Expanders – Apply equivalent subjects Search modes Boolean/Phrase	65,097
S2	TI (anxi* or panic* or phobia* or phobic* or sociophobi* or socio-phobi* or GAD or 'self esteem' or 'self concept' or 'self-esteem' or 'self-concept' or 'self image' or 'self-image' or 'emotional health')	Expanders – Apply equivalent subjects Search modes Boolean/Phrase	12,320
S1	DE 'Anxiety Disorders' OR DE 'Anxiety' OR DE 'Separation Anxiety' OR DE 'Test Anxiety' OR DE 'Self Esteem' OR DE 'Self Concept'	Expanders – Apply equivalent subjects Search modes Boolean/Phrase	62,563

## 5.2.2 | Searching other resources

### Hand-Search

We conducted a hand search of specific journals, to make sure that all relevant articles were found. The hand search focused on editions published between 2021 and 2023 to secure recently published articles which have not yet been indexed in the bibliographic databases. Based on the identified records from the electronic searches we hand searched the following journals in the time period between 2021 and October 2023:

- *British Journal of Clinical Psychology*
- *Behaviour Change*
- *Journal of Clinical Child and Adolescent Psychology*
- *Clinical Child Psychology and Psychiatry*
- *Journal of Primary Prevention*
- *Child and Adolescent Mental Health*
- *European Journal of Child and Adolescent Psychiatry Advances in School Mental Health Promotion*

### Grey literature searches

The searches on other resources and for unpublished literature were done between 13 October 2023 and 24 October 2023.

We searched for such references as dissertations, working papers and conference proceedings, reports, and EGM's or systematic reviews. Most of the resources searched included multiple types of references, both published and unpublished. In general, there was a great amount of overlap between the types of references in the chosen resources. The resources are listed once under the category of literature we expect to be most prevalent in the resource, even though multiple types of unpublished/published literature might be identified in the resource. Terms used to search other resources were based on the general search strategy. All of these searches can be seen in Supporting Information S5: Appendix 2.

### Artificial intelligence search for references on the Internet

- [Elicit.org](https://elicit.org)

## Dissertations

We searched the following resources for dissertations:

- Open Access Theses and Dissertations
- EBSCO Open Dissertations (EBSCO-host)

## Working papers and conference proceedings

We searched the following resources for working papers/conference proceedings:

- Google Scholar: <https://scholar.google.com/>
- Social Science Research Network: <https://www.ssrn.com/index.cfm/en/Conference>
- Proceedings Citation Index

For each resource, we screened the first 100 hits.

## Evidence and gap maps and systematic reviews

We searched the following resources:

- Campbell Systematic Reviews Journal: <https://onlinelibrary.wiley.com/journal/18911803?af=R>
- EPPI-Centre publications: <https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=116>
- PROSPERO: <https://www.crd.york.ac.uk/prospero/> Epistemonikos: <https://www.epistemonikos.org/>

Relevant systematic reviews or EGMs were used for citation-tracking.

## Trial registries

- CENTRAL Trials Register within the Cochrane Library: <https://www.cochranelibrary.com/central> (includes ClinicalTrials.gov: <https://clinicaltrials.gov/> and WHO International Clinical Trials Registry Platform: <https://www.who.int/clinical-trials-registry-platform>).

## Citation-tracking

To identify both published studies and grey literature, we utilised citation-tracking/snowballing strategies. Our primary strategy was to citation-track related systematic-reviews and meta-analyses. The review team also checked reference lists of included primary studies for new leads.

## Contact to experts

By e-mail during October 2023, we contacted international experts, identified during the literature search, to identify unpublished and ongoing studies.

## Other criteria

Studies were not excluded based on publication status or language (although the ability to assess the relevance of studies was limited by the language skills in the review team). Studies authored before 1998 were not eligible.

## 5.3 | Data collection and analysis

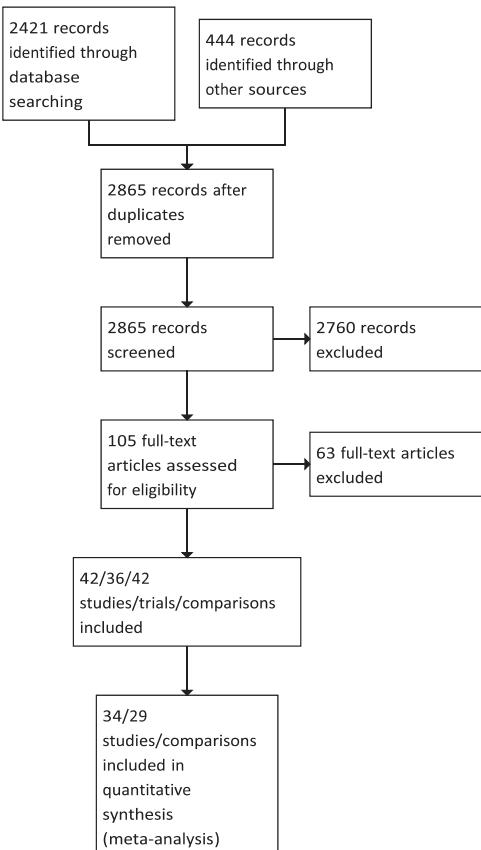
### 5.3.1 | Selection of studies

Two review authors (TF, MWK) independently screened titles and abstracts to exclude studies that were clearly irrelevant. Studies considered eligible by at least one author, or studies where there was insufficient information in the title and abstract to judge eligibility, were retrieved in full text. The full texts were then screened independently by two review authors (TF, MWK). Any disagreement of eligibility were resolved by discussion. Exclusion reasons for studies that otherwise might be expected to be eligible are documented.

The study inclusion criteria were piloted (see Supporting Information S5: Appendix 3). The overall search and screening process is illustrated in Figure 2. None of the review authors were blind to the authors, institutions, or the journals responsible for the publication of the articles.

### 5.3.2 | Data extraction and management

In pairs of two, review authors independently coded and extracted data from all included studies. A coding sheet was piloted on several studies and revised as necessary (see Supporting Information S5: Appendix 4). Disagreements were minor and resolved by discussion.



**FIGURE 2** Flow diagram.

Data and information was extracted on: available characteristics of participants, intervention characteristics and control conditions, research design, sample size, risk of bias and potential confounding factors, outcomes, and results. Analysis was conducted using RevMan Web and Stata software. Extracted numerical and descriptive data, and the risk of bias assessments described in the next section, can be found in the supplementary documents.

### 5.3.3 | Assessment of risk of bias in included studies

We assessed the risk of bias in randomised studies using Cochrane's revised risk of bias tool, ROB 2 (Higgins et al., 2019).

The tool is structured into five domains, each with a set of signalling questions to be answered for a specific outcome. The five domains cover all types of bias that can affect results of randomised trials.

The five domains for individually randomised trials are:

- (1) bias arising from the randomisation process;
- (2) bias due to deviations from intended interventions (separate signalling questions for effect of assignment and adhering to intervention);
- (3) bias due to missing outcome data;
- (4) bias in measurement of the outcome;
- (5) bias in selection of the reported result.

For cluster-randomised trials, an additional domain is included (1b) Bias arising from identification or recruitment of individual participants within clusters). We used the latest template for completion (currently it is the version of 15 March 2019 for individually randomised parallel-group trials and 20 October 2016 for cluster randomised parallel-group trials). In the cluster randomised template however, only the risk of bias due to deviation from the intended intervention (effect of assignment to intervention; intention to treat ITT) is present, and the signalling question concerning the appropriateness of the analysis used to estimate the effect is missing. Therefore, for cluster randomised trials we only used the signalling questions concerning the bias arising from identification or recruitment of individual participants within clusters from the template for cluster randomised parallel-group trials; otherwise we used the template and signalling questions for individually randomised parallel-group trials.

We assessed the risk of bias in non-randomised studies, using the model ROBINS-I, developed by members of the Cochrane Bias Methods Group and the Cochrane Non-Randomised Studies Methods Group (Sterne, Hernán, et al., 2016a). We used the latest template for completion (currently it is the version of 19 September 2016).

The ROBINS-I tool is based on the Cochrane RoB2 tool for randomised trials, which was launched in 2008 and modified in 2011 (Higgins et al., 2011).

The ROBINS-I tool covers seven domains (each with a set of signalling questions to be answered for a specific outcome) through which bias might be introduced into non-randomised studies:

- (1) bias due to confounding.
- (2) bias in selection of participants.
- (3) bias in classification of interventions.
- (4) bias due to deviations from intended interventions;
- (5) bias due to missing outcome data;
- (6) bias in measurement of the outcome;
- (7) bias in selection of the reported result.

The first two domains address issues before the start of the interventions and the third domain addresses classification of the interventions themselves. The last four domains address issues after the start of interventions, and there is substantial overlap of these four domains between bias in randomised studies and bias in non-randomised study trials (although signalling questions are somewhat different in several places, see Sterne, Higgins, et al., 2016b and Higgins et al., 2019).

Randomised study outcomes were rated on a 'Low/Some concerns/High' scale on each domain; whereas non-randomised study outcomes were rated on a 'Low/Moderate/Serious/Critical/No Information' scale on each domain. The level 'Critical' means: the study (outcome) is too problematic in this domain to provide any useful evidence on the effects of intervention, and it is excluded from the data synthesis. The same critical level of risk of bias (excluding the result from the data synthesis) is not directly present in the RoB 2 tool, according to the guidance to the tool (Higgins et al., 2019).

In the case of an RCT, where there is evidence that the randomisation has gone wrong or is no longer valid, we assessed the risk of bias of the outcome measures using ROBINS-I instead of ROB 2. One trial (Lawson, 2023) was moved from CRCT ROB 2 tool to ROBINS-I due to the unfortunate premature stop in March of 2020 due to the COVID-19 pandemic and associated school closures.

We stopped the assessment of a non-randomised study outcome as soon as one domain in the ROBINS-I was judged as 'Critical'.

#### Confounding

An important part of the risk of bias assessment of non-randomised studies is consideration of how the studies deal with confounding factors. Systematic baseline differences between groups can compromise comparability between groups. Baseline differences can be observable (e.g., age and gender) and unobservable (to the researcher; e.g., motivation and ability). There is no single non-randomised study design that always solves the selection problem. Different designs represent different approaches to dealing with selection problems under different assumptions, and consequently require different types of data. There can be particularly great variations in how different designs deal with selection on unobservables. The 'adequate' method depends on the model generating participation, that is, assumptions about the nature of the process by which participants are selected into a programme.

As there is no universal correct way to construct counterfactuals for non-randomised designs, we looked for evidence that identification was achieved, and that the authors of the primary studies justified their choice of method in a convincing manner by discussing

the assumption(s) leading to identification (the assumption(s) that make it possible to identify the counterfactual). Preferably the authors should make an effort to justify their choice of method and convince the reader that the only difference between a treated individual and a non-treated individual is the treatment. The judgement is reflected in the assessment of the confounder 'unobservables' in the list of confounders considered important at the outset (see Supporting Information S5: Appendix 5).

In addition to unobservables, we had (at the protocol stage) identified the following observable confounding factors to be most relevant: age, gender, SES and anxiety symptoms at baseline. In each study, we assessed whether these indicators had been considered, and in addition we assessed other factors likely to be a source of confounding within the individual included studies.

#### *Importance of pre-specified confounding factors*

The motivation for focusing on age, gender, SES and anxiety symptoms at baseline is given below.

The prevalence of different types of psychological problems, coping skills, cognitive and emotional ability vary throughout a child's development through puberty and into adulthood (Cole et al., 2005), and therefore we consider age to be a potential confounding factor.

Furthermore, there are substantial (although inconsistent) gender differences in fear reporting, coping and risk of different types of anxiety disorders, which is why we also include gender as a potential confounding factor (Dalsgaard et al., 2020; Hampel & Petermann, 2005; McLean & Anderson, 2009).

Low childhood SES, in particular financial hardship, is associated with increased exposure to a range of childhood adversities (CAs) such as parental psychopathology, maltreatment, and family violence. Exposure to CAs as well as financial hardship have been associated with increased levels of anxiety in children and onset of anxiety disorders in childhood (Green et al., 2010; McLaughlin et al., 2011).

Finally, pre-treatment group equivalence of anxiety symptoms is indisputably an important confounder. Therefore, the accuracy of the estimated effects of FRIENDS programmes will depend crucially on how well pre-treatment anxiety symptoms are controlled for.

#### *Effect of primary interest and important co-interventions*

We were mainly interested in the effect of starting and adhering to the intended intervention, that is, the treatment on the treated effect. The risk of bias was therefore assessed in relation to this specific effect.

The risk of bias assessments of both randomised trials and non-randomised studies considered adherence and differences in additional interventions ('co-interventions') between intervention groups. Relevant cointerventions are those that individuals might receive with or after starting the intervention of interest, and that are both related to the intervention received and prognostic for the outcome of interest. Important co-interventions we considered were any kind of mental health treatments delivered on an individual basis.

#### *Assessment*

In pairs of two, review authors independently assessed the risk of bias for each relevant outcome in the included studies. We discussed all initial disagreements and were able to reach a consensus in all cases. We report the risk of bias assessment in risk of bias tables for each included study outcome in a supplementary document.

#### 5.3.4 | Measures of treatment effect

Effect sizes that could not be pooled (were reported in a single study only) were reported in as much detail as possible. Software for storing data (including calculating effect sizes and standard errors) and statistical analyses were RevMan 5.0 and Excel.

#### *Continuous outcomes*

We calculated effect sizes with 95% confidence intervals (CIs) where means and standard deviations were available; or alternatively from mean differences and standard deviations of the mean or reported effect sizes and 95% CIs (whichever were available), using the methods suggested by Lipsey and Wilson (2001). If insufficient information was yielded, the review authors requested this information from the principal investigators. Hedges'  $g$  was used for estimating standardised mean differences (SMDs). Any measures of anxiety and self-esteem outcomes, were continuous outcomes in this review.

#### *Dichotomous outcomes*

There were no dichotomous outcomes.

#### 5.3.5 | Unit of analysis issues

We checked for consistency in the unit of allocation and the unit of analysis, as statistical analysis errors can occur when they are different.

We adjusted the effect size and its standard error using the methods suggested by Hedges (2007), using an ICC of 0.02 (we searched the literature for estimates of relevant ICCs; Campbell, 2000, Connolly, 2018, Health Services Research Unit, 2004, Parker, 2021, Stallard, 2012), and assumed equal cluster sizes. To calculate an average cluster size, we divided the total sample size in the study by the number of clusters. In the sensitivity analysis, we report both results using unadjusted effect sizes and using a substantially higher ICC (0.1) than in the primary analysis.

We performed analyses separated by the time points stated in Section 5.1.5.

#### 5.3.6 | Criteria for determination of independent findings

To determine the independence of results in included studies, we considered whether individuals had undergone multiple interventions, whether there were multiple treatment groups, whether several

studies were based on the same data source and whether studies reported multiple conceptually similar outcomes.

#### *Multiple interventions groups and multiple interventions per individuals*

A few studies reported on multiple trials or multiple intervention and/or control groups. To avoid problems with dependence between effect sizes we applied robust standard errors (Hedges et al., 2010) and used the small sample adjustment to the estimator itself (Tipton, 2015). We used the results in Tanner-Smith and Tipton (2014) and Tipton (2015) to evaluate if there were enough studies for this method to consistently estimate the standard errors. When there were not enough studies, we used a synthetic effect size (the average) to avoid dependence between effect sizes. See Section 5.3.10 below for more details about the data synthesis.

#### *Multiple studies using the same sample of data*

A few trials were reported in multiple studies. We reviewed all studies but included only one estimate of the effect from each trial in a particular meta analysis (outcome metric and time point).

#### *Multiple time points*

When the results were measured at multiple points in time, each outcome at each time point was analysed in a separate meta-analysis with other comparable studies reporting measurements at a similar time point. They were grouped together as stated in Section 5.1.5.

#### *Multiple conceptually similar outcomes*

Meta-analysis of outcomes was conducted on each metric (as outlined in Section 5.1.4) separately. A few studies reported multiple estimates of effects regarding the same/similar outcome (anxiety symptoms measured with both the SCAS and the RCMAS measures). We extracted (and report) all outcomes, but included only one measure in the meta-analysis. We included the most frequently used measure in the analysis.

Multiple reporters were used in a few studies; we analysed reduction in anxiety symptoms separately for (1) self-reported and (2) parent-reported measures.

### 5.3.7 | Dealing with missing data

Missing data and attrition rates were assessed in the included studies; see Section 5.3.3. Where studies had missing summary data, such as missing standard deviations or numbers of participants, we requested this information from the principal investigators. We only received an answer in a few instances. The study results from these trials were reported in as much detail as possible.

### 5.3.8 | Assessment of heterogeneity

Heterogeneity amongst primary outcome studies were assessed with Chi-squared ( $\chi^2$ ) test, and the  $I^2$ , and  $\tau^2$  statistics (Higgins et al., 2003).

The estimation of  $\tau^2$  was the DerSimonian and Laird estimate (DerSimonian & Laird, 1986). Any interpretation of the Chi-squared test was made cautiously on account of its low statistical power.

### 5.3.9 | Assessment of reporting biases

Reporting bias refers to both publication bias and selective reporting of outcome data and results. Here, we state how we planned to assess publication bias.

Asymmetric funnel plots are not necessarily caused by publication bias (and publication bias does not necessarily cause asymmetry in a funnel plot). In general, asymmetry is a sign of small-study effects, of which there can be many causes beside publication bias (Sterne et al., 2005).

Instead of trying to interpret the funnel plots as direct evidence of publication bias, or the lack thereof, we performed sensitivity analyses for publication bias in meta-analyses as suggested by Mathur and VanderWeele (2020). This method gives a value of how large ratios of publication probabilities (i.e., the likelihood of affirmative results to be published relative to non-affirmative results) would have to be to alter the results and therefore indicate how robust the meta-analysis is to publication bias.

We further used significance funnel plots for information about possible publication bias.

### 5.3.10 | Data synthesis

Meta-analysis of outcomes was conducted on each metric (as outlined in Section 5.1.4) separately, and we performed separate analyses for post-intervention, short-term, medium term, and long-term outcomes. Studies that were coded with a critical risk of bias were not included in the data synthesis.

As the intervention dealt with diverse populations of participants (from different countries, facing different life circumstances, etc.), and we therefore expected heterogeneity amongst primary study outcomes, all analyses of the overall effect were inverse variance weighted random-effects statistical models.

Not all effect estimates were independent. One study reported on two trials; three (groups of) studies had either multiple intervention or multiple control groups; one study reported results on different age groups within the same trial; and two (groups of) studies most likely shared intervention and control schools. To take these dependences into account, we applied the robust variance estimation (RVE) approach (Hedges et al., 2010). An important feature of this analysis is that the results are valid regardless of the weights used. For efficiency purposes, we calculated the weights using a method proposed by Hedges et al. (2010). This method assumes a simple random-effects model in which study average effect sizes vary across studies ( $\tau^2$ ) and the effect sizes within each study are equi correlated ( $\rho$ ). The method is approximately efficient, since it uses approximate inverse-variance weights: they are approximate given that  $\rho$  is, in fact,

unknown and the correlation structure may be more complex. We calculated weights using estimates of  $\tau^2$ , setting  $p = 0.80$  and conducted sensitivity tests using a variety of  $p$  values, to assess if the general results and estimates of the heterogeneity were robust to the choice of  $p$ . We used the small sample adjustment to the residuals used in RVE as proposed by Bell and McCaffrey (2002) and extended by McCaffrey et al. (2001) and by Tipton (2015). We used the Satterthwaite degrees of freedom (Satterthwaite, 1946) for tests as proposed by Bell and McCaffrey (2002) and extended by Tipton (2015). We used the guidelines provided in Tanner-Smith and Tipton (2014) to evaluate if there were enough studies for this method to consistently estimate the standard errors in each particular meta-analysis. In a few meta-analyses there was not a sufficient number of studies to use RVE, and we conducted the data synthesis using a synthetic effect size (the average) to avoid dependence between effect sizes.

Random effects weighted mean effect sizes were calculated using 95% CIs, and we provide graphical displays (forest plots) of effect sizes.

In addition to 95% CIs, we reported 95% prediction intervals.

All meta-analyses were carried out in STATA and Revman 5.4.

