Outcomes of routine mental health care for young people with depression/ anxiety

Systematic review and meta-analysis: Outcomes of routine specialist mental health care for young people with depression and/or anxiety

Holly Alice Bear, Julian Edbrooke-Childs, Sam Norton, Karolin Rose Krause, Miranda Wolpert

Clinical, Educational and Health Psychology, University College London, Gower Street, London, WC1E 6BT, UK (HA Bear MSc, J Edbrooke-Childs PhD, KR Krause MSc, Professor M Wolpert PsychD)

Anna Freud National Centre for Children and Families; The Kantor Centre of Excellence, 4-8 Rodney Street, London, N1 9JH, UK (HA Bear MSc, J Edbrooke-Childs PhD, KR Krause MSc, Professor M Wolpert PsychD)

Child Outcomes Research Consortium, 1st Floor Jordan House, Brunswick Place London, N1 6EB, UK (J Edbrooke-Childs PhD, Professor M Wolpert PsychD)

Health Psychology Section, Institute of Psychiatry, Psychology & Neuroscience, King's College London, Guy's Hospital Campus, London Bridge, London, SE1 9RT, UK (S Norton PhD)

Correspondence to: Holly Bear, Evidence Based Practice Unit, The Kantor Centre for Excellence, 4-8 Rodney Street, London, N1 9JH, UK, holly.bear.16@ucl.ac.uk

Summary

Objective: Depression and anxiety are the most prevalent mental health problems in youth, yet almost nothing is known about what outcomes are to be expected at the individual level following routine treatment. This paper sets out to address this gap by undertaking a systematic review of outcomes following treatment as usual (TAU) with a particular focus on individual-level outcomes.

Method: MEDLINE, Embase and PsycInfo were searched for articles published between 1980 and January 2019 that assessed TAU outcomes for youth depression and anxiety accessing specialist mental healthcare. Meta-analysis considered change at both group-level pre-post effect size (ES) and individual-level recovery, reliable change and reliable recovery. Temporal analysis considered stability of primary and secondary outcomes over time. Sub-group analysis considered the moderating effect of informant; presenting problem; study design; study year; mean age of youth; use of medication; intervention dosage and type of treatment offered on outcomes. A protocol was pre-registered on PROSPERO (CRD42017063914).

Results: Initial screening of 6,350 publications resulted in 38 which met the inclusion criteria, and which were subsequently included in meta-analyses. This resulted in a final full pooled sample of 11,739 young people (61% of which were female, mean age 13.8 years). The pre-post ES (Hedges' *g*) at first/final outcome (13/ 26 weeks) was -0.74/- 0.87. The individual-level change on measures of self-report was 38% reliable improvement, 44% no reliable change and 6% reliable deterioration. Outcomes varied according to moderators, informant, problem type and dosage.

Conclusions: Poor data quantity and quality are limitations, but this is the first study that indicates likely rates of reliable improvement for those accessing TAU. We propose the need for improved reporting of both individual-level metrics and details of TAU to enable greater

understanding of likely current outcomes from routine care for youth with depression and anxiety in order to allow the potential for further improvement of impact.

Keywords: *Children; young people; anxiety; depression; usual care; treatment outcome*
Abbreviations: CYP, children and young people; RCI, reliable change index

Introduction

Internationally, anxiety and depression are among the most prevalent and widely diagnosed psychiatric conditions in childhood and adolescence. [1, 2] The experience of depression and anxiety can fluctuate over the life course but if not successfully resolved can lead to major negative impacts well into adulthood.[3-6] These conditions frequently present together with comorbid conditions showing a worse prognostic outlook than either presentation alone. [7, 8]

There is evidence that rates of depression and anxiety may be increasing in children and young people. [1, 9, 10] This has been accompanied by calls for greater services to address their needs. Substantial and sustained effort has been made over many decades to develop and evaluate evidence-based treatments for anxiety and depression in children and young people. This has culminated in a body of evidence supporting several efficacious treatments which indicate improved outcomes when compared to control groups.[7, 11-14] However, there is a paucity of literature evaluating outcomes achieved by children and young people accessing routine care or treatment as usual (TAU) in relation to specialist mental health care.

Evidence suggests that outcomes achieved in routine care may be less positive than those achieved under experimental conditions,[7, 14-16] though other findings suggest the difference may be less pronounced.[17] The majority of studies report the average mean difference between two treatments at a group-level, with minimal consideration given to outcomes at an individual-level, using metrics such as recovery or reliable change. [18, 19] Using pre-post effect size (ES), with the inappropriate application of large, moderate and small ES parameters, may contribute to overinflated estimates of impact as there is no consideration of clinically meaningful metrics of change. The use of ES as the key pre-post measure has been critiqued as potentially misleading and there is an increasing call for a focus on individual-level outcomes. [18-23]

There have been some attempts to measure individual-level change in analyses of routine data,[22, 24, 25] but no systematic review been undertaken to synthesize this information. The fact that in routine practice a wide range of different outcome measures are used and informed by different perspectives adds to the complexity of comparing across treatments. Symptom change has traditionally been the primary focus of outcome studies, however, there is evidence that some domains that may be of relevance to service users, such as functioning, remain under-investigated. [26]

There is a dearth of research examining the mechanisms or predictors of outcomes, and efforts to determine the key moderating factors in treatment outcomes following routine practice have provided an inconsistent picture.[27] There may be several person and treatment-related variables which act as moderators, including but not limited to, problem type, comorbidity, treatment length, outcome informant. [28-33] Outcome moderation within this context warrants further investigation and is crucial for delineating the factors which may enhance or reduce clinical effectiveness in routine practice. This information