### 5.3.11 | Moderator analysis and investigation of heterogeneity

We planned to investigate the following factors with the aim of explaining observed heterogeneity:

Type of programme, that is, whether it is a universal, indicated or selective intervention; and the three different age appropriate programmes: 'Fun FRIENDS' (4–7 year olds), 'FRIENDS for Life' (8–11 year olds) and 'My FRIENDS Youth' (12–16 year olds), or the corresponding before 2005 versions 'FRIENDS for Children' (7–11 year olds), and 'FRIENDS for Youth' (12–16 year-olds). Other study-level summaries of participant characteristics (e.g., studies considering a specific gender or studies where separate effects for girls/boys are available) and SES indicator (e.g., studies considering a specific SES indicator or studies where separate effects for low/high SES are available). In addition, we investigated other programme characteristics such as type of provider (lay/teacher or mental health provider), country of implementation (Australia/other countries) and implementation issues in relation to the booster sessions and parent sessions (implemented, partly implemented or not at all) and other implementation issues.

In meta-analyses where the number of included studies was sufficient and where there was variation in the covariates, we performed moderator analyses (multiple meta-regression using the mixed model) to explore how observed variables were related to heterogeneity.

We applied the RVE approach when there was enough studies and used approximately inverse variance weights calculated using a method proposed by Hedges et al. (2010). This technique calculates standard errors using an empirical estimate of the variance: it does

not require any assumptions regarding the distribution of the effect size estimates. The assumptions that are required to meet the regularity conditions are minimal and generally met in practice. This more robust technique is beneficial because it takes into account the possible correlation between effect sizes separated by the covariates within the same study (e.g., age or gender-separated effects) and allows all the effect size estimates to be included in meta-regression. We calculated weights using estimates of  $\tau^2$ , setting  $p = 0.80$  and conducted sensitivity tests using a variety of  $p$  values, to assess if the general results were robust to the choice of  $p$ . We used the small sample adjustment to the residuals used in RVE and the Satterthwaite degrees of freedom (Satterthwaite, 1946) for tests (Tipton, 2015). The results in Tipton (2015) suggest that the degrees of freedom depend on not only the number of studies, but also on the type of covariates included in the meta-regression. The degrees of freedom can be small, even when the number of studies is large if a covariate is highly unbalanced or a covariate with very high leverage is included. The degrees of freedom varied from coefficient to coefficient. The corrections to the degrees of freedom enable us to assess when the RVE method performs well. As suggested by Tanner-Smith and Tipton (2014) and Tipton (2015) if the degrees of freedom are smaller than four, the RVE results should not be trusted.

We reported 95% CIs for regression parameters. We estimated the correlations between the covariates and considered the possibility of confounding. Conclusions from meta-regression analysis was cautiously drawn and not solely based on significance tests. The magnitude of the coefficients and width of the CIs was taken into account as well.

If there were not enough studies, single factor subgroup analysis was performed. The assessment of any difference between subgroups was based on 95% CIs. Interpretation of relationships was cautious, as they were based on subdivision of studies, and indirect comparisons.

In general, the strength of inference regarding differences in treatment effects amongst subgroups is controversial. However, making inferences about different effect sizes amongst subgroups on the basis of between-study differences entails a higher risk compared to inferences made on the basis of within study differences (see Schandlmaier et al., 2020). However, there were no within study differences to be used.

### 5.3.12 | Sensitivity analysis

Sensitivity analysis was carried out where possible, by restricting the meta-analysis to a subset of all studies included in the original meta-analysis. We considered sensitivity analysis with regard to research design and restricted the analysis to studies using randomisation. We further considered sensitivity analyses for each domain of the risk of bias checklists and restricted the analysis to studies with a low risk of bias. Sensitivity analysis was only conducted when there were more than two studies left in the analyses. We tested sensitivity to

clustered delivery of treatment by reporting results using both unadjusted effect sizes and using a substantially higher ICC (0.3) than in the primary analysis.

Finally, we conducted sensitivity tests using a variety of  $\rho$  values, to assess if the general results and estimates of the heterogeneity were robust to the choice of  $\rho$  in the analyses using RVE.

### 5.3.13 | Treatment of qualitative research

We did not include qualitative research.

### 5.3.14 | Summary of findings and assessment of the certainty of the evidence

The GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) system was used to assess the certainty of the body of evidence as it relates to the studies that contribute data to the meta-analyses for the pre-specified outcomes (Guyatt et al., 2008). The system classifies certainty of evidence as high, moderate, low, or very low (Guyatt et al., 2013). Five of the eight criteria proposed in the GRADE method have the potential to decrease one's confidence in the correctness of the effect estimates: risk of bias, inconsistency of results across studies, indirectness of evidence, imprecision, and publication bias. Three further criteria are proposed that have the potential to increase this confidence: a large magnitude of effect with no plausible confounders, a dose-response gradient, and a conclusion that all plausible residual confounding would further support inferences regarding treatment effect. GRADE proposes that these three criteria should be considered particularly in observational studies (Guyatt et al., 2011). However, none of these criteria for upgrading the evidence was relevant. We justified all decisions to downgrade the certainty of outcomes, and made comments to aid readers' understanding of the review where necessary.

The outcomes were graded as follows.

- High certainty: further research is very unlikely to change our confidence in the effect estimate.
- Moderate certainty: further research is likely to have an important impact on our confidence in the effect estimate and may change the estimate.
- Low certainty: further research is very likely to have an important impact on our confidence in the effect estimate and may change the estimate.
- Very low certainty: we are very uncertain about the effect estimate.

We produced a summary of findings table, presenting the overall certainty of the body of evidence according to GRADE criteria for mental health outcomes.

## 6 | RESULTS

### 6.1 | Description of studies

#### 6.1.1 | Results of the search

We summarise the search results in a flow chart in Figure 2. The total number of potentially relevant studies was 2865 after excluding duplicates (database: 2421, grey, snowballing and other resources: 444). We screened all studies based on title and abstract; 2760 were excluded for not fulfilling the screening criteria, and 105 studies were ordered, retrieved, and screened in full text. Of these, 63 did not fulfil the screening criteria and were excluded. We included a total of 42 studies in the review. The references are listed in the section *References to included studies*.

#### 6.1.2 | Included studies

The 42 included studies, reported on a total of 36 trials containing 42 comparisons. Unfortunately, two studies (Ghajareih, 2018; Haidari, 2016) with a title and an abstract in English indicating they were eligible, could not be assessed as, except for the title and abstract, they were written in Persian.

The remaining 40 studies reported on a total of 34 trials containing a total of 40 comparisons. Some trials were reported in multiple studies and some studies had multiple comparisons.

One study reported outcomes at follow-up only: Ahlen (2019) reported follow-up outcomes to Ahlen (2018).

Four trials were reported in multiple studies (same outcomes and time points): the same outcomes from the same trial was reported in both Gallegos (2008) and Gallegos (2013); the three studies Lowry-Webster (2001), Lowry-Webster (2002) and Lowry-Webster (2003) reported on the same trial; Pahl (2009) and Pahl (2010) reported on the same trial; the three studies Skryabina (2016), Stallard (2014) and Stallard (2015) reported on the same trial.

One study, Miller (2011a), reported on two trials.

Three (groups of) studies had either multiple-interventions or multiple control groups: Anticich (2013) had both a wait-list control and an active control; Barrett and Turner (2001) reported on a teacher-led intervention and on a psychologist-led intervention; Stallard (2014) and Stallard (2015) reported on both a teacher led and a health led intervention.

Two studies reported on different age groups: Barrett (2003), reported results on FRIENDS for Children (children) and FRIENDS for Youth (youth) separately within the same trial and Sabey (2019) reported on FunFriends (pre-schoolers) and FRIENDS for Life (children) separately within the same trial.

Three (groups of) studies most likely shared intervention and control schools and probably partly reported on the same children: the studies Lowry-Webster (2001/2002/2003) and Barrett and Turner (2001) shared three of the control schools and most likely some of the intervention schools too (it is unclear which schools were allocated to

intervention and when). One study (Barrett & Turner, 2001) analysed grade 6 students and the other (Lowry-Webster, 2001/2002/2003) analysed grade 5–7 students. We have e-mailed (August 2023) the first author of Barrett and Turner (2001), asking for an explanation (see the risk of bias tables provided in the supplementary documents) but have received no answer. Finally, in Barrett (2005) it was reported that they used part of the same data as Barrett and Turner (2001) (p. 543): ‘This study was designed to examine the effects of a universal schoolbased cognitive behavioural intervention for child anxiety at two developmental levels. Utilising data from our previous study...’. Thus, Barrett (2005) and Barrett and Turner (2001) probably analyse the same grade 6 students as Barrett (2005) analyse grade 6 and 9 students, whereas Barrett and Turner (2001) analyse grade 6 students only. As there is data overlap with both Barrett and Turner (2001) and Lowry-Webster (2001/2002/2003) and further, there are problems with numbers that do not match (see the risk of bias assessment in the supplementary document), leading us to chose not to use this study (Barrett, 2005) in the data synthesis.

In the following we will refer to number of comparisons and not number of studies, as comparisons potentially available for analysis is the relevant measure and not the number of studies reporting on the comparisons.

Thirty-one comparisons were obtained using randomised designs, 7 of them individual randomised and 24 cluster randomised. The remaining nine were non-randomised comparisons. The participants were from 15 different countries (Australia, Brasil, Canada, Germany, Hong Kong, Iran, Ireland, Lebanon, Mexico, the Netherlands, Scotland, Slovenia, Sweden, UK and USA). Descriptions of the intervention and control conditions of each comparison were extracted in as much detail as possible and can be found in the supplementary descriptive table.

In Table 1, we show the total number of comparisons from studies that met the inclusion criteria for this review. The first column shows the total number of comparisons grouped by country of origin. The second column shows the number of these comparisons where not enough data was provided to calculate an effect estimate (and requests to provide the data from the study authors failed). The third column gives the number of comparisons that were coded with Critical risk of bias. The fourth column gives the number of comparisons not used due to (partly) overlap of data. The last column gives the total number of comparisons used in the data synthesis.

Four comparisons could not be used in the data synthesis as all reported outcomes were judged to have a Critical risk of bias (see supplementary documents for the detailed risk of bias assessments). In accordance with the protocol, we excluded studies with an overall rating ‘Critical risk of bias’ from the data synthesis on the basis that they would be more likely to mislead than inform. Six comparisons were from studies that did not report data in a form that enabled the calculation of effect sizes and standard errors, and our attempts to request the data from study authors produced no answers, and finally, two comparisons were not used as there was partial overlap of data to other comparisons used. All comparisons are listed in Table 2 along with the reason why the comparison was not used in the data synthesis. A total of 28 comparisons were available for data synthesis.

**TABLE 1** Number of included comparisons by country.

Country	Total	Reduction due to			Used in data synthesis
		Critical risk of bias	Missing data	Overlap of data	
Australia	13	2	1	2	8
Brasil	1				1
Canada	5				5
Germany	1				1
Hong Kong	3		2		1
Iran	3	1	2		0
Ireland	2				2
Lebanon	1				1
Mexico	1				1
NL	1				1
Scotland	1				1
Slovenia	2	1			1
Sweden	1				1
UK	3				3
USA	2		1		1
Total	40	4	6	2	28

The main characteristics of the 28 comparisons used in the data synthesis are shown in Table 3. The majority of studies did not report the start year of the interventions analysed. Of those who did, the timespan of intervention start was 15 years, from 2001 to 2016 and, on average, the intervention start year was 2010. The average number of treated participants analysed was 240, ranging from 8 to 1476 per comparison. The average number of controls was 212, ranging from 10 to 1534 per comparison. The average age of treated participants was 10.4 years, ranging from 4.6 to 15.8 years. On average, females constituted more than half of treated participants, 54%, ranging from 43% to 100%. SES of participants was reported for only 11 comparisons and 3 (27%) reported they were low SES. The average child reported pre-anxiety scores, calculated as percent of the maximum scale because different scales were used, was 30.9, ranging from 7.8 to 59.7. The majority of programmes were FRIENDS for Life (64%) followed by FRIENDS for youth (25%), two comparisons considered Fun FRIENDS (7%) and one comparison was a mix of FRIENDS for children and FRIENDS for youth (4%). Three quarters of programmes were universal (75%), around one fifth were Indicated (21%) and one programme was selected (4%). All programmes, except two, were delivered at schools (93%). The number of sessions offered was 10 in 79% of programmes, less than 10 in 18% of the programmes and more than 10 in 4% of the programmes. The mean length of sessions was 63.6 min, ranging from 45 to 90 min. Booster sessions was reported to be offered in 38% of programmes, and parent sessions was reported offered in 56% of programmes. The

**TABLE 2** Comparisons used/not used in data synthesis.

Study	Comparison (if more than one)	Country	Used in data synthesis/reason not used
Ahlen (2018/2019)		Sweden	Used in data synthesis
Ahmari (2020)	Active control	Iran	Numbers are wrong
Ahmari (2020)	TAU control	Iran	Numbers are wrong
Anticich (2013)	Active control	Australia	Cannot calculate ES and SE
Anticich (2013)	Waitlist control	Australia	Cannot calculate ES and SE
Barker (2008)		Canada	Used in data synthesis
Barrett and Turner (2001)	Psychologist-led intervention	Australia	Used in data synthesis
Barrett and Turner (2001)	Teacher-led intervention	Australia	Data used in another study
Barrett (2000)		Australia	Used in data synthesis
Barrett (2003)	Friends for life (children)	Australia	Used in data synthesis
Barrett (2003)	Friends for youth (youth)	Australia	Used in data synthesis
Barrett (2005)		Australia	Part of data used in another study and something wrong with reported means
Barrett (2001)		Australia	Critical risk of bias
Cooley-Strickland (2011)		USA	Used in data synthesis
Doyle (2016)		Canada	Used in data synthesis
Essau (2012)		Germany	Used in data synthesis
Gallegos (2008/2013)		Mexico	Used in data synthesis
Ghajareih (2018)		Iran	Written in Persian
Haidari (2016)		Iran	Written in Persian
Hunt (2009)		Australia	Used in data synthesis
Kozina (2020)		Slovenia	Used in data synthesis
Kozina (2021)		Slovenia	Cannot calculate ES and SE
Kösters (2015)		the Netherlands	Used in data synthesis
Lawson (2023)		USA	Critical risk of bias
Liddle (2010)		Scotland	Used in data synthesis
Lock (2003)		Australia	Used in data synthesis
Lowry-Webster (2001/2002/2003)	Australia	Used in data synthesis	
Miller (2011a)	Trial 1	Canada	Used in data synthesis
Miller (2011a)	Trial 2	Canada	Used in data synthesis
Miller (2011b)		Canada	Used in data synthesis
Moharreri (2017)		Iran	Numbers are wrong
Maalouf (2020)		Lebanon	Used in data synthesis
Pahl (2009/2010)		Australia	Used in data synthesis
Rivero (2020)		Brasil	Used in data synthesis
Rodgers (2015)		Ireland	Used in data synthesis
Ruttledge (2016)		Ireland	Used in data synthesis

(Continues)

**TABLE 2** (Continued)

Study	Comparison (if more than one)	Country	Used in data synthesis/reason not used
Sabey (2019)	Friends for life	Hong Kong	Critical risk of bias
Sabey (2019)	FunFriends	Hong Kong	Critical risk of bias
Siu (2007)		Hong Kong	Used in data synthesis
Stallard (2014, 2015)/ Skryabina (2016)	Health-led intervention	UK	Used in data synthesis
Stallard (2014, 2015)/ Skryabina (2016)	Teacher-led intervention	UK	Used in data synthesis
Wigelsworth (2018)		UK	Used in data synthesis

intervention was delivered by teachers in 11 comparisons (39%), by psychologists in another 10 comparisons (36%) and the remaining programmes were delivered by a counsellor/other facilitator (25%).

### 6.1.3 | Excluded studies

In addition to the 42 studies that met the inclusion criteria for this review, 49 studies appeared relevant at first sight but did not meet our criteria for inclusion. The studies and reasons for exclusion are given in Table 4. More than half (27 studies) were excluded because they had no comparison group, seven studies were excluded because they were treatment and not prevention studies, and seven studies were excluded because they compared one unit (school or class) to one or more units.

Other reasons were, more than 70% of participants had an anxiety diagnosis (one study), no relevant outcome (two studies), the intervention analysed was not the FRIENDS intervention (one study), compared FRIENDS with and without an add on (two studies), the control group had received the intervention (one study) and wrong age (one study).

## 6.2 | Risk of bias in included studies

The risk of bias coding was carried out in accordance with the protocol (Filges, 2023) and the assessment of each of the 40 comparisons (excluding the two studies written in Persian which could not be assessed) is available in a supplementary document. For a summary of the risk of bias assessments, see Table 5 for the randomised trials (individual and cluster randomised) and Table 6 for the non-randomised trials. Thirty-one comparisons were obtained through randomising participants, either individually (7) or in clusters (24) and nine comparisons were obtained through a non-randomised allocation of participants. We used the ROB 2 tool for the individual randomised comparisons, and for cluster-randomised trials, an additional domain was included, as described in Section 5.3.3. The non-randomised comparisons were rated using the ROBINS-I tool.

None of the randomised comparisons had an overall low risk of bias. As it is not possible to blind participants or the provider of the intervention in these kinds of studies, it is not possible for a comparison to achieve a low risk of bias. The highest overall rating a comparison can acquire is *Some concerns*, which was acquired by only six randomised comparisons. One of these comparisons (reported in Wigelsworth, 2018) rated *Some concerns* overall and was rated *Low risk of bias* on all domains except in the domain Deviations from intervention, where the only concern was lack of blinding of participants and the provider of the intervention. Twenty-three comparisons had an overall rating of *High risk of bias* and two comparisons (both reported in the Ahmari, 2020 study) could not be rated overall as there were some serious unresolved issues, see notes in the supplementary document.

The majority of randomised comparisons were rated either *Some concerns* (11) or *High risk of bias* (11) on the randomisation process. Five were rated *Low risk of bias*, two offered no information and one could not be rated (Moharreri 2017) as we could not evaluate the pre-test imbalance as the data reported in the study seems wrong, see notes in the supplementary document.

On the Deviation from interventions, no comparisons were rated *Low risk of bias*. The majority, 18 comparisons, were rated *Some concerns*, nine were rated *High risk of bias* and four offered no information. On the Missing outcome data domain the majority were rated either *Low risk of bias* (14 comparisons) or *Some concerns* (9 comparisons), five were rated *High risk of bias* and the remaining three offered no information in this domain. Note that for several outcomes, the risk of bias in this domain were rated differently depending on the point in time the outcome is measured (post or follow-up). In these cases, the summary risk of bias scores are based on the time points with the most favourable rating. In some cases the differently rated Missing outcome data domain also led to a different overall rating of the outcome depending on the point in time of measurement. The summary overall risk of bias scores are based on the time points with the most favourable overall rating. On the Measurement of outcome domain only four comparisons were rated *Low risk of bias*, 17 were rated *Some concerns* and the remaining ten comparisons offered no information. Finally, on the Selection of

**TABLE 3** Characteristics of comparisons used in data synthesis.

Characteristic (Number of comparisons reporting)		
Year start of participation (9)	Average (SD)	2010 (4.3)
	Range	2001–2016
Number of participants, treated (28)	Average (SD)	240 (283.6)
	Range	8–1476
Number of participants, control (28)	Average (SD)	212 (286.8)
	Range	10–1534
Percent female (23)	Average (SD)	54 (11.4)
	Range	43–100
Age (21)	Average (SD)	10.4 (2.5)
	Range	4.6–15.8
Socioeconomic status (11)	Low SES	3 (27%)
Child reported pre-test scores, percent of max scale (27)	Mean (SD)	30.9 (12.6)
	Range	7.8–59.7
Programme type (28)	FunFriends	2 (7%)
	Friends for children	18 (64%)
	Friends for youth	7 (25%)
	Mix	1 (4%)
	Universal	21 (75%)
	Selected	1 (4%)
	Indicated	6 (21%)
Place of delivery (28)	In schools	26 (93%)
	Not reported	2 (7%)
Number of sessions offered (28)	Less than 10 sessions	5 (18%)
	10 sessions	22 (79%)
	More than 10 sessions	1 (4%)
Length of sessions (21)	Mean (SD)	63.6 (11.7)
	Range	45–90
Booster sessions offered (24)	Yes	9 (38%)
Parent sessions offered (27)	Yes	15 (56%)
Facilitator (28)	Counsellor/other facilitators	7 (25%)
	Psychologist	10 (36%)
	Teacher	11 (39%)

reported results domain, one comparison was rated *Low risk of bias* as it had a published study protocol and analysis plan (the Wigelsworth, 2018 study). The majority (24 comparisons) were rated *Some concerns*, in 22 cases because we could not locate a published a priori study protocol and analysis plan, while for two comparisons (reported in the Stallard, 2015 study) a protocol with analysis plan

**TABLE 4** Studies excluded with reason.

Study	Reason
Ahlen (2012)	Before-after study
Balle (2010)	Intervention is not FRIENDS but 'largely grounded on FRIENDS child in-session contents'.
Barrett (2006)	Follow-up to Lock 2003 but control had received the intervention most likely after the 12 months follow-up
Barrett (2015)	Treatment and before-after study
Bernstein (2005/2008)	Treatment study
Carlyle (2014)	Before-after study
Cooley (2004)	Before-after study
Desousa (2016)	Before-after study
Dohl (2013)	Two classrooms from one school compared to two classrooms from another school
Doyle (2020)	Before-after study
Eiraldi (2023)	Both groups receive FRIENDS, one group in addition to the traditional training of therapist implementors receives an added ongoing remote online consultation for supervisors
Farrell (2005)	Before-after study
Fisak (2018)	Treatment and before-after study
Fjermestad (2020)	Before-after study and compares to Wergeland 2014
Fukushima-Flores (2011)	Compares programme with and without parent participation
Gallegos (2012)	Before-after study
Gallegos (2015)	Before-after study
Gallegos-Guajardo (2020)	Before-after study
Games (2020)	University students
Garcia (2019)	Before-after study
Green (2013)	Before-after study
Hosokawa (2023)	Social skills is the only outcome
Iizuka (2014a)	Before-after study
Iizuka (2014b)	Before-after study
Iizuka (2015)	Before-after study
Kato (2017)	Compares one class to another class
Kavanagh (2014)	Before-after study
LaRose (2017)	Proposal for a trial
Lewis (2016)	One school compared to another school
Lewis (2023)	Before-after study
Martinsen (2009)	Treatment before-after study

(Continues)

**TABLE 4** (Continued)

Study	Reason
Matsumoto (2016)	Two schools compared to one control school
Mims (2015)	Before-after study
Moharer (2021)	Treatment study
Mostert (2008/2007)	Compares one school class to another school class
O'Brien (2007)	Treatment study
Onge (2016)	Before-after study
Paul (2011)	One class compared to another class
Pereira (2014)	70.3% had an anxiety diagnosis
Rose (2009)	Compares one school class to another school class
Salari (2018)	Treatment study
Salum (2018)	Treatment study
Shortt (2001)/Barrett (2001)	Treatment study
Stallard (2005, 2008)	Before-after study
Stallard (2007)	Before-after study
Stopa (2010)	Before-after study
Sykes (2009)	Before-after study
van der Mheen (2019)	Before-after study
Wergeland (2014)	Treatment study

was published. However there was a discrepancy between primary outcome according to the protocol and the primary outcome reported in the published studies. The remaining six comparisons were rated *High risk of bias* as they had other concerns in addition to not having a published *a priori* study protocol and analysis plan.

Concerning the nine non-randomised comparisons, four had an overall rating of *Critical risk of bias*, corresponding to a risk of bias so high that the findings should not be considered in the data synthesis. The overall *Critical risk of bias* rating was mainly due to issues on the Confounding domain, two were rated *Serious risk of bias* and as stated above three were rated *Critical risk of bias*. Eight comparisons were rated *Low risk of bias* on the Selection of participants domain and one was rated *Moderate risk of bias*. None of the non-randomised comparisons had issues on the Classification of intervention status domain, they were all rated *Low risk of bias*.

On the remaining domains (common to randomised and non-randomised comparisons) one comparison was rated *Low risk of bias* on the Deviation from intervention domain, four were rated *Moderate risk of bias*, three did not provide any information for us to rate on this domain and as stated above, one comparison was rated *Critical risk of bias* on the Deviation from intervention domain. On the Missing outcome data domain the majority were rated either *Low risk of bias* (4 comparisons), or *Moderate risk of bias* (4 comparisons), and one was rated *Serious risk of bias*. On the Measurement of outcome domain, six comparisons were rated *Moderate risk of bias*, and the remaining three did not provide any information for us to rate on this domain. Finally, on the Selection of reported results domain no comparisons had a published *a priori* study protocol and analysis plan and all were rated *Moderate risk of bias*.

### 6.3 | Effects of interventions

Note: 'Effects of interventions' heading will be removed at publication stage.

### 6.4 | Synthesis of results

All outcomes are measured such that a positive effect size favours the treated population, that is, a higher score for all outcomes means fewer anxiety symptoms.

#### 6.4.1 | Primary outcomes

Twenty-eight comparisons were not rated *Critical risk of bias*, and reported data that permitted calculation of an effect size and standard error and could, thus, be used in the data synthesis. All comparisons reported data on anxiety, however, unfortunately not all reported results post-intervention (Hunt, 2009) did not report the data for the control group at post-intervention. Further, two comparisons (Pahl, 2010 and Rivero, 2020) only reported anxiety measured by parents as the intervention was Fun FRIENDS for children aged 4–7 years and hence, no child reported data was collected). All continuous outcomes (effect sizes measured as Hedges  $g$ ) were coded such that a larger effect size indicated better outcomes for the treated group.

#### Anxiety post-intervention

Child-reported. Twenty-five effect sizes from 22 clusters were available for analysis. As mentioned above, the psychologist-led intervention in Barrett and Turner (2001) shared control group

On the three domains only assessed for the non-randomised comparisons, three were rated *Moderate risk of bias* on the

**TABLE 5** Summary risk of bias assessment, randomised comparisons.

Judgement:	Low risk of bias	Some concerns	High risk of bias	No information	Not rated	Number of comparisons
Risk of bias domain						
Overall Judgement	0	6	23	0	2	31
Randomisation process	5	12	11	2	1	31
Deviation from intervention	0	18	9	4	0	31
Missing Outcome Data	14	9	5	3	0	31
Measurement of Outcome	4	17	0	10	0	31
Selection of Reported Results	1	24	6	0	0	31

**TABLE 6** Summary risk of bias assessment, non-randomised comparisons.

Judgement:	Low risk of bias	Moderate risk of bias	Serious risk of bias	Critical risk of bias	No information	Number of comparisons
Risk of bias domain						
Overall Judgement	0	3	2	4	0	9
Confounding	0	4	2	3	0	9
Selection of participants	8	1	0	0	0	9
Classification of intervention status	9	0	0	0	0	9
Deviation from intervention	1	4	0	1	3	9
Missing Outcome Data	4	4	1	0	0	9
Measurement of Outcome	0	6	0	0	3	9
Selection of Reported Results	0	9	0	0	0	9

schools with Lowry-Webster (2003) and were considered one cluster; results were reported separately by age (children/youth) in Barrett (2003) and were considered one cluster; and the health-led and the teacher-led comparisons reported in Stallard (2015) were considered one cluster.

The random effects weighted SMD obtained using RVE was 0.13 [95% CI 0.04 to 0.22] and statistically significant with an adjusted degree of freedom level of 17.46. The forest plot is displayed in Figure 3. There was some heterogeneity between comparisons; the estimated  $\tau^2$  was 0.02,  $Q = 48.87$  and  $I^2$  was 51%. The prediction interval was [95% PI -0.16 to 0.42].

*Parent-reported.* Five effect sizes from five clusters were available for analysis and RVE was not necessary.

The random effects weighted SMD was -0.08 [95% CI -0.20 to 0.04] and not statistically significant. The forest plot is displayed in Figure 4. None of the three heterogeneity measures indicated any heterogeneity; the estimated  $\tau^2$  was 0.00,  $Q = 3.33.65$  and  $I^2$  was 0%. However, interpretations concerning heterogeneity should be made cautiously on account of the low statistical power available to detect it. Prediction intervals could not be calculated as the estimated  $\tau^2$  was 0.00 (Higgins 2009).

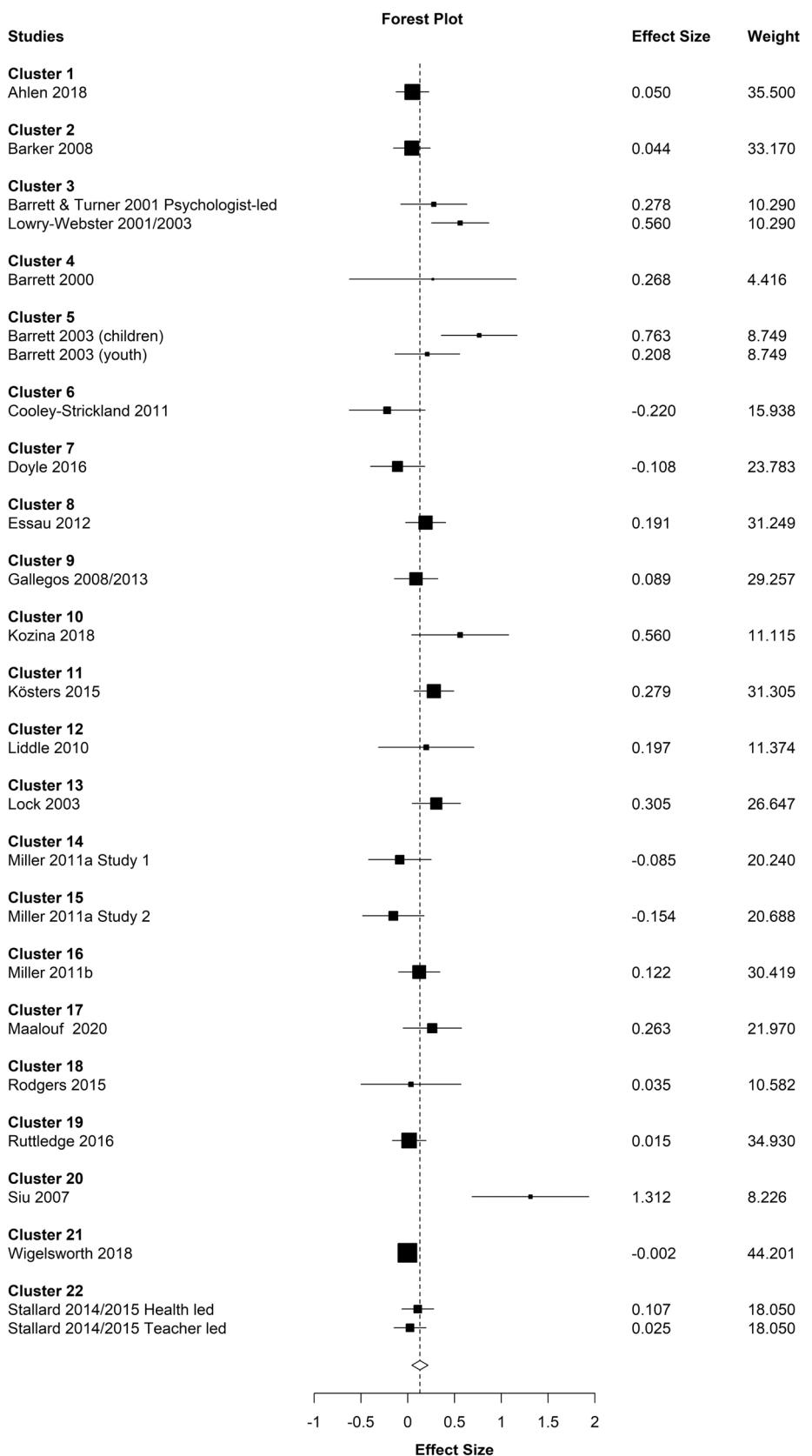
*Anxiety 12 months follow-up*

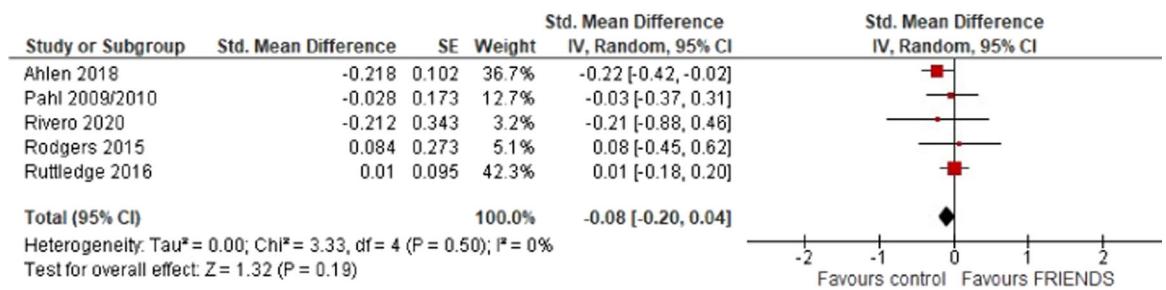
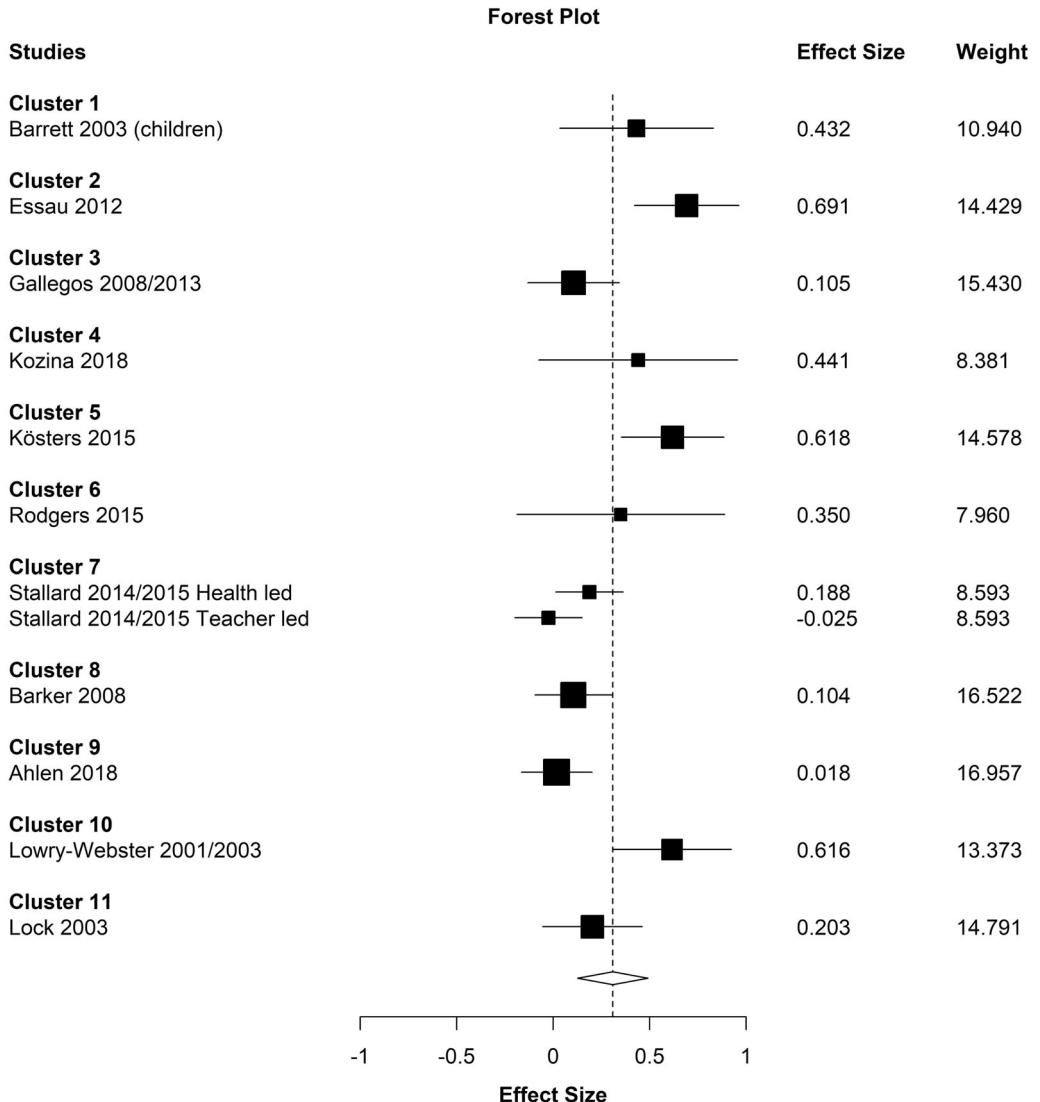
*Child-reported.* Twelve effect sizes from 11 clusters were available for analysis (the health-led and the teacher-led comparisons in Stallard, 2015 were considered one cluster).

The random effects weighted SMD obtained using RVE was 0.31 [95% CI 0.13 to 0.49] and statistically significant. The forest plot is displayed in Figure 5. All three heterogeneity measures indicated substantial heterogeneity; the estimated  $\tau^2$  was 0.05,  $Q = 39.07$  and  $I^2$  was 74%. The prediction interval was [95% PI -0.17 to 0.79].

*Parent-reported.* Four effect sizes from three clusters were available for analysis (the health-led and the teacher-led comparisons in Stallard, 2015 were considered one cluster).

The number of comparisons and the degrees of freedom were too small for the RVE procedure to yield trustworthy results regarding the standard errors (Tanner-Smith & Tipton, 2014; Tipton, 2015). The meta-analysis was therefore performed using Revman 5.4 and the synthetic average from the cluster with two effect estimates. The random effects weighted SMD was -0.10 [95% CI -0.26 to 0.06] and not statistically significant. The forest plot is displayed in Figure 6. None of the three heterogeneity measures

**FIGURE 3** Forest plot anxiety child reported post-intervention.

**FIGURE 4** Forest plot anxiety parent reported post-intervention.**FIGURE 5** Forest plot Anxiety Child reported 12 months follow-up.

indicated any heterogeneity; the estimated  $\tau^2$  was 0.00,  $Q = 0.54$  and  $I^2$  was 0%. However, interpretations concerning heterogeneity shall be made cautiously on account of the low statistical power to detect it. Prediction intervals could not be calculated as the estimated  $\tau^2$  was 0.00 (Higgins, 2009).

#### Anxiety 24 months follow-up

Child-reported. Only three effect sizes from two clusters were available for analysis (the health-led and the teacher-led comparisons in Stallard, 2015 were considered one cluster).

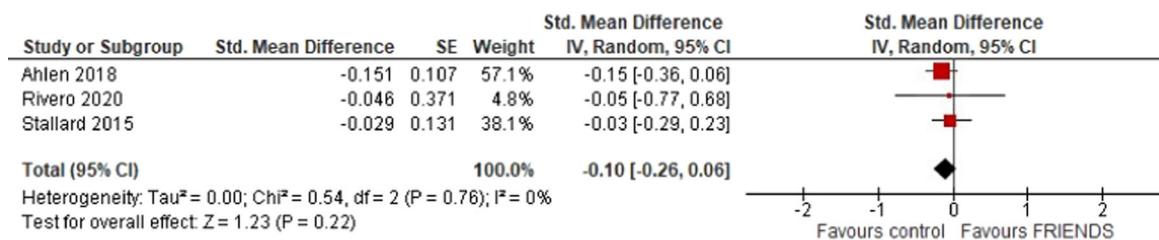


FIGURE 6 Forest plot Anxiety Parent reported 12 months follow-up.

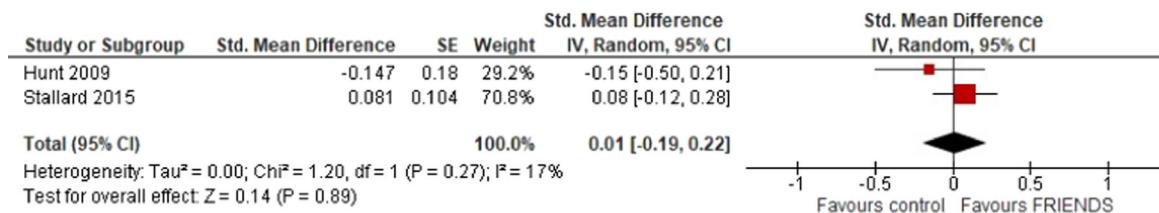


FIGURE 7 Forest plot anxiety child reported 24 months follow-up.

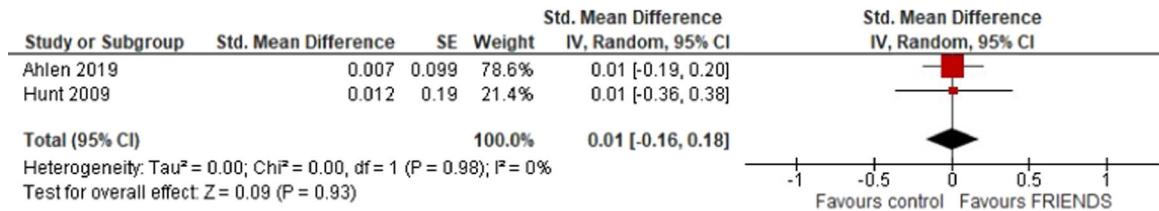


FIGURE 8 Forest plot anxiety child reported 24+ months follow-up.

The number of comparisons and the degrees of freedom were too low for the RVE procedure to yield trustworthy results regarding the standard errors. The random effects weighted SMD was 0.01 [95% CI -0.19 to 0.22] and not statistically significant. The forest plot is displayed in Figure 7. All three heterogeneity measures indicated there was a very small amount of heterogeneity; the estimated  $\tau^2$  was 0.00,  $Q = 1.20$  and  $I^2 = 17\%$ . However, interpretations concerning heterogeneity should be made cautiously on account of the low statistical power to detect it. Prediction intervals were not calculated with only two effect estimates in the analysis (Higgins, 2009).

*Parent-reported.* Only two effect sizes from the same cluster (the health-led and the teacher-led comparisons in Stallard, 2015) were available. We report the synthetic average.

The SMD was -0.12 [95% CI -0.34 to 0.10] and not statistically significant.

#### Anxiety 24+ months follow-up

*Child-reported.* Only two effect sizes from two clusters were available for analysis.

The random effects weighted SMD was 0.01 [95% CI -0.16 to 0.18] and not statistically significant. The forest plot is displayed in

Figure 8. All three heterogeneity measures indicated there was no heterogeneity; the estimated  $\tau^2$  was 0.00,  $Q = 0.0$  and  $I^2 = 0\%$ . However, interpretations concerning heterogeneity shall be made cautiously on account of the low statistical power to detect it. Prediction intervals were not calculated with only two effect estimates in the analysis and an estimated  $\tau^2$  of 0.00 (Higgins, 2009).

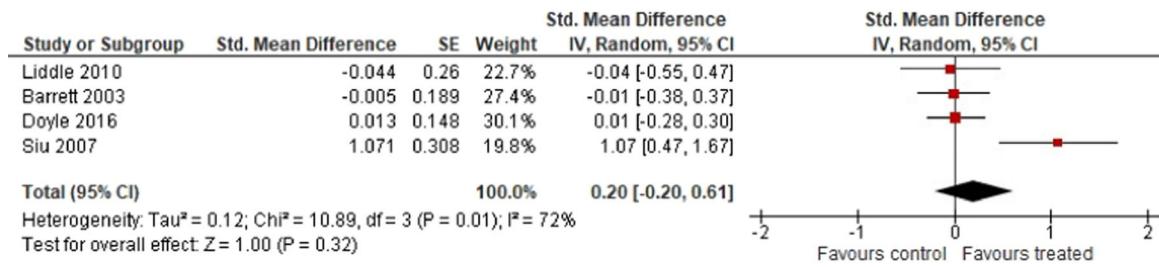
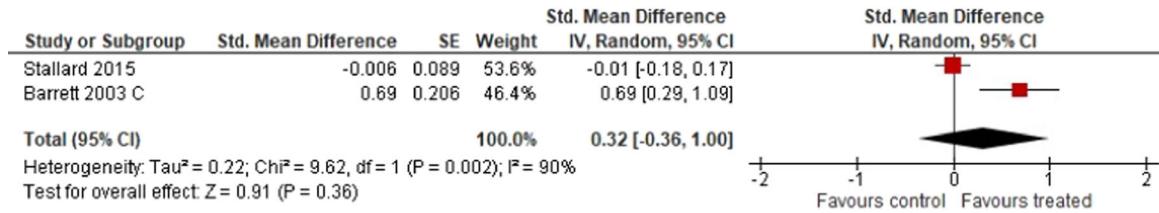
*Parent-reported.* There was one effect size available at more than 24 months follow-up. The SMD was -0.05 [95% CI -0.28 to 0.18] and not statistically significant.

#### 6.4.2 | Secondary outcomes

##### Self-esteem post-intervention

Five effect estimates from four clusters (results were reported separately by age (children/youth) in Barrett (2003) and were considered one cluster). The number of comparisons and the degrees of freedom were too small for the RVE procedure to yield trustworthy results regarding the standard errors (Tanner-Smith & Tipton, 2014; Tipton, 2015).

The random effects weighted SMD was 0.20 [95% CI -0.20 to 0.61] and not statistically significant. The forest plot is displayed in Figure 9. The three heterogeneity measures indicated substantial

**FIGURE 9** Forest plot self-esteem post-intervention.**FIGURE 10** Forest plot Self-esteem 12 months follow-up.

heterogeneity; the estimated  $\tau^2$  was 0.12,  $Q = 10.89$  and  $I^2$  was 72%. The prediction interval was [95% PI -0.58 to 1.00].

#### Self-esteem 12 months follow-up

Only three effect sizes, of which two were from the same cluster (the health-led and the teacher-led comparisons in Stallard, 2015) were available. We used the synthetic average.

The random effects weighted SMD was 0.32 [95% CI -0.36 to 1.00] and not statistically significant. The forest plot is displayed in Figure 10. The three heterogeneity measures indicated substantial heterogeneity; the estimated  $\tau^2$  was 0.22,  $Q = 9.62$  and  $I^2$  was 90%.

#### Self-esteem 24 months follow-up

Only two effect sizes from the same cluster (the health-led and the teacher-led comparisons in Stallard, 2015) were available. We report the synthetic average.

The SMD was 0.06 [95% CI -0.14 to 0.27] and not statistically significant.

#### Self-esteem 24+ months follow-up

No effect estimates were available more than 24 months after intervention.

### 6.4.3 | Moderator analysis

The included comparisons differed in terms of their participant characteristics and programme characteristics. Evidence of statistical heterogeneity was found in several meta-analyses.

We therefore performed moderator analysis to attempt to identify the characteristics of interventions that were associated with effect sizes. We used the RVE procedure robumeta in STATA and

reported 95% CIs for regression coefficients. We performed multiple meta-regression analyses where possible, in which we estimated RVE models including moderators besides the intercept. Below we describe the variables we used to define potential moderators.

To reduce the number of moderators, we first excluded moderators with very low variation (i.e., for which nearly all observations have the same value) or where information was missing from studies.

Due to missing data on the moderators gender and SES, we could not explore the impact of these covariates. The remaining potential moderators (7) were defined as follows:

- Indicated: Type of programme, that is, whether it is a universal, indicated or selective intervention was grouped as indicated versus not indicated as only one intervention was characterised as selected and the remaining (not indicated) were universal (21 comparisons).
- Youth: My FRIENDS Youth (12–16 year olds) versus FRIENDS for Life (8–11 year olds).
- Teacher: Type of provider, teacher versus not teacher (mental health provider or counsellor).
- Australia: Country of implementation Australia versus other countries.
- Booster: Booster sessions implemented versus not offered or not reported if they were implemented.
- Parent: Parent sessions offered versus not offered or not reported if they were offered.
- Noplml: No indication there were other implementation problems (full programme (10 sessions) implemented and programme integrity measures reported showing high concordance between session and manual content) versus implementation problems (full programme (10 sessions) not implemented, or programme integrity measures reported not showing high concordance between session and manual content or measures not reported).

### Anxiety post-intervention

*Child-reported.* To retain enough (adjusted) degrees of freedom, we included only study characteristics where at least four comparisons had the same characteristics. All seven possible moderators were available. However, some of the potential moderators were highly correlated (see Table 7). High correlations between moderators increase the risk of multicollinearity and make it less likely that we can estimate meaningful separate associations.

Considering the correlations in Table 7, this problem seems to be particularly high for indicated programmes and parent sessions. Booster sessions and parent sessions were highly correlated and indicated programmes were highly correlated with programmes implemented without any further implementation problems. We chose to leave out parent sessions and indicated programmes as moderators as they were to a higher extent correlated with the remaining covariates than the booster sessions and no further implementation problems covariates.

Table 8 displays the results from the meta-regression in column (1). Although there is a relatively large number of moderators compared to the number of clusters, the small sample corrected degrees of freedom were not close to the level where RVE starts to perform poorly, except for the covariates Youth and Australia. The adjusted degrees of freedom for Youth is, however, above 4, whereas it is just below 4 for Australia and thereby low enough that the RVE procedure may yield untrustworthy results regarding the standard errors for this covariate (Tanner-Smith & Tipton, 2014; Tipton, 2015). In column (2) we have added a meta-regression with only Australia as included covariate to investigate if the number of covariates included has an impact on the low level of adjusted degrees of freedom. The adjusted degrees of freedom is still below 4, and the result is not trustworthy. Most likely it is due to the covariate being unbalanced, only 6 comparisons are from Australia and this includes two of the clusters as well.

There were no statistically significant moderators, except Australia, with the coefficient also being relatively large in relation to the baseline effect.

**TABLE 7** Anxiety child reported post-intervention correlation matrix for potential moderators.

	Indicated	Youth	Australia	Teacher	Booster sessions offered	Parent sessions offered	No implementation problems
Indicated	1						
Youth	-0.25	1					
Australia	-0.28	0.19	1				
Teacher	-0.20	-0.20	-0.08	1			
Booster sessions offered	-0.09	-0.09	0.07	0.04	1		
Parent sessions offered	0.32	-0.28	-0.16	0.03	0.47	1	
No implementation problems	-0.44	0.16	0.26	-0.07	-0.01	-0.37	1

**TABLE 8** Anxiety child reported post-intervention meta regression (RVE).

Moderator	(1)				(2)			
	Coefficient	95% CI lower	95% CI higher	Adj. df	Coefficient	95% CI lower	95% CI higher	Adj. df
Youth	0.10	-0.18	0.37	4.39				
Australia <sup>a</sup>	0.28	0.08	0.47	3.31	0.29	0.09	0.50	2.92
Teacher	-0.10	-0.27	0.08	13.56				
Booster	0.08	-0.12	0.27	9.99				
Noplml	-0.02	-0.20	0.16	10.84				
Constant	0.11	-0.15	0.37	7.91	0.09	0.01	0.17	13.66
# Clusters	22				22			
# Effect sizes	25				25			
$\tau^2$	0.02				0.01			
Q-statistic	34.96				37.84			
$I^2$	43%				37%			

Note: Italics for this covariate the adjusted degrees of freedom are low enough that the RVE procedure may yield untrustworthy results regarding the standard errors.

<sup>a</sup>Note the adjusted degrees of freedom are low enough that the RVE procedure may yield untrustworthy results regarding the standard errors.

The Q statistic was reduced, however, there was still heterogeneity;  $\tau^2$  increased whereas  $I^2$  was reduced slightly.

*Parent-reported.* There was no heterogeneity to investigate.

#### Anxiety 12 months follow-up

*Child-reported.* To retain enough (adjusted) degrees of freedom, we included only study characteristics where at least four comparisons had the same characteristics. Only one programme was indicated, two were programmes for youth and three were from Australia, leaving four possible moderators available for analysis. Some of the potential moderators were correlated. In particular, booster sessions were highly correlated with parent sessions and to a lesser extent further implementation problems, see Table 9. We therefore chose to leave booster sessions out of the analysis, leaving three moderators to be analysed. Table 10 displays the results from the meta-regression. Although there is a relatively large number of moderators compared to the number of clusters, the small sample corrected degrees of freedom were high enough for RVE to perform well. The coefficient on Teacher was negative, statistically significant and relatively large in relation to the baseline effect. This implies that programmes implemented by providers other than teachers (mental health provider or counsellor) performed better than programmes

implemented by teachers. There was still heterogeneity although  $\tau^2$  and  $I^2$  were reduced slightly.

*Parent-reported.* There was no heterogeneity to investigate.

#### Anxiety 24 and 24+ months follow-up

*Child-reported.* There was no heterogeneity to investigate.

*Parent-reported.* There was no heterogeneity to investigate.

#### Secondary outcomes

*Self-esteem post-intervention.* Due to invariability between comparisons on most moderators, it was only possible to investigate the impact of the moderator Indicated. With five effect estimates and four clusters, it was only possible to perform a subgroup analysis using the synthetic average of the two effects from the one cluster. The random effects weighted SMDs from the two subgroups is reported in Table 11. Interpretation of relationships is cautious and not based on statistical significance of subgroup average effect sizes but based on 95% CIs (if there is overlap of CIs or not). The CIs of the subgroups overlapped. There was no evidence to support the hypothesis that the effect differs by type of programme.

*Self-esteem 12, 24 and 24+ months follow-up.* There were too few comparisons to perform any moderator analyses.

#### 6.4.4 | Sensitivity analysis

We performed a number of sensitivity analyses.

Sensitivity analysis was carried out where possible, by restricting each meta-analysis to a subset of all studies included in the original meta-analysis. We performed sensitivity analysis with regard to research design and restricted the analyses to studies using randomisation. We further performed sensitivity analyses for each domain of the risk of bias checklists and restricted the analysis to studies with either a *low risk of bias* or *some concerns*. Sensitivity analysis was only conducted when there were at least two comparisons left in the analyses. We tested sensitivity to clustered delivery of treatment by reporting results using both unadjusted effect sizes and using a substantially higher ICC (0.1) than in the primary analysis.

**TABLE 9** Anxiety child reported 12 months follow-up correlation matrix f or potential moderators.

	Teacher	Booster	Parent	Nolmpl
Teacher	1			
Booster sessions offered	0.00	1		
Parent sessions offered	0.00	<b>0.71</b>	1	
No implementation problems	0.00	0.35	0.13	1

Note: Bold used for value greater than 0.5.

**TABLE 10** Anxiety Child reported 12 months follow-up Meta-regression (RVE).

Moderator	Coefficient	95% CI lower bound	95% CI upper bound	Adj. df
Teacher	-0.31	-0.55	-0.06	6.93
Parent	0.15	-0.20	0.50	4.92
Nolmpl	0.02	-0.35	0.38	5.26
Constant	0.37	0.19	0.55	5.19
# Clusters	11			
# Effect sizes	12			
$\tau^2$	0.03			
Q-statistic	19.25			
$I^2$	64%			

**TABLE 11** Self-esteem post-intervention subgroup.

Moderator	SMD [95% CI]	# Effect sizes
Indicated	0.50 [-0.59, 1.59]	2
Universal	0.01 [-0.22, 0.23]	2

Note: SMD is random effects weighted standardised mean difference.

Finally, we conducted sensitivity tests using a variety of  $p$  values, to assess if the general results and estimates of the heterogeneity were robust to the choice of  $p$  in the two analyses using RVE.

The results of these sensitivity analyses are reported in Supporting Information S5: Appendix 6 (Figures S11–S15, Tables S12–S20). There were no appreciable changes to any of the results.

#### 6.4.5 | Publication bias

##### Anxiety child-reported post-intervention

We conducted sensitivity analyses for publication bias in which affirmative studies (i.e., those with statistically significant positive estimates) are more likely to be published than non-affirmative studies (i.e., those with non-significant estimates or negative estimates) by a certain ratio, called selection ratio. We used a selection ratio of 4 and 6; and in addition, we performed a worst-case scenario where affirmative studies are infinitely more likely to be published than non-affirmative studies, see Table 12.

**TABLE 12** Publication bias anxiety child reported post-intervention.

Anxiety child reported post-intervention	SMD	95% CI
Main analysis	0.13	[0.04, 0.22]
Selection ratio 4	0.09	[0.02, 0.15]
Selection ratio 6	0.08	[0.01, 0.14]
Worst case	0.06	[0.002, 0.12]

We also attempted to calculate the S-value, defined as the severity of publication bias (i.e., the ratio by which affirmative studies are more likely to be published than non-affirmative studies) that would be required to shift the pooled point estimate or its CI limit to the null.

The result of the analysis was that there was no amount of publication bias that would shift the pooled point estimate or its CI limit to the null. We thus consider the meta-analysis to be relatively robust.

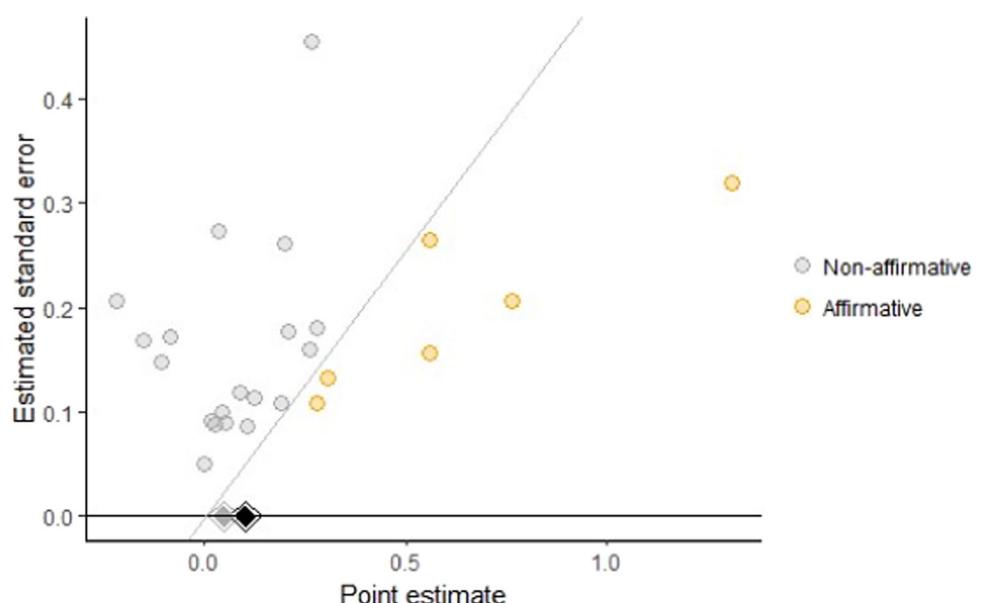
Finally, as a visual supplement, we present a modified funnel plot called the 'significance funnel plot'. This plot distinguishes between affirmative and non-affirmative studies, helping to detect the extent to which the non-affirmative studies' point estimates are systematically smaller than the entire set of point estimates. The estimate amongst only non-affirmative studies (grey diamond) represents a corrected estimate under worst-case publication bias. The significance funnel plot confirms that the meta-analysis seems robust to extreme publication bias, see Figure 11.

##### Anxiety child-reported 12 months follow-up

We conducted sensitivity analyses for publication bias using a selection ratio of 4 and 6; and in addition, we performed a worst-case scenario, see Table 13.

We also attempted to calculate the S-value. The result of the analysis was that there was no amount of publication bias that would shift the pooled point estimate or its CI limit to the null. We thus consider the meta-analysis to be relatively robust.

Finally, as a visual supplement, the significance funnel plot confirms that the meta-analysis seems robust to extreme publication bias, see Figure 12.



**FIGURE 11** Significance funnel plot Anxiety child reported post-intervention.

## 7 | DISCUSSION

### 7.1 | Summary of main results

Based on the 28 comparisons which could be used in the data synthesis, meta-analyses were conducted on the outcomes anxiety child-reported, anxiety parent-reported and the secondary outcome self-esteem.

#### 7.1.1 | Anxiety child-reported post-intervention

The available evidence does suggest that the FRIENDS programme may reduce anxiety symptoms in children and adolescents. We found a statistically significant positive effect of small size. The weighted average effect (using the 25 comparisons from 22 clusters reporting this outcome) was 0.13 [95% CI 0.04 to 0.22].

To ease interpretation, we translate the SMD into the probability-of benefit (POB) statistic. POB is also known as the Common Language effect size statistic (CL) defined in McGraw and Wong (1992). To calculate the POB for continuous data, we

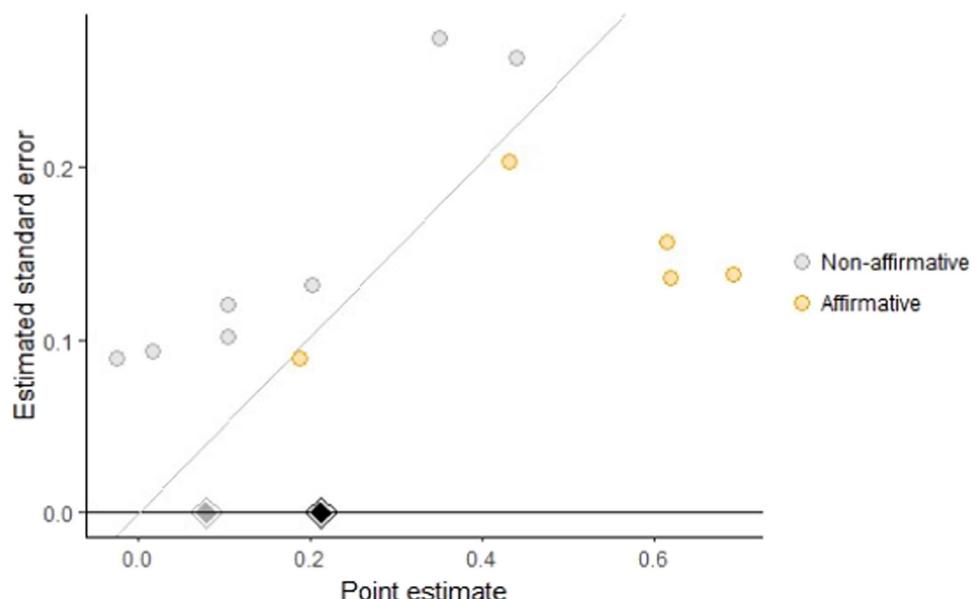
first compute:  $Z = \text{SMD}/\sqrt{2}$ . This Z statistic is normally expressed as a standard normal distribution, and the POB is computed as the probability that a randomly selected standard normal variable is less than Z. Measured as the POB statistic, defined as the probability that a randomly selected score from the treated population of children would be better (note that a higher score for this outcome means fewer anxiety symptoms) than a randomly selected score from the comparison population of children, the overall anxiety symptom reduction of children and adolescents POB was 0.54. A SMD of 0.13, therefore, corresponds to a 54% chance that a randomly selected score of a child from the treated population is better than the score of a randomly selected child from the comparison population. Note that this probability should be compared to a fifty-fifty chance, which is 50%, had the intervention been ineffective. The lower and upper 95% CI corresponds to 51% and 56%, respectively, chance of a randomly selected score of the treated being higher than a score from the comparison population.

Some heterogeneity was present and the 95% prediction interval was [95% PI -0.16 to 0.42] reflecting that the distribution of effect sizes spanned both positive and negative effect sizes.

We examined the impact of the moderators youth, Australia, teacher, booster and no further implementation problems. None of the moderators were statistically significant, except Australia, with the coefficient also being relatively large in relation to the baseline effect. Unfortunately, the adjusted degrees of freedom for Australia is, however, just below 4 and thereby low enough that the RVE procedure may yield untrustworthy results regarding the standard errors for this covariate. Heterogeneity was only reduced slightly.

**TABLE 13** Publication bias anxiety child reported 12 months follow-up.

Anxiety child reported 12 months follow-up	SMD	95% CI
Main analysis	0.31	[0.13, 0.49]
Selection ratio 4	0.19	[0.05, 0.33]
Selection ratio 6	0.17	[0.04, 0.30]
Worst case	0.13	[0.00, 0.26]



**FIGURE 12** Significance funnel plot anxiety child reported 12 months follow-up.

## 7.1.2 | Anxiety parent-reported post-intervention

We are uncertain about the effect of the FRIENDS programme on anxiety symptoms in children and adolescents when measured by parents. We found a statistically non-significant negative effect. The weighted average effect (using the 5 comparisons from 5 clusters reporting this outcome) was  $-0.08$  [95% CI  $-0.20$  to  $0.04$ ].

However, given the relatively few comparisons, some caution is needed in making an assumption that the FRIENDS programme does not decrease children and adolescents' anxiety symptoms when measured by parents. This corresponds to a 48% chance that a randomly selected score of a child from the treated population is better (note that a higher score for this outcome means fewer anxiety symptoms) than the score of a randomly selected child from the comparison population. The lower and upper 95% CI corresponds to 44% and 51%, respectively.

No heterogeneity was found. However, interpretations concerning heterogeneity should be made cautiously on account of the low statistical power to detect it.

## 7.1.3 | Anxiety child-reported medium term follow-up

The available evidence does suggest that the FRIENDS programme may reduce anxiety symptoms in children and adolescents. We found a statistically significant positive effect of moderate size. The weighted average effect (using the 12 comparisons from 11 clusters reporting this outcome) was  $0.31$  [95% CI  $0.13$  to  $0.49$ ].

This corresponds to a 59% chance that a randomly selected score of a child from the treated population is better (note that a higher score for this outcome means fewer anxiety symptoms) than the score of a randomly selected child from the comparison population. The lower and upper 95% CI corresponds to 54% and 64%, respectively.

A large amount of heterogeneity was present and the 95% prediction interval was [95% PI  $-0.17$  to  $0.79$ ] reflecting that the distribution of effect sizes span both positive and negative effect sizes.

We examined the impact of the moderators teacher, parent and no further implementation problems. The metaregression coefficient on teacher was negative, statistically significant and relatively large in relation to the baseline effect. This implies that programmes implemented by providers other than teachers (mental health provider or counsellor) performed better than programmes implemented by teachers. There was still heterogeneity, although slightly reduced.

## 7.1.4 | Other outcomes

In the remaining meta-analyses there were too few comparisons included (at most four in each meta-analysis) and/or the degree of heterogeneity was too high in order for us to draw any conclusion concerning the effectiveness of the FRIENDS programme.

## 7.2 | Overall completeness and applicability of evidence

A total of 28 comparisons were available for data synthesis. Four comparisons could not be used in the data synthesis as all reported outcomes were judged to have a *Critical risk of bias*. In accordance with the protocol, we excluded studies rated overall *Critical risk of bias* items from the data synthesis on the basis that they would be more likely to mislead than inform. Six comparisons were from studies that did not report data in a form that enabled the calculation of effect sizes and standard errors. Attempts to request the data from study authors produced no answers and, finally, two comparisons were not used as there was partial overlap of data to other comparisons used.

The 28 comparisons used in the data synthesis covered 14 different countries (Australia, Brasil, Canada, Germany, Hong Kong, Ireland, Lebanon, Mexico, the Netherlands, Scotland, Slovenia, Sweden, UK and USA) whereas 15 countries (the comparisons from Iran could not be used) were represented by the total 40 comparisons included in the review. The geographical coverage is not considered to limit the applicability of the evidence.

Evidence of statistical heterogeneity was found in several meta-analyses indicating inconsistency of results across comparisons.

Due to missing data on the moderators gender and SES, we could not explore the impact of these covariates.

It was possible to study the impact of seven moderators:

(1) Indicated (type of programme, universal, indicated or selective intervention), (2) youth (My FRIENDS Youth [12–16 year olds] vs. FRIENDS for Life [8–11 year olds]), (3) teacher (type of provider, teacher vs. mental health provider or counsellor), (4) Australia (country of implementation Australia vs. other countries), (5) booster: booster sessions implemented versus not offered or not reported if they were implemented, (6) parent (parent sessions offered vs. not offered or not reported if they were offered) and (7) no further implementation problems (no indication there were other implementation problems [full programme (10 sessions) implemented, and programme integrity measures reported showing high concordance between session and manual content] vs. implementation problems [full programme (10 sessions) not implemented, or programme integrity measures reported not showing high concordance between session and manual content or measures not reported]).

However, due to high correlations between some of the moderators they could not all be included and further, the number of comparisons in each meta-analysis restricted the number of moderators included.

## 7.3 | Quality of the evidence

The overall certainty of evidence varied from low to very low. A grading of low certainty means further research is very likely to have an important impact on our confidence in the effect estimate and

may change the estimate. A grading of very low certainty means we are very uncertain about the effect estimate.

Thirty-one comparisons were obtained through randomising participants, either individual (7) or in clusters (24) and nine comparisons were obtained through a non-randomised allocation of participants. We used the ROB 2 tool for the individual randomised comparisons (RCT) and for cluster-randomised trials (CRCT), an additional domain was included, as described in Section 5.3.3. The non-randomised comparisons (NRS) were rated using the ROBINS-I tool.

We attempted to enhance the quality of the evidence in this review by excluding NRS comparisons judged to be of *Critical risk of bias* using this tool. We believe this process excluded those studies that are more likely to mislead than inform.

Twenty-three randomised comparisons had an overall rating of *High risk of bias* and two comparisons could not be rated overall as there were some serious unresolved issues. The remaining six comparisons had an overall rating of *Some concerns*. Four non-randomised comparisons had an overall rating of *Critical risk of bias*, two were rated *Serious risk of bias* overall and the remaining three were rated *Moderate*.

Concerning the overall certainty of the evidence, the GRADE evidence profile (Summary of findings Table 1) indicates that the certainty of evidence was low for all pooled effect sizes except the self-esteem at post-intervention, which was very low.

Downgrading of evidence was undertaken for all pooled effect sizes because of risk of bias in the design of the studies. Apart from the problems with risk of bias, the directness, consistency, precision and publication bias were not downgraded, except for the self-esteem at post-intervention where there was important unexplained inconsistency (heterogeneity) in the results.

Furthermore, we performed a number of sensitivity analyses to check whether the obtained results were robust across study design and methodological quality.

We performed sensitivity analyses with regard to research design and restricted the analysis to studies using randomisation. We further performed sensitivity analyses for each domain of the risk of bias checklists and restricted the analysis to studies with either a *low risk of bias* or *some concerns*. We tested sensitivity to clustered delivery of treatment and finally, we conducted sensitivity tests using a variety of  $p$  values, to assess if the general results and estimates of the heterogeneity were robust to the choice of  $p$  in the two analyses using RVE.

There were no appreciable changes to any of the results.

## 7.4 | Potential biases in the review process

We performed a comprehensive electronic database search, combined with grey literature searching, and hand searching for key journals. All citations were screened in teams by two independent review authors (TF and MWK).

We believe that all the publicly available studies on the effect of the FRIENDS programme on reducing anxiety symptoms in children and adolescents up to the censor date were identified during the review process.

Unfortunately, two studies with a title and an abstract in English indicating they were eligible, could not be assessed as, except for the title and abstract, they were written in Persian.

We are confident that the two meta-analyses with more than five comparisons are robust to extreme publication bias. We are unable to comment on the possibility of publication bias in the remaining meta-analyses as at most five comparisons were included in each meta-analysis.

We believe that there are no other potential biases in the review process as three review authors in pairs of two (TF, MWK and GS) independently coded the included studies. Any disagreements were resolved by discussion. Further, decisions about inclusion of studies were made by the team of screeners (TF and MWK) and one further review author (GS). Assessment of study quality and numeric data extraction was made by one review author (TF) and each study was checked by one other review author (MWK and GS).

## 7.5 | Agreements and disagreements with other studies or reviews

We located three systematic reviews on anxiety preventive programmes for children and adolescents in general (including FRIENDS) and three reviews on the specific programme FRIENDS. In addition, after we finished our search (and after the first submission of the current review), we located a systematic review on the FRIENDS programme (Fisak, 2023).

Werner-Seidler et al. (2017) and the update Werner-Seidler et al. (2021):

Performed a systematic review on school-based depression and anxiety prevention programmes for young people (i.e., children or adolescents with a mean age between 5 and 19 years met their inclusion criteria). In the update, the authors searched up to October 2020. Of the 118 included studies, 34 studies were classified as focused on preventing anxiety, and 30 were classified as being mixed depression/anxiety prevention programmes. A total of 18 evaluations of the FRIENDS anxiety prevention programme were included (and four follow-up studies), the majority (15) were classified as anxiety prevention programmes, but three were classified as mixed depression/anxiety prevention programmes. The classification of the focus of the studies, however, plays no role in the analyses as all studies reporting anxiety outcomes (72) were used in the analyses of anxiety, regardless of the focus of the study as classified by the review authors. Separate meta-analyses were performed for depression and anxiety outcomes. Single factor subgroup analyses and multiple meta-regressions were performed with a number of study characteristics entered as predictors of the outcomes. In particular, programme content was examined, but the impact of FRIENDS was not investigated, rather the impact of CBT-based programmes versus other therapeutic approaches was investigated. None of the effect sizes analysed were corrected for clustering at neither the school nor the class level (if needed). The review included both universal prevention programmes, which typically are both assigned and delivered

at the class or school level, and targeted (selective/indicative) prevention programmes which may be assigned at the school or class level but are not delivered to whole classes.

**Ahlen et al. (2015):**

Performed a systematic review on universal prevention programmes targeting anxiety or/and depression in school-aged children 6–18 years. The searches were performed in July 2012. Of the 30 included studies, the primary aim of the intervention was to prevent depression in 13 studies, to prevent anxiety in 10 studies, and to prevent both anxiety and depression in 7.

A total of six evaluations of the FRIENDS anxiety prevention programme were included (and two follow-up studies), all were classified as anxiety prevention programmes.

All studies reporting anxiety outcomes (18) were used in the analyses of anxiety regardless of the primary aim of the intervention as classified by the review authors. Separate meta-analyses were performed for depression and anxiety outcomes. The intervention 'FRIENDS for life' was examined as moderator for anxiety symptoms using single factor subgroups. No significant differences were found. Based on six effect sizes, they found a post-intervention effect of 0.21 and based on five effect sizes, they found a follow-up effect of 0.31; both statistically non-significant.

**Higgins and O'Sullivan (2015):**

The review provided a narrative summary of five randomised controlled trials (and two follow-up studies) which examined the effectiveness of the FRIENDS programme as a preventative universal intervention for children and youth (aged 4–16 years) anxiety. Studies published in peer-reviewed journals between 2000 and 2013 were eligible.

**Maggin and Johnson (2014):**

This is a systematic review with meta-analysis on school-based FRIENDS for students enroled in Kindergarten to grade 12. Only group based experimental or quasi-experimental studies with a control group reporting on standardised measures of anxiety were eligible. The review included 17 studies (reported in 16 manuscripts), of which two were follow-up studies.

The final search date (reported in Maggin & Johnson, 2019, which was a reply to a critique provided in Barrett et al., 2017) was October 2010. All analyses in the review were separated by the review authors' definition of low-risk students and students with elevated risk (defined as pre-test scores within the clinical range, though the cut-off scores used were not reported). There are further a number of other ambiguities. First, it is unclear which studies provide effect estimates to which analyses (low/high risk students and post/follow-up) as the numbers available (according to tab. 1 in Maggin & Johnson, 2014) do not add up to the numbers reported used in tab. 4 in Maggin and Johnson (2014). Second, it is unclear which anxiety measures are used. Five studies reported results for two or more standardised anxiety measures. The review authors

stated that they randomly selected one measure from each of these studies to include in the meta-analyses. It was not reported which ones were used.

According to their Table 4, based on 13 effect sizes for low-risk students, they found a post-intervention effect of 0.26 and based on six effect sizes for low-risk students, they found a follow-up effect of 0.31; both statistically significant. Based on five effect estimates, the corresponding effects at post-intervention and follow-up for students at elevated risk are 0.37 and 0.21 respectively, none of them statistically significant.

Last, it was reported that 'a series of moderator analyses of student characteristics and programme features failed to predict treatment outcomes' (p. 295), however, the analyses were not shown, and it was not reported which student characteristics and programme features were used nor which model was used other than 'a mixed effects framework' was used.

**Fisak et al. (2011):**

The purpose of this review was to provide a comprehensive review of the effectiveness of child and adolescent (below the age of 18) anxiety prevention programmes, including universal as well as targeted (selective/indicative) programmes. Programmes in which depression or general stress management was the primary goal, and where anxiety was only measured as a secondary variable, were not eligible. The date of database search is not reported, but it is reported that the hand searches (of selected journals) were performed from 1970 to the end of 2009. A total of 30 studies (of which four were follow-up studies to previously published studies) with a comparison group was found, of which 10 were evaluating the FRIENDS intervention. Based on moderator analyses (single factor subgroups), it was found that studies utilising the FRIENDS programme were more effective than programmes not utilising FRIENDS; the weighted average effect of the FRIENDS intervention at post-intervention was 0.25 and statistically significant. The effect of the FRIENDS intervention at follow-up was not reported. As noted by the authors, more research is needed to determine the degree to which the effectiveness of the programme is generalisable to nations other than Australia (8 of the 10 evaluations on FRIENDS were performed in Australia). None of the effect sizes analysed were corrected for clustering at either the school nor the class level (if needed). The review included both universal prevention programmes and targeted (selective/indicative) prevention programmes.

**Briesch et al. (2010):**

This was a report on a literature search, conducted to identify all empirical studies of the FRIENDS programme published in peer-reviewed journals. The search was not described nor documented other than that they screened the list of research abstracts provided on the programme developers' website (no web address was reported), and conducted literature searches using the PsycINFO and MEDLINE databases. No search terms or date were reported. The review cannot be labelled systematic and further, no meta-analysis was performed. The authors reported the range of effect sizes in the studies (14 studies were found) and an average (simple average)

effect size but no other statistics such as standard errors, CIs or p-values.

We specifically searched the Cochrane Library for Cochrane systematic reviews and located one marginally relevant for the current review. James et al. (2005) and the updates James et al. (2015, 2020), examined the effect of CBT treatment interventions for childhood anxiety disorders. Eligible participants were children and adolescents younger than age 19, who met diagnostic criteria for an anxiety disorder diagnosis. Primary outcomes were remission of primary anxiety diagnosis post-treatment and number of participants lost to post-treatment assessment. Secondary outcomes included remission of all anxiety diagnoses, reduction in anxiety symptoms and depressive symptoms and improvements in global functioning. In the most updated (James et al., 2020), 87 studies were included, of which 5 evaluated FRIENDS. No separate analysis of FRIENDS was provided nor was FRIENDS included as a moderator.

Besides being up-to-date, a major difference between these systematic reviews and the current review is that we focused on the FRIENDS intervention delivered as universal, selective and indicated preventive programmes. We only included studies with a control group and with participants from at least two units (e.g., school or class) in each of the groups (treatment and control). All relevant outcome areas were analysed separately in a meta-analysis taking into consideration the unit of analysis (cluster or individual) and the dependencies between effect sizes. We were transparent concerning which studies and which measures were used in each analysis. In addition, the specific data used for any cluster correction and any moderator analyses were reported in detail.

Only three of the above-mentioned reviews reported specific results for the FRIENDS intervention (Ahlen et al., 2015; Fisak et al., 2011; Maggin & Johnson, 2014) of which one (Maggin & Johnson, 2014) only reported results separated by an unclear division into subgroups of risk. Our findings agree with the findings concerning overall weighted average effects reported in Ahlen et al. (2015) and Fisak et al. (2011).

#### Fisak et al. (2023):

This review, a systematic review with meta-analysis on all types of FRIENDS interventions, was published after we finished our search. Both RCTs as well as NRCTs of the FRIENDS programme with a simultaneously assessed comparison group where eligible. Further, eligible trials were required to have authorised and licensed use of the programme provided by the FRIENDS organisation. All types of programmes were included, that is, universal, selected, indicated and treatment programmes were eligible. The only participant restriction imposed was that samples were required to consist of predominantly neurotypical children and/or adolescents.

The date of database search is not reported. It is reported that they searched the database maintained by the FRIENDS research team. Further, a rather narrow search in the PsychInfo database was performed, resulting in only 68 hits of which three were additional to the search in the database maintained by the FRIENDS research team.

A total of 37 papers with 41 comparisons were included (25 based on RCTs and 16 based on NRCTs). Of these, four were treatment studies, a further five compared one unit to another unit (school or class) and one compared a programme with and without parent participation. Thus, these 10 studies were all excluded from our review. On the other hand, we located and included 7 studies not included in the Fisak et al. (2023) review.

The risk of bias was assessed using the same tools as in our review (ROB-2 tool for RCTs and the ROBINS-I tool for NRCTs). The risk of bias assessments is overall far more favourable than ours. The individual assessments of each outcome reported for a particular comparison are, however, not reported and neither are the comments or answers to signalling questions associated with each domain in the risk of bias tool. It is thus not possible for us to explain why the ratings differ so much from ours. An aggregate summary of risk of bias is reported and based on the ROB-2 tool, all RCTs were overall rated either low risk of bias or some concerns. More than 50% of the RCTs were overall rated low risk of bias even though only one RCT was rated low risk of bias in the measurement of outcome domain. This is not in accordance with the risk of bias tool used, as an outcome has to be rated low risk of bias in all domains to achieve an overall rating of low risk of bias. Concerning NRCTs, all were overall rated low or moderate risk of bias except one which was rated serious risk of bias.

Meta-analysis was performed using aggregated effect sizes for those comparisons where more than one measure was reported. The method of aggregation is not reported, neither is the level of correlation used, if any. The random effects average effect size for anxiety was -0.20 (note negative effects sizes favours treated) at post-intervention and -0.21 at 6–12 months follow-up, both are significant. We find a somewhat smaller average effect size at post (0.13) and a somewhat larger average at follow-up (0.31). At post there is more heterogeneity than we find ( $I^2$  is 84.49% compared to our  $I^2$  of 51%). They do not report any heterogeneity measures at follow-up. There are no analyses of parent reported measures as they (most likely) are included in their aggregated measure. Thus, there is some discrepancy in the results concerning the overall average effect. The reason may be due to the difference in studies included but also due to different approaches. The fact that we take into consideration that some studies use overlapping data (excluding some effect sizes) and that not all effect sizes are independent may produce different results from the ones reported in Fisak et al. (2023). In addition, we downgraded the evidence for all pooled effect sizes to a grading of very low certainty mainly because of risk of bias in the design of the studies, making us conclude that further research is very likely to have an important impact on our confidence in the effect estimate and may change the estimate. Such reservations are not present in Fisak et al. (2023) with their risk of bias assessments being very different from ours.

A series of moderator analyses of setting, student characteristics and programme features were performed, although only for the post-intervention effects. Although not reported, the moderator analyses were most likely single factor subgroup analyses (except the

continuous covariates average age and percent female were meta-regression was performed, most likely not multiple regression though). Two covariates reached significance, although analysed in only a subset of included programmes (universal programmes). A significantly larger average effect size was found for programmes facilitated by mental health professionals compared to non-mental health professionals (e.g., nurses, teachers, other school staff), and a significantly larger average effect size was also found for trials conducted by the FRIENDS Research team (all of which were in Australia) relative to trials led by other research teams (all of which were conducted in other countries).

In comparison, at post-intervention, in a multiple meta-regression model using the RVE procedure, we examined the impact of the moderators youth, Australia, teacher (provider is teacher vs. not teacher), booster sessions and no further implementation problems. We found that none of the moderators were statistically significant, except Australia, with the coefficient also being relatively large in relation to the baseline effect. Unfortunately, the adjusted degrees of freedom for Australia was just below 4 and thereby low enough that the RVE procedure we used may yield untrustworthy results regarding the standard errors for this covariate. At the 12 months follow-up we examined the impact of the moderators teacher, parent and no further implementation problems, also in a multiple meta-regression model using the RVE procedure. The meta-regression coefficient on teacher was negative, statistically significant and relatively large in relation to the baseline effect. This implies that programmes implemented by providers other than teachers (mental health provider or counsellor) performed better than programmes implemented by teachers.

Thus contrary to the results in Fisak et al. (2023), we can not be sure that the average effect of trials conducted in Australia (all involving the FRIENDS research team) is larger than the average effect of trials conducted outside of Australia and although we find programmes implemented by teachers are significantly smaller than the average effect of programmes implemented by non-teachers (mental health professionals or counsellors) this only applies to the 12 months follow-up and not to the post-intervention effect.

The reason for these different results may be that we take into consideration that some studies use overlapping data (excluding some effect sizes) and that not all effect sizes are independent. Further we perform multiple meta-regression where the impact of other moderators is taken into account.

## 8 | AUTHORS' CONCLUSIONS

### 8.1 | Implications for practice and policy

Our results indicate that the FRIENDS intervention may reduce anxiety symptoms in children and adolescent when reported by children and adolescents themselves. We found a positive and statistically significant average effect both in the short-term and at 12 months follow-up. We believe these average effect sizes are

of a meaningful magnitude as the intervention was primarily given as a universal intervention, that is, to children and adolescents with no immediate anxiety problems. The average effect at post-intervention corresponds to a 54% chance that a randomly selected score of a child from the treated population is better (note that a higher score for this outcome means fewer anxiety symptoms) than the score of a randomly selected child from the comparison population and increases to a 59% chance at 12 months follow-up. Note that these probabilities should be compared to a fifty-fifty chance, which is 50%, had the intervention been ineffective.

Heterogeneity was present in the analyses and the 95% prediction intervals included negative values, reflecting that the distribution of effect sizes spanned both positive and negative effect sizes. The probability of a negative effect size in a future trial was not high though. At post-intervention it was 19% and at 12 months follow-up it was 10%.

We investigated plausible moderators at both post-intervention and follow-up and only found evidence that programmes implemented by providers other than teachers (mental health provider or counsellor) may perform better than programmes implemented by teachers at follow-up.

The available evidence does not suggest that the FRIENDS programme reduces anxiety symptoms in children and adolescents when reported by parents. However, given the relatively few comparisons, some caution is needed in making an assumption that the FRIENDS programme do not decrease children and adolescents' anxiety symptoms when measured by parents.

Few studies reported on the long-term effects of the FRIENDS intervention. Only half the number of studies reporting post-intervention effects reported effects at 12 months follow-up, and even fewer reported on follow-up effects beyond 12 months. Our findings suggest that the FRIENDS intervention increase the reduction in anxiety symptoms 12 months after the intervention. This may have important implications for practice, as there is a need to continuously evaluate the sustainability of treatment results. The evidence was inconclusive beyond 12 months follow-up due to few studies.

### 8.2 | Implications for research

Further research is required to fully address the effects of the FRIENDS intervention in reducing anxiety symptoms in children and adolescent. The overall certainty of evidence varied from low to very low. The GRADE evidence profile indicates that the certainty of evidence was low for all pooled effect sizes except the self-esteem at post-intervention which was very low.

As it is not possible to blind participants or the provider of the intervention in these kinds of studies, it is not possible for a comparison to achieve an overall low risk of bias. The highest overall risk of bias rating a comparison can acquire is Some concerns, which was acquired by only six randomised comparisons and five of them had issues on several domains. Only one trial rated overall Some concerns

was rated Low risk of bias on all domains except in the domain Deviations from intervention where the only concern was lack of blinding of participants and the provider of the intervention.

These considerations point to the need for more rigorously conducted studies.

The majority of trials employed a waitlist design, implying only a few studies reported on the long-term effects of the FRIENDS intervention. Our findings suggest that the FRIENDS intervention increase the reduction in anxiety symptoms up to 12 months after the intervention. This emphasises the need for future research not using waitlist design so more long-run follow-up effects are available.

## ACKNOWLEDGEMENTS

We would like to thank Jette Ní Mhaolcatha for being so accommodating and providing invaluable help with describing the intervention.

## CONTRIBUTIONS OF AUTHORS

- Content: Trine Filges, Geir Smedslund, Tine L. Mundbjerg Eriksen
- Systematic review methods: Trine Filges, Geir Smedslund, Malene Wallach Kildemoes
- Statistical analysis: Trine Filges, Geir Smedslund, Tine L. Mundbjerg EriksenInformation
- Information retrieval: Kirsten Birkefoss

## DECLARATIONS OF INTEREST

None.

## SOURCES OF SUPPORT

### Internal sources

- VIVE Campbell, Denmark.

### External sources

- New Source of support, Other.

## DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We modified the time points for analysis slightly.

We did not use the GRADEpro GDT software (available at <https://gdt.graderpro.org/app/>) to produce a summary of findings table, as stated in the protocol as not all analyses were performed in RevMan.

## REFERENCES

### REFERENCES TO STUDIES INCLUDED IN THIS REVIEW

#### Ahlen 2018 {published data only}

- \*Ahlen, J., Hursti, T., Tanner, L., Tokay, Z., & Ghaderi, A. (2018). Prevention of anxiety and depression in Swedish school children: A cluster-randomized effectiveness study. *Prevention Science*, 19(2), 147–158. <https://doi.org/10.1007/s11121-017-0821-1>

#### Ahlen 2019 {published data only}

- \*Ahlen, J., Lenhard, F., & Ghaderi, A. (2019). Long term outcomes of a cluster randomized universal prevention trial targeting anxiety and depression in school children. *Behavior Therapy*, 1, 200–213.

#### Ahmarijan 2020 {published data only}

- \*Ahmarijan, F., Khodabakhshi-Koolaee, A., Taghvaei, D., & Ahghar, G. (2020). The effect of "FRIENDS for Life" and "Coping Cat" programs for reduction of anxiety symptoms on male children. *Caspian Journal of Pediatrics*, 6(2), 434–441.

#### Anticich 2013 {published data only}

- \*Anticich, S. J., Barrett, P. M., Silverman, W., Lacherez, P., & Gillies, R. (2013). The prevention of childhood anxiety and promotion of resilience among preschool-aged children: A universal school-based trial. *Advances in School Mental Health Promotion*, 6, 93–21. <https://doi.org/10.1080/1754730X.2013.784616>

#### Barker 2008 {published data only}

- \*Barker, L. J. (2008). *Preventing anxiety disorders in youth: universal school-based intervention* [PhD Thesis, University of British Columbia].

#### Barrett 2000 {published data only}

- Barrett, P., Moore, A., & Sonderegger, R. (2000). The FRIENDS program for young Former-Yugoslavian refugees in Australia: A pilot study. *Behaviour Change*, 17. <https://doi.org/10.1375/bech.17.3.124>

#### Barrett 2001 {published data only}

- \*Barrett, P. M., Sonderegger, R., & Sonderegger, S. L. (2001). Evaluation of an anxiety-prevention and positive-coping program (FRIENDS) for children and adolescents of non-English speaking backgrounds. *Behaviour Change*, 18, 78–91.

#### Barrett 2003 {published data only}

- \*Barrett, P. M., Sonderegger, R., & Xenos, S. (2003). Using FRIENDS to combat anxiety and adjustment problems among young migrants to Australia: A national trial. *Clinical Child Psychology and Psychiatry*, 8, 241–260.

#### Barrett 2005 {published data only}

- \*Barrett, P. M., Lock, S., & Farrell, L. J. (2005). Developmental differences in universal preventive intervention for child anxiety. *Clinical Child Psychology and Psychiatry*, 10, 539–555.

#### Barrett and Turner 2001 {published data only}

- \*Barrett, P. M., & Turner, C. (2001). Prevention of anxiety symptoms in primary school children: Preliminary results from a universal school-based trial. *British Journal of Clinical Psychology*, 40, 399–410.

#### Cooley-Strickland 2011 {published data only}

- \*Cooley-Strickland, M. R., Griffin, R. S., Darney, D., Otte, K., & Ko, J. (2011). Urban African American youth exposed to community violence: A school-based anxiety preventive intervention efficacy study. *Journal of Prevention and Intervention in the Community*, 39, 149–166.

### Doyle 2016 {published data only}

Doyle, S. (2016). *Child anxiety, resiliency, and the FRIENDS for Life program* [Thesis]. Department of Psychology Memorial University of Newfoundland.

### Essau 2012 {published data only}

\*Essau, C. A., Conradt, J., Sasagawa, S., & Ollendick, T. H. (2012). Prevention of anxiety symptoms in children: Results from a universal school trial. *Behavior Therapy*, 43, 450–464. <https://doi.org/10.1016/j.beth.2011.08.003>

### Gallegos 2008 {published data only}

Gallegos, J. (2008). *Preventing childhood anxiety and depression: Testing the effectiveness of a school-based program in Mexico* [Doctoral dissertation, University of Texas at Austin]. Available from ProQuest Dissertations and Thesis database (UMI No. 3341564)/Psicología Educativa 19 (pp. 37–44).

### Gallegos 2013 {published data only}

\*Gallegos, J., Linan-Thompson, S., Stark, K., & Ruvalcaba, N. (2013). Preventing childhood anxiety and depression: Testing the effectiveness of a schoolbased program in Mexico. *Psicología Educativa*, 19(1), 37–44. <https://doi.org/10.5093/ed2013a6>

### Ghajareih 2018 {published data only}

\*Ghajareih, M., Allipour, S., shehni yailagh, M., & Davoodi, I. (2018). The effects of the FRIENDS for youth program on anxiety of Ahvaz female students. *Psychological Methods and Models*, 9(31), 69–84.

### Haidari 2016 {published data only}

\*Haidari, M., Bakhtiyarpoor, S., Makvandi, B., Naderi, F., & hafezi, F. (2016). Investigating the effectiveness of FRIENDS program training on the anxiety of children in shiraz. *Psychological Methods and Models*, 7(24), 23–38.

### Hunt 2009 {published data only}

\*Hunt, C., Andrews, G., Sakashita, C., Crino, R., & Erskine, A. (2009). Randomized controlled trial of an early intervention programme for adolescent anxiety disorders. *Australian & New Zealand Journal of Psychiatry*, 43(4), 300–304. <https://doi.org/10.1080/00048670902721152>

### Kösters 2015 {published data only}

\*Kösters, M. P., Chinapaw, M. J., Zwaanswijk, M., Van der Wal, M. F., & Koot, H. M. (2015). Indicated prevention of childhood anxiety and depression: Results from a practice-based study up to 12 months after intervention. *American Journal of Public Health*, 105, 2005–2013. <https://doi.org/10.2105/AJPH.2015.302742>

Kösters, M. P., Chinapaw, M. J., Zwaanswijk, M., van der Wal, M. F., Utens, E. M. W. J., & Koot, H. M. (2012). Study design of 'FRIENDS for Life': Process and effect evaluation of an indicated school-based prevention programme for childhood anxiety and

depression. *BMC Public Health*, 12, 86. <https://doi.org/10.1186/1471-2458-12-86>

### Kozina 2020 {published data only}

\*Kozina, A. (2020). School-Based prevention of anxiety using the "My FRIENDS" emotional resilience program: Six month follow-up. *International Journal of Psychology*, 55(S1), 70–77. <https://doi.org/10.1002/ijop.12553>

### Kozina 2021 {published data only}

Kozina, A. (2021). Can FRIENDS for Life social-emotional learning programme be used for preventing anxiety and aggression in a school environment: 6 months, 1-year and 1-and-a-half-year follow-up. *European Journal of Developmental Psychology*, 18(2), 214–229. <https://doi.org/10.1080/17405629.2020.1776103>

### Lawson 2023 {published data only}

Eiraldi, R., Khanna, M. S., Jawad, A. F., Fishman, J., Glick, H. A., Schwartz, B. S., Cacia, J., Wandersman, A., & Beidas, R. (2016). A hybrid effectiveness-implementation cluster randomized trial of group CBT for anxiety in urban schools: Rationale, design, and methods. *Implementation Science*, 11, 92. <https://doi.org/10.1186/s13012-016-0453-z>

\*Lawson, G. M., Jawad, A. F., Comly, R., Khanna, M., Glick, H. A., Beidas, R. S., Goldstein, J., Brizzolara-Dove, S., Wilson, T., Rabenau-McDonnell, Q., & Eiraldi, R. (2023). A comparison of two group cognitive behavioral therapy protocols for anxiety in urban schools: Appropriateness, child outcomes, and cost-effectiveness. *Frontiers in Psychiatry*, 14, 1105630. <https://doi.org/10.3389/fpsy.2023.1105630>

### Liddle 2010 {published data only}

\*Liddle, I., & Macmillan, S. (2010). Evaluating the FRIENDS programme in a Scottish setting. *Educational Psychology in Practice*, 26, 53–67.

### Lock 2003 {published data only}

\*Lock, S., & Barrett, P. M. (2003). A longitudinal study of developmental differences in universal preventive intervention for child anxiety. *Behaviour Change*, 20, 183–199.

### Lowry-Webster 2001 {published data only}

\*Lowry-Webster, H. M., Barrett, P. M., & Dadds, M. R. (2001). A universal prevention trial of anxiety and depressive symptomatology in childhood: Preliminary data from an Australian study. *Behaviour Change*, 18, 36–50.

Lowry-Webster, H. M. (2002). *An Ecological Approach to the Prevention of Anxiety Disorders during Childhood* [PhD Thesis School of Applied Psychology, Griffith University]. <https://doi.org/10.25904/1912/2035>

### Lowry-Webster 2003 {published data only}

\*Lowry-Webster, H. M., Barrett, P. M., & Lock, S. (2003). A universal prevention trial of anxiety symptomatology during childhood: Results at a 1-year follow up. *Behaviour Change*, 20, 25–43.

**Maalouf 2020 {published data only}**

\*Maalouf, F. T., Alrojolah, L., Ghandour, L., Afifi, R., Dirani, L. A., Barrett, P., Nakkash, R., Shamseddeen, W., Tabaja, F., Yuen, C. M., & Becker, A. E. (2020). Building emotional resilience in youth in Lebanon: A school-based randomized controlled trial of the FRIENDS intervention. *Prevention Science*, 22(5), 650–660. <https://doi.org/10.1007/s11121-02001123-5>

**Miller 2011a {published data only}**

\*Miller, L. D., Laye-Gindhu, A., Liu, Y., March, J. S., Thordarson, D. S., & Garland, E. J. (2011). Evaluation of a preventive intervention for child anxiety in two randomized attention-control school trials. *Behaviour Research and Therapy*, 49(5), 315–323. <https://doi.org/10.1016/j.brat.2011.02.006>

**Miller 2011b {published data only}**

\*Miller, L. D., Laye-Gindhu, A., Bennett, J. L., Liu, Y., Gold, S., March, J. S., Olson, B. F., & Waechtler, V. E. (2011). An effectiveness study of a culturally enriched school-based CBT anxiety prevention program. *Journal of Clinical Child and Adolescent Psychology*, 40(4), 618–629. <https://doi.org/10.1080/15374416.2011.581619>

**Moharreri 2017 {published data only}**

\*Moharreri, F., & Yazdi, A.-S. H. (2017). Evaluation of the effectiveness of the friends for life program on children's anxiety and depression. *Iranian Journal of Psychiatry*, 12(4), 269–277.

Rahaleh, N. (2017). Evaluation of the Effectivenesseffectiveness of the "Friends for Life" program on children's anxiety and Depressiondepression. Clinical Trial Protocol Iranian Registry of Clinical Trials.

**Pahl 2010 {published data only}**

Pahl, K. M. (2009). *Preventing anxiety and promoting social and emotional strength in early childhood: An investigation of aetiological risk factors* [Doctoral dissertation. University of Queensland, Australia].

\*Pahl, K. M., & Barrett, P. M. (2010). Preventing anxiety and promoting social and emotional strength in preschool children: A universal evaluation of the Fun FRIENDS program. *Advances in School Mental Health Promotion*, 3, 14–25. <https://doi.org/10.1080/1754730X.2010.9715683>

**Rivero 2020 {published data only}**

\*Rivero, L. M. H. N., Andrade, A. L. M., Figueiredo, L. Z. P., Pinheiro, B. O., & Micheli, D. (2020). Evaluation of FunFRIENDS program in prevention of anxiety in Brazilian children: A randomized controlled pilot trial. *Cien Saude Colet*, 25(11), 4497–4508. <https://doi.org/10.1590/1413-812320202511.33072018>

**Rodgers 2009 {published data only}**

\*Rodgers, A., & Dunsmuir, S. (2015). A controlled evaluation of the FRIENDS for Life emotional resiliency programme on overall anxiety levels, anxiety subtype levels and school adjustment. *Child and Adolescent Mental Health*, 20, 13–19. <https://doi.org/10.1111/camh.12030>

**Rutledge 2016 {published data only}**

\*Rutledge, R., Devitt, E., Greene, G., Mullany, M., Charles, E., Frehill, J., & Moriarty, M. (2016). A randomised controlled trial of the FRIENDS for Life emotional resilience programme delivered by teachers in Irish primary schools. *Educational and Child Psychology*, 33, 69–89.

**Sabey 2019 {published data only}**

\*Sabey, P. B., Zelman, D. C., & Tori, C. D. (2019). Chapter: 19. Evaluation of the FRIENDS Intervention for Childhood Anxiety in Hong Kong. In A. Shukla, A. Dubey, & N. S. Thagunna (Eds.), *The psyche of Asian society* (pp. 310–328). Concept Publishers.

**Siu 2007 {published data only}**

\*Siu, A. F. Y. (2007). Using friends to combat internalizing problems among primary school children in Hong Kong. *Journal of Cognitive and Behavioral Psychotherapies*, 7(1), 11–26.

**Stallard 2015 {published data only}**

Skryabina, E., Taylor, G., & Stallard, P. (2016). Effect of a universal anxiety prevention programme (FRIENDS) on children's academic performance: results from a randomised controlled trial. *The Journal of Child Psychology and Psychiatry*, 57(11), 1297–1307. <https://doi.org/10.1111/jcpp.12593>

\*Stallard, P., Skryabina, E., Taylor, G., Anderson, R., Ukoumunne, O. C., Daniels, H., Phillips, R., & Simpson, N. (2015). A cluster randomised controlled trial comparing the effectiveness and cost-effectiveness of a school-based cognitive-behavioural therapy programme (FRIENDS) in the reduction of anxiety and improvement in mood in children aged 9/10 years. *Public Health Research*, 3(14), 1–115. <https://doi.org/10.3310/phr03140>

Stallard, P., Skryabina, E., Taylor, G., Phillips, R., Daniels, H., Anderson, R., & Simpson, N. (2014). Classroom-based cognitive behaviour therapy (FRIENDS): A cluster randomised controlled trial to Prevent Anxiety in Children through Education in Schools (PACES). *The Lancet Psychiatry*, 1(3), 185–192. [https://doi.org/10.1016/S22150366\(14\)70244-5](https://doi.org/10.1016/S22150366(14)70244-5)

Stallard, P., Taylor, G., Anderson, R., Daniels, H., Simpson, N., Phillips, R., & Skryabina, E. (2012). School-based intervention to reduce anxiety in children: Study protocol for a randomized controlled trial (PACES). *Trials*, 13, 227. <https://doi.org/10.1186/1745-6215-13-227>

Stallard, P., Taylor, G., Anderson, R., Daniels, H., Simpson, N., Phillips, R., & Skryabina, E. (2014). The prevention of anxiety in children through school-based interventions: Study protocol for a 24-month follow-up of the PACES project. *Trials*, 15, 77. <https://doi.org/10.1186/1745-6215-15-77>

**Wigelsworth 2018 {published data only}**

Wigelsworth, M. (2017). *FRIENDS for life study – Statistical analysis plan*. The University of Manchester.

Wigelsworth, M., Squires, G., & Birchall, E. (2015). *The FRIENDS Programme: An evaluation of academic and emotional health outcomes. An outline evaluation strategy submitted to the Education Endowment Foundation*. The University of Manchester <https://doi.org/10.1186/ISRCTN13721202>.

\*Wigelsworth, M., Squires, G., Birchinall, E., Kalambouka, A., Lendrum, A., Black, L., Troncoso, P., Santos, J., Ashworth, E., & Britteon, P. (2018). *FRIENDS for life: Evaluation report and executive summary*. The University of Manchester.

## REFERENCES TO STUDIES EXCLUDED FROM THIS REVIEW

### Ahlen 2012 {published data only}

Ahlen, J., Breitholtz, E., Barrett, P. M., & Gallegos, J. (2012). School-based prevention of anxiety and depression: A pilot study in Sweden. *Advances in School Mental Health Promotion*, 5(4), 246–257. <https://doi.org/10.1080/1754730X.2012.730352>

### Balle 2010 {published data only}

Balle, M., & Tortella-Feliu, M. (2010). Efficacy of a brief school-based program for selective prevention of childhood anxiety. *Anxiety, Stress and Coping: An International Journal*, 23, 71–85. <https://doi.org/10.1080/10615800802590652>

### Barrett 2006 {published data only}

Barrett, P. M., Farrell, L. J., Ollendick, T. H., & Dadds, M. (2006). Long-term outcomes of an Australian universal prevention trial of anxiety and depression symptoms in children and youth: An evaluation of the FRIENDS program. *Journal of Clinical Child and Adolescent Psychology*, 35, 403–411.

### Barrett 2015 {published data only}

Barrett, P., Fisak, B., & Cooper, M. (2015). The treatment of anxiety in young children: Results of an open trial of the Fun FRIENDS program. *Behaviour Change*, 32, 1–12. <https://doi.org/10.1017/bec.2015.12>

### Bernstein 2005 {published data only}

Bernstein, G. A., Layne, A. E., Egan, E. A., & Tennison, D. M. (2005). School-based interventions for anxious children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44, 1118–1127.

Bernstein, G., Bernat, D., Victor, A., & Layne, A. (2008). School-based interventions for anxious children: 3-, 6-, and 12-month follow-ups. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47, 1039–1047. <https://doi.org/10.1097/CHI.0b013e31817eecc0>

### Carlyle 2014 {published data only}

Carlyle, D. (2014). With a little help from FUN FRIENDS young children can overcome anxiety. *Community Practitioner*, 87, 26–29.

### Cooley 2004 {published data only}

Cooley, M. R., Boyd, R. C., & Grados, J. J. (2004). Feasibility of an anxiety preventive intervention for community violence exposed African American children. *Journal of Primary Prevention*, 25, 105–123.

### Desousa 2016 {published data only}

Desousa, D., Vilanova, F., Ecker, K., Koller, S., Salcides, A., & Petersen, C. (2016). Community trial of the FRIENDS for life prevention program with children at risk for internalizing disorders. *Clinica & Cultura*, 5, 87–106.

### Dohl 2013 {published data only}

Dohl, A. H. (2013). *Managing anxiety through childhood social-emotional development* [Master's thesis, School of Psychology, University of British Columbia].

### Doyle 2020 {published data only}

Doyle, S., Francis, S., & Joy, R. (2020). Treatment integrity and social validity of the FRIENDS for life programme in a northeastern Canadian school system. *Behaviour Change*, 37, 1–19. <https://doi.org/10.1017/bec.2020.4>

### Eiraldi 2023 {published data only}

Eiraldi, R., Lawson, G. M., Glick, H. A., Khanna, M. S., Beidas, R., Fishman, J., Rabenau-McDonnell, Q., Wilson, T., Comly, R., Schwartz, B. S., & Jawad, A. F. (2023). Implementation fidelity, student outcomes, and cost-effectiveness of train-the-trainer strategies for Masters-level therapists in urban schools: A cluster randomized trial (Under review). <https://doi.org/10.21203/rs.3.rs-2540115/v1>

### Farrell 2005 {published data only}

Farrell, L., Barrett, P., & Claassens, S. (2005). Community trial of an evidence-based anxiety intervention for children and adolescents (the FRIENDS Program): A pilot study. *Behaviour Change*, 22, 236–248. <https://doi.org/10.1375/bech.22.4.236>

### Fisak 2018 {published data only}

Fisak, B., Gallegos, J., Verreyne, M., & Barrett, P. (2018). The results of a targeted open trial of the Fun FRIENDS combined with a concurrent parent-based intervention. *Mental Health & Prevention*, 10, 35–41. <https://doi.org/10.1016/j.mhp.2018.03.001>

### Fjermestad 2020 {published data only}

Fjermestad, K. W., Wergeland, G. J., Rogde, A., Bjaastad, J. F., Heiervang, E., & Haugland, B. S. M. (2020). School-based targeted prevention compared to specialist mental health treatment for youth anxiety. *Child and Adolescent Mental Health*, 25(2), 102–109. <https://doi.org/10.1111/camh.12366>

### Fukushima-Flores 2011 {published data only}

Fukushima-Flores, M., & Miller, L. (2011). FRIENDS parent project: Effectiveness of parent training in reducing parent anxiety in a universal prevention program for anxiety symptoms in school children. *Behaviour Change*, 28(2), 57–73.

### Gallegos 2012 {published data only}

Gallegos, J., Rodríguez, A., Gómez, G., Rabelo, M., & Gutiérrez, M. (2012). The FRIENDS for life program for Mexican girls living in an

- orphanage: A pilot study. *Behaviour Change*, 29(1), 1–14. <https://doi.org/10.1017/bec.2012.8>
- Gallegos 2015 {published data only}**  
 Gallegos, J., Ruvalcaba Romero, N., Langley, A., & Villegas, D. (2015). Selective prevention for anxiety and resilience promotion: Outcomes of an anxiety prevention and resilience program with girls at risk. *Pensando Psicología*, 11, 11–23. <https://doi.org/10.16925/pe.v11i18.1001>
- Gallegos-Guajardo 2020 {published data only}**  
 Gallegos-Guajardo, J., Ruvalcaba-Romero, N. A., Chavez-Amavizca, A., & Barrett, P. (2020). Evaluation of the Fun FRIENDS program with a sample of Mexican children. *Pensando Psicología*, 16(1), 1–21. <https://doi.org/10.16925/2382-3984.2020.01.03>
- Games 2020 {published data only}**  
 Games, N., Thompson, C. L., & Barrett, P. (2020). A randomised controlled trial of the Adult Resilience Program: A universal prevention program. *International Journal of Psychology: Journal international de psychologie*, 55(Suppl. 1), 78–87. <https://doi.org/10.1002/ijop.12587>
- Garcia 2019 {published data only}**  
 Rinaldin Garcia, L., Salvo Toni, C., Batista, A., & Zeggio, L. (2019). Evaluation of the effectiveness of the Fun FRIENDS program. *Trends in Psychology*, 27, 925–941. <https://doi.org/10.9788/tp2019.4-08>
- Green 2013 {published data only}**  
 Green, S. (2013). An Evaluation of a FRIENDS for Life Programme in a mainstream secondary school and its impact on emotional distress, anxiety and coping skills [PhD Thesis, University of Nottingham].
- Hosokawa 2023 {published data only}**  
 Hosokawa, R., Matsumoto, Y., Nishida, C., Funato, K., & Mitani, A. (2023). Evaluating the effectiveness of a social and emotional learning program among preschool children in Japan: An experimental cohort study. *Child and Adolescent Psychiatry and Mental Health*, 17, 93. <https://doi.org/10.1186/s13034-023-00643-6>
- Iizuka 2014a {published data only}**  
 Iizuka, C., Barrett, P., Gillies, R., Cook, C., & Miller, D. (2014). The FRIENDS emotional health program for minority groups at risk. *Journal of School Health*, 84, 124–132. <https://doi.org/10.1111/josh.12127>
- Iizuka 2014b {published data only}**  
 Iizuka, C., Barrett, P., Gillies, R., Cook, C., & Marinovic, W. (2014). A combined intervention targeting both teachers' and students' social-emotional skills: Preliminary evaluation of students' outcomes. *Australian Journal of Guidance and Counselling*, 24(2), 152–166.
- Iizuka 2015 {published data only}**  
 Iizuka, C., Barrett, P., Gillies, R., Cook, C., & Marinovic, W. (2015). Preliminary evaluation of the FRIENDS for life program on students' and teachers' emotional states for a school in a low socio-economic status area. *Australian Journal of Teacher Education*, 40, 1–21. <https://doi.org/10.14221/ajte.2014v40n3.1>
- Kato 2017 {published data only}**  
 Kato, S., & Shimizu, E. (2017). A pilot study on the effectiveness of a school-based cognitive-behavioral anxiety intervention for 8- and 9-year-old children: A controlled trial in Japan. *Mental Health and Prevention*, 8, 32–38. <https://doi.org/10.1016/j.mhp.2017.10.001>
- Kavanagh 2014 {published data only}**  
 Kavanagh, C. (2014). Effects of the FRIENDS for Life program on anxiety, school functioning, and social functioning. A Thesis Submitted to the Faculty of Graduate and Postdoctoral Studies University of Ottawa.
- LaRose 2017 {published data only}**  
 LaRose, C. N. (2017). *A proposal to study the effectiveness of the Friends for Life program* [Master thesis, University of Lethbridge, Canada].
- Lewis 2016 {published data only}**  
 Lewis, K. M. (2016). An ounce of prevention: Evaluation of the Fun FRIENDS program for kindergarteners in a rural school. In *Dissertation Abstracts International: Section B: The Sciences and Engineering* (Vol. 76, Issue 10-B(E)).
- Lewis 2023 {published data only}**  
 Lewis, K. M., Barrett, P., Freitag, G., & Ollendick, T. H. (2023). An ounce of prevention: Building resilience and targeting anxiety in young children. *Clinical Child Psychology and Psychiatry*, 28(2), 795–809. <https://doi.org/10.1177/13591045221121595>
- Martinsen 2009 {published data only}**  
 Martinsen, K., Aalberg, M., Narayanan, M., & Neumer, S.-P. (2009). Using a structured treatment, Friends for Life, in Norwegian outpatient clinics: Results from a pilot study. *The Cognitive Behaviour Therapist*, 2, 10–19. <https://doi.org/10.1017/S1754470X08000160>
- Matsumoto 2016 {published data only}**  
 Matsumoto, Y., & Shimizu, E. (2016). The FRIENDS cognitive behavioral program in Japanese schools: An examination of the treatment effects. *School Psychology International*, 37(4), 397–409. <https://doi.org/10.1177/0143034316649639>
- Mims 2015 {published data only}**  
 Mims, C. (2015). *Effectiveness of the friends program in reducing anxiety symptoms among female high school students* [PhD Thesis, WILLIAM JAMES COLLEGE].

**Moharer 2021 {published data only}**

Moharer, G. S., Tabatabaeineya, S. H., Dovvom, M. H. Z., & Jafari, H. (2021). The effect FRIENDS program on the anxiety of adolescent girls. *Social Determinants of Health*, 7, 1–7.

Y., Pine, D. S., Koller, S. H., & Manfro, G. G. (2018). Group cognitive behavioral therapy and attention bias modification for childhood anxiety disorders: A factorial randomized trial of efficacy. *Journal of Child and Adolescent Psychopharmacology*, 28(9), 620–630. <https://doi.org/10.1089/cap.2018.0022>

**Mostert 2008 {published data only}**

- Mostert, J. (2007). Pilot study of the effectiveness of the FRIENDS programme in a South African sample [PhD Thesis, University of Stellenbosch].
- \*Mostert, J., & Loxon, H. (2008). Exploring the effectiveness of the FRIENDS program in reducing anxiety symptoms among South African children. *Behaviour Change*, 25(2), 85–96.

**O'Brien 2007 {published data only}**

O'Brien, F., Olden, N., Migone, M., Dooley, B., Atkins, L., Ganter, K., & Bates, A. (2007). Group cognitive behavioural therapy for children with anxiety disorder – An evaluation of the 'Friends for Youth' programme. *Irish Journal of Psychological Medicine*, 24(1), 5–12.

**Onge 2016 {published data only}**

Onge, J. R. S., Stephenson, R., & Kumar Bindu, S. (2016). Validation of the FRIENDS anxiety prevention program for children in Canada. *Canadian Journal of Community Mental Health*, 35(3), 1–16. <https://doi.org/10.7870/cjcmh-2016-036>

**Paul 2011 {published data only}**

Paul, J. (2011). Evaluating the impact of a universal FRIENDS for Life Programme on emotional distress and academic self-perceptions. <https://eprints.nottingham.ac.uk/12978/1/555382.pdf>

**Pereira 2014 {published data only}**

Pereira, A. I., Marques, T., Russo, V., Barros, L., & Barrett, P. (2014). Effectiveness of the friends for life program in Portuguese schools: Study with a sample of highly anxious children. *Psychology in the Schools*, 51(6), 647–657. <https://doi.org/10.1002/pits.21767>

**Rose 2009 {published data only}**

Rose, H., Miller, L., & Martinez, Y. (2009). FRIENDS for life: The results of a resilience-building, anxiety-prevention program in a Canadian elementary school. *Professional School Counseling*, 12(6), 400–407. <https://doi.org/10.5330/PSC.n.2010-12.400>

**Salari 2018 {published data only}**

Salari, E., Shahrivar, Z., Mahmoudi-Gharaei, J., Shirazi, E., & Sepasi, M. (2018). Parent-only group cognitive behavioral intervention for children with anxiety disorders: A control group study. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 27(2), 130–136.

**Salum 2018 {published data only}**

Salum, G. A., Petersen, C. S., Jarros, R. B., Toazza, R., DeSousa, D., Borba, L. N., Castro, S., Gallegos, J., Barrett, P., Abend, R., Bar-Haim,

**Shortt 2001 {published data only}**

- Barrett, P. M., Shortt, A. L., & Wescombe, K. (2001). Examining the social validity of the FRIENDS treatment program for anxious children. *Behaviour Change*, 18(2), 63–77. <https://doi.org/10.1375/bech.18.2.63>
- \*Shortt, A. L., Barrett, P. M., & Fox, T. L. (2001). Evaluating the FRIENDS program: A cognitive-behavioral group treatment for anxious children and their parents. *Journal of Clinical Child and Adolescent Psychology*, 30(4), 525–535. [https://doi.org/10.1207/S15374424JCCP3004\\_09](https://doi.org/10.1207/S15374424JCCP3004_09)

**Stallard 2005 {published data only}**

- Stallard, P., Simpson, N., Anderson, S., & Goddard, M. (2008). The FRIENDS emotional health prevention programme: 12 month follow-up of a universal UK school based trial. *European Journal of Child and Adolescent Psychiatry*, 17, 283–289.
- \*Stallard, P., Simpson, N., Anderson, S., Carter, T., Osborn, C., & Bush, S. (2005). An evaluation of the FRIENDS programme: A cognitive behavior therapy intervention to promote emotional resilience. *Archives of Disease in Childhood*, 90, 1016–1019.

**Stallard 2007 {published data only}**

Stallard, P., Simpson, N., Anderson, S., Hibbert, S., & Osborn, C. (2007). The FRIENDS emotional health programme: Initial findings from a school-based project. *Child and Adolescent Mental Health*, 12, 32–37.

**Stopa 2010 {published data only}**

Stopa, J. E., Barrett, P. M., & Golingi, F. (2010). The prevention of childhood anxiety in socioeconomically disadvantaged communities: A universal school-based trial. *Advances in School Mental Health Promotion*, 3(4), 5–24. <https://doi.org/10.1080/1754730X.2010.9715688>

**Sykes 2009 {published data only}**

Sykes, M. (2009). An evaluation of the FRIENDS program: A therapeutic intervention for anxious young people [PhD Thesis, University of Exeter].

**van der Mheen 2019 {published data only}**

van der Mheen, M., Legerstee, J., Dieleman, G., Hillegers, M., & Utens, E. (2019). Cognitive behavioural therapy for anxiety disorders in young children: A Dutch open trial of the Fun FRIENDS program. *Behaviour Change*, 37, 1–12. <https://doi.org/10.1017/bec.2019.16>

**Wergeland 2014 {published data only}**

Wergeland, G. J., Fjermestad, K. W., Marin, C. E., Haugland, B. S., Bjaastad, J. F., Oeding, K., Bjelland, I., Silverman, W. K., Öst, L.-S.,

- Havik, O. E., & Heiervang, E. R. (2014). An effectiveness study of individual vs. group cognitive behavioral therapy for anxiety disorders in youth. *Behaviour Research and Therapy*, 57, 1–12. <https://doi.org/10.1016/j.brat.2014.03.007>
- ADDITIONAL REFERENCES**
- Ahlen 2015**  
Ahlen, J., Lenhard, F., & Ghaderi, A. (2015). Universal prevention for anxiety and depressive symptoms in children: A meta-analysis of randomized and cluster-randomized trials. *Journal of Primary Prevention*, 36(6), 387–403. <https://doi.org/10.1007/s10935-015-0405-4>
- Angrist and Pischke 2009**  
Angrist, J. D., & Pischke, J. S. (2009). Princeton University Press.
- Barican 2022**  
Barican, J. L., Yung, D., Schwartz, C., Zheng, Y., Georgiades, K., & Waddell, C. (2022). Prevalence of childhood mental disorders in high-income countries: A systematic review and meta-analysis to inform policymaking. *Evidence Based Mental Health*, 25, 36–44. <https://doi.org/10.1136/ebmental-2021-300277>
- Barrett and May 2007**  
Ahlen 2019Barrett, P. M., & May, S. (2007). Introduction to FRIENDS: Anxiety Prevention and Treatment for children aged 7–11 and youth aged 12–16. Retrieved from <http://www.mentalhealthpromotion.net/resources/friendsintroduction.pdf>
- Barrett 2014**  
Barrett, P. M., Cooper, M., & Guajardo, J. G. (2014). Using the FRIENDS programs to promote resilience in cross-cultural populations. In S. Prince-Embury, & D. H. Saklofske (Eds.), *Resilience interventions for youth in diverse populations* (pp. 85–108). Springer. <https://doi.org/10.1007/978-1-4939-0542-3-5>
- Barrett 2017**  
Barrett, P., Cooper, M., Stallard, P., Zeggio, L., & Gallegos-Guajardo, J. (2017). Effective evaluation of the FRIENDS anxiety prevention program in school settings: A response to Maggin and Johnson. *Education and Treatment of Children*, 40(1), 97–110.
- Barrett 1999**  
Barrett, P. (1999). Interventions for child and youth anxiety disorders: Involving parents, teachers, and peers. *The Australian Educational and Developmental Psychologist*, 16(1), 5–24. <https://doi.org/10.1017/S0816512200050045>
- Barrett 2020**  
Barrett, P. M. (2020). *FRIENDS for life: Facilitator manual*. Friends Resilience Pty Ltd.
- Battle 1992**  
Battle, J. (1992). *Culture-free self esteem inventories* (2 edition.). ProEd.
- Baxter 2012**  
Baxter, A., Scott, K., Vos, T., & Whiteford, H. (2012). Global prevalence of anxiety disorders: A systematic review and meta-regression. *Psychological Medicine*, 43, 1–14. <https://doi.org/10.1017/S003329171200147X>
- Beesdo 2009**  
Beesdo, K., Knappe, S., & Pine, D. S. (2009). Anxiety and anxiety disorders in children and adolescents: developmental issues and implications for DSM-V. *Psychiatric Clinics*, 32(3), 483–524.
- Bell and McCaffrey 2002**  
Bell, R. M., & McCaffrey, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, 28(2), 169–181.
- Biddle 2004**  
Biddle, L., Gunnell, D., Sharp, D., & Donovan, J. L. (2004). Factors influencing help seeking in mentally distressed young adults: A cross-sectional survey. *The British Journal of General Practice: The Journal of the Royal College of General Practitioners*, 54(501), 248–253.
- Birmaher 1999**  
Birmaher, B. (1999). *Screen for Child Anxiety Related Emotional Disorders (SCARED)*. Western Psychiatric Institute and Clinic.
- Briesch 2010**  
Briesch, A. M., Hagermoser Sanetti, L. M., & Briesch, J. M. (2010). Reducing the prevalence of anxiety in children and adolescents: An evaluation of the evidence base for the FRIENDS for life program. *School Mental Health*, 2, 155–165. <https://doi.org/10.1007/s12310-010-9042-5>
- Campbell 2000**  
Campbell, M., Grimshaw, J., & Steen, N. (2000). Sample size calculations for cluster randomised trials. *Journal of Health Services Research & Policy*, 5(1), 12–16.
- Chorpita 2000**  
Chorpita, B. F., Yim, L., MoHitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. *Behaviour Research and Therapy*, 38(8), 835–855.
- Cole 2005**  
Cole, M., Cole, S. R., & Lightfoot, C. (2005). *The development of children*. Worth Publishers.

### Connolly 2018

Connolly, P., Miller, S., Kee, F., Sloan, S., Gildea, A., McIntosh, E., Boyer, N., & Bland, M. (2018). A cluster randomised controlled trial and evaluation and cost-effectiveness analysis of the Roots of Empathy schools-based programme for improving social and emotional well-being outcomes among 8- to 9-year olds in Northern Ireland. *NIHR Journals Library*, 6(4), 1–140. <https://doi.org/10.3310/phr06040>

adult psychopathology in the National Comorbidity Survey Replication (NCS-R) I: Associations with first onset of DSM-IV disorders. *Archives of General Psychiatry*, 62, 113–123.

### Coopersmith 1989

Coopersmith, S. (1989). *Self-esteem inventories*. Consulting Psychologists Press.

### Dalsgaard 2020

Dalsgaard, S., Thorsteinsson, E., Trabjerg, B. B., Schullehner, J., Plana-Ripoll, O., Brikell, I., Wimberley, T., Thygesen, M., Madsen, K. B., Timmerman, A., Schendel, D., McGrath, J. J., Mortensen, P. B., & Pedersen, C. B. (2020). Incidence rates and cumulative incidences of the full spectrum of diagnosed mental disorders in childhood and adolescence. *JAMA Psychiatry*, 77(2), 155–164. <https://doi.org/10.1001/jamapsychiatry.2019.3523>

### Guyatt 2008

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, 336(7650), 924–926. <https://doi.org/10.1136/bmj.39489.470347.A>

### Guyatt 2011

Guyatt, G. H., Oxman, A. D., Sultan, S., Glasziou, P., Akl, E. A., Alonso-Coello, P., & Schünemann, H. J. (2011). GRADE guidelines: 9. Rating up the quality of evidence. *Journal of Clinical Epidemiology*, 64(12), 1311–1316. <https://doi.org/10.1016/j.jclinepi.2011.06.004>

### Guyatt 2013

Guyatt, G., Oxman, A. D., Sultan, S., Brozek, J., Glasziou, P., Alonso-Coello, P., Atkins, D., Kunz, R., Montori, V., Jaeschke, R., Rind, D., Dahm, P., Akl, E. A., Meerpohl, J., Vist, G., Berliner, E., Norris, S., Falck-Ytter, Y., & Schünemann, H. J. (2013). GRADE guidelines: 11: Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *Journal of Clinical Epidemiology*, 66(2), 151–157.

### DerSimonian and Laird 1986

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 177–188.

### Donovan and Spence 2000

Donovan, C. L., & Spence, S. H. (2000). Prevention of childhood anxiety disorders. *Clinical Psychology Review*, 20(4), 509–531. [https://doi.org/10.1016/s0272-7358\(99\)00040-9](https://doi.org/10.1016/s0272-7358(99)00040-9)

### Hampel and Petermann 2005

Hampel, P., & Petermann, F. (2005). Age and gender effects on coping in children and adolescents. *Journal of Youth and Adolescence*, 34(2), 73–83.

### Filges 2023

Filges, T., Smedslund, G., Eriksen, T., & Birkefoss, K. (2023). PROTOCOL: The FRIENDS preventive programme for reducing anxiety symptoms in children and adolescents: A systematic review. *Campbell Systematic Reviews*, 19(4), 1–20. <https://doi.org/10.1002/cl2.1374>

### Health Services Research Unit 2004

Health Services Research Unit. (2004). Database of ICCs: Spreadsheet (empirical estimates of ICCs from changing professional practice studies). Retrieved May 1, 2023, from <https://www.abdn.ac.uk/hsru/what-we-do/tools/index.php>

### Fisak 2011

Fisak, Jr., B. J., Richard, D., & Mann, A. (2011). The prevention of child and adolescent anxiety: A meta-analytic review. *Prevention Science*, 12(3), 255–268. <https://doi.org/10.1007/s11121-011-0210-0>

Heckman, J. J., & Urzúa, S. (2010). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics*, 156, 27–37.

### Fisak 2023

Fisak, B., Griffin, K., Nelson, C., Gallegos-Guajardo, J., & Davila, S. (2023). The effectiveness of the FRIENDS programs for children and adolescents: A meta-analytic review. *Mental Health & Prevention*, 30, 200271. <https://doi.org/10.1016/j.mhp.2023.200271>

### Heckman 2006

Heckman, J. J., Urzúa, S., & Vytlacil, E. (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics*, 88(3), 389–432.

### Green 2010

Green, J. G., McLaughlin, K. A., Berglund, P., Gruber, M. J., Sampson, N. A., Zaslavsky, A. M., & Kessler, R. C. (2010). Childhood adversities and

### Hedges 2007

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, 32, 341–370.

### Guyatt 2008

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, 336(7650), 924–926. <https://doi.org/10.1136/bmj.39489.470347.A>

### Guyatt 2011

Guyatt, G. H., Oxman, A. D., Sultan, S., Glasziou, P., Akl, E. A., Alonso-Coello, P., & Schünemann, H. J. (2011). GRADE guidelines: 9. Rating up the quality of evidence. *Journal of Clinical Epidemiology*, 64(12), 1311–1316. <https://doi.org/10.1016/j.jclinepi.2011.06.004>

### Hampel and Petermann 2005

Hampel, P., & Petermann, F. (2005). Age and gender effects on coping in children and adolescents. *Journal of Youth and Adolescence*, 34(2), 73–83.

### Health Services Research Unit 2004

Health Services Research Unit. (2004). Database of ICCs: Spreadsheet (empirical estimates of ICCs from changing professional practice studies). Retrieved May 1, 2023, from <https://www.abdn.ac.uk/hsru/what-we-do/tools/index.php>

### Heckman 2010

Heckman, J. J., & Urzúa, S. (2010). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics*, 156, 27–37.

### Heckman 2006

Heckman, J. J., Urzúa, S., & Vytlacil, E. (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics*, 88(3), 389–432.

### Hedges 2007

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, 32, 341–370.

**Hedges 2010**

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta regression with dependent effect size estimates. *Research Synthesis Methods*, 1, 39–65.

*Cochrane Database of Systematic Reviews*, 2005(4), CD004690. <https://doi.org/10.1002/14651858.CD004690.pub2>

**Higgins 2009**

Higgins, J., Thompson, S., & Spiegelhalter, D. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 172, 137–159. <https://doi.org/10.1111/j.1467-985X.2008.00552.x>

**Higgins and Green 2011**

Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions*. Version 5.1.0 [updated March 2011]. Wiley-Blackwell The Cochrane Collaboration.

**Higgins and O'Sullivan 2015**

Higgins, E., & O'Sullivan, S. (2015). "What Works": Systematic review of the "FRIENDS for Life" programme as a universal school-based intervention programme for the prevention of child and youth anxiety. *Educational Psychology in Practice*, 31(4), 424–438. <https://doi.org/10.1080/02667363.2015.1086977>

**Higgins 2003**

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327(7414), 557–560.

**Higgins 2011**

Higgins, J. P. T., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., & Sterne, J. A. C. (2011). The Cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 343, 1–9.

**Higgins 2019**

Higgins, J. P. T., Savović, J., Page, M. J., & Sterne, J. A. C. (editors on behalf of the ROB2 Development Group). (2019). Revised Cochrane risk-of-bias tool for randomized trials (RoB 2): Detailed guidance. WP Updated August 22.

**Hills 2011**

Hills, P., Francis, L., & Jennings, P. (2011). The school short-form Coopersmith self-esteem inventory: Revised and improved. *Canadian Journal of School Psychology*, 26, 62–71. <https://doi.org/10.1177/0829573510397127>

**James 2005**

James, A. A. C. J., Soler, A., & Weatherall, R. R. W. (2005). Cognitive behavioural therapy for anxiety disorders in children and adolescent.

**James 2015**

James, A. C., James, G., Cowdrey, F. A., Soler, A., & Choke, A. (2015). Cognitive behavioural therapy for anxiety disorders in children and adolescents. *Cochrane Database of Systematic Reviews*, 2015(2), CD004690. <https://doi.org/10.1002/14651858.CD004690.pub4>

**James 2020**

James, A. C., Reardon, T., Soler, A., James, G., & Creswell, C. (2020). Cognitive behavioural therapy for anxiety disorders in children and adolescents. *Cochrane Database of Systematic Reviews*, 2020(11), CD013162. <https://doi.org/10.1002/14651858.CD013162.pub2>

**Kozina 2012**

Kozina, A. (2012). The LAOM multidimensional anxiety scale for measuring anxiety in children and adolescents: Addressing the psychometric properties of the scale. *Journal of Psychoeducational Assessment*, 30(3), 264–273. <https://doi.org/10.1177/07342911423362>

**Lipsey and Wilson 2001**

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (Vol. 49). Sage Publications Inc.

**Mack 2014**

Mack, S., Jacobi, F., Gerschler, A., Strehle, J., Höfler, M., Busch, M. A., & Gaebel, W. (2014). Self-reported utilization of mental health services in the adult German population—Evidence for unmet needs? Results of the DEGS1Mental health module (DEGS1-MH). *International Journal of Methods in Psychiatric Research*, 23(3), 289–303.

**Maggin and Johnson 2014**

Maggin, D., & Johnson, A. (2014). A meta-analytic evaluation of the FRIENDS program for preventing anxiety in student population. *Education and Treatment of Children*, 37, 277–306. <https://doi.org/10.1353/etc.2014.0018>

**Maggin and Johnson 2019**

Maggin, D., & Johnson, A. (2019). Meta-Analytic evaluation of FRIENDS: A rejoinder of Barrett, Cooper, Stallard, Zeggio, & Gallegos-Guajardo (2017). *Education and Treatment of Children*, 42(1), 129–146.

**Majcher and Pollack 1996**

Majcher, D., & Pollack, M. H. (1996). Childhood anxiety disorders. In T. L. Hechtman (Ed.), *Do they grow out of it? Long-term outcomes of childhood disorders* (pp. 139–169). American Psychiatric Association.

**March 1997**

- March, J. S., Parker, J. D., Sullivan, K., Stallings, P., & Conners, C. K. (1997). The Multidimensional Anxiety Scale for Children (MASC): Factor structure, reliability, and validity. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36(4), 554–565.

**Mathur and VanderWeele 2020**

- Mathur, M. B., & VanderWeele, T. J. (2020). Sensitivity analysis for publication bias in meta-analyses. *Applied Statistics*, 69(Pt. 5), 1091–1119.

**McCaffrey 2001**

- McCaffrey, D. F., Bell, R. M., & Botts, C. H. (2001). Generalizations of biased reduced linearization. In Proceedings of the Annual Meeting of the American Statistical Association. August 5–9.

**McGraw and Wong 1992**

- McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, 111(2), 361–365.

**McLaughlin 2011**

- McLaughlin, K. A., Breslau, J., Green, J. G., Lakoma, M. D., Sampson, N. A., Zaslavsky, A. M., & Kessler, R. C. (2011). Childhood socio-economic status and the onset, persistence, and severity of DSM-IV mental disorders in a US national sample. *Social Science & Medicine*, 73(7), 1088–1096.

**McLean and Anderson 2009**

- McLean, C. P., & Anderson, E. R. (2009). Brave men and timid women? A review of the gender differences in fear and anxiety. *Clinical Psychology Review*, 29(6), 496–505. <https://doi.org/10.1016/j.cpr.2009.05.003>

**Merikangas 2011**

- Merikangas, K. R., He, J., Burstein, M., Swendsen, J., Avenevoli, S., Case, B., & Olfson, M. (2011). Service utilization for lifetime mental disorders in US adolescents: Results of the national comorbidity Survey-Adolescent supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, 50(1), 3245.

**Mrazek and Haggerty 1994**

- Mrazek, P. J., & Haggerty, R. J. (Eds.). (1994). *Reducing the risks for mental disorders: Frontiers from preventative intervention research*. National Academy.

**Myers and Winters 2002**

- Myers, K., & Winters, N. C. (2002). Ten year review of rating scales II: Scales for internalising disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41, 634–659.

**O'Connell 2009**

- O'Connell, M. E., Boat, T., & Warner, K. E. (2009). *Preventing mental, emotional, and behavioral disorders among young people. Progress and possibilities*. The National Academies Press.

**Pahl and Barrett 2007**

- Pahl, K. M., & Barrett, P. M. (2007). The development of social-emotional competence in preschool aged children: An introduction of the Fun FRIENDS program. *Australian Journal of Guidance and Counselling*, 17(1), 81–90.

**Parker 2021**

- Parker, K., Nunns, M., Xiao, Z., Ford, T., & Ukoumunne, O. C. (2021). Characteristics and practices of school-based cluster randomised controlled trials for improving health outcomes in pupils in the United Kingdom: A methodological systematic review. *BMC Medical Research Methodology*, 21(152), 1–17. <https://doi.org/10.1186/s12874-021-01348-0>

**Reynolds and Richmond 1985**

- Reynolds, C. R., & Richmond, B. O. (1985). *Western Psychological Services*.

**Rosenberg 1965**

- Rosenberg, M. (1965). Princeton University Press.

**Satterthwaite 1946**

- Satterthwaite, F. (1946). An approximate distribution of estimates of variance components. *Biometrics*, 2, 110–114.

**Schandelmaier 2020**

- Schandelmaier, S., Briel, M., Varadhan, R., Sauerbrei, W., Devasenapathy, N., Hayward, R. A., Gagnier, J., Borenstein, M., VanderHeijden, G. J. M. G., Dahabreh, I., Schmid, C. H., Ioannidis, J. P. A., Walsh, M., Thabane, L., & Guyatt, G. H. (2020). Development of a new Instrument to assess the Credibility of Effect Modification ANalyses (ICEMAN) in randomized controlled trials and meta-analyses. *CMAJ*, 192(32), E901–E906.

**Sowislo 2013**

- Sowislo, J. F., & Orth, U. (2013). Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. *Psychological Bulletin*, 139, 213–240. <https://doi.org/10.1037/a0028931>

**Spence 2001**

- Spence, S. H., Rapee, R., McDonald, C., & Ingram, M. (2001). The structure of anxiety symptoms among preschoolers. *Behavior Research and Therapy*, 39, 1293–1316.

**Spence 1997**

Spence, S. (1997). Structure of anxiety symptoms amongst children: A confirmatory factor-analytic study. *Journal of Abnormal Psychology*, 106, 280–297.

**Stallard 2012**

Stallard, P., Sayal, K., Phillips, R., Taylor, J., Gannon, M., Anderson, R., Araya, R., Lewis, G., Millings, A., & Montgomery, A. (2012). Classroom based cognitivebehavioural therapy in reducing symptoms of depression in high risk adolescents: Pragmatic clusterrandomised controlled trial. *BMJ (Clinical research ed.)*, 345, 1–13. <https://doi.org/10.1136/bmj.e6058>

**Sterne 2005**

Sterne, J. A., Becker, B. J., & Egger, M. (2005). Chapter 5. The funnel plot. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 75–98). John Wiley & Sons Ltd.

**Sterne 2016a**

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., & Higgins, J. P. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomized studies of interventions. *BMJ*, 355, 1–7.

**Sterne 2016b**

Sterne, J. A. C., Higgins, J. P. T., Elbers, R. G., & Reeves, B. C. (2016). The development group for ROBINS-I. Risk of bias in non-randomized studies of interventions (ROBINS-I): Detailed guidance. WP Updated 12 October, 2016 (pp. 1–53).

**Tanner-Smith and Tipton 2014**

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, 5(1), 13–30.

**Thompson 2004**

Thompson, A. E., Hunt, C., & Issakidis, C. (2004). Why wait? Reasons for delay and prompts to seek help for mental health problems in an Australian clinical sample. *Social Psychiatry and Psychiatric Epidemiology*, 39, 810–817.

**Thompson 2008**

Thompson, A., Issakidis, C., & Hunt, C. (2008). Delay to seek treatment for anxiety and mood disorders in an Australian clinical sample. *Behaviour Change*, 25(2), 71–84. <https://doi.org/10.1375/bech.25.2.71>

**Tipton 2015**

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, 20(3), 375–393.

**Wang 2007**

Wang, P. S., Aguilar-Gaxiola, S., Alonso, J., Angermeyer, M. C., Borges, G., Bromet, E. J., & Gureje, O. (2007). Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. *The Lancet*, 370(9590), 841–850.

**Warren and Sroufe 2004**

Warren, S. L., & Sroufe, L. A. (2004). Developmental issues. In T. H. Ollendick, & J. S. March (Eds.), *Phobic and anxiety disorders in children and adolescents: A clinician's guide to effective psychosocial and pharmacological interventions* (pp. 92–115). Oxford University Press.

**Weems and Stickle 2005**

Weems, C. F., & Stickle, T. R. (2005). Anxiety disorders in childhood: Casting a nomological net. *Clinical Child and Family Psychology Review*, 8(2), 107–134. <https://doi.org/10.1007/s10567-005-4751-2>

**Werner-Seidler 2017**

Werner-Seidler, A., Perry, Y., Calear, A. L., Newby, J. M., & Christensen, H. (2017). School-based depression and anxiety prevention programs for young people: A systematic review and meta-analysis. *Clinical Psychology Review*, 51, 30–47. <https://doi.org/10.1016/j.cpr.2016.10.005>

**Werner-Seidler 2021**

Werner-Seidler, A., Spanos, S., Calear, A. L., Perry, Y., Torok, M., O'Dea, B., Christensen, H., & Newby, J. M. (2021). Schoolbased depression and anxiety prevention programs: An updated systematic review and meta-analysis. *Clinical Psychology Review*, 89, 1–24. <https://doi.org/10.1016/j.cpr.2021.102079>

**Wigelsworth 2018**

Wigelsworth, M., Squires, G., Birchall, E., Kalambouka, A., Lendrum, A., Black, L., Troncoso, P., Santos, J., Ashworth, E., & Britton, P. (2018). *FRIENDS for life: Evaluation report and executive summary*. Manchester Institute of Education Cathie Marsh Institute for Social Research Division of Population Health, Health Services Research & Primary Care.

**World Bank Group 2021**

World Bank Group. (2021). High income. <https://data.worldbank.org/country/XD>

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Filges, T., Smedslund, G., Eriksen, T., Birkefoss, K., & Kildemoes, M. W. (2024). The FRIENDS preventive programme for reducing anxiety symptoms in children and adolescents: A systematic review and meta-analysis. *Campbell Systematic Reviews*, 20, e1443. <https://doi.org/10.1002/cl2.1443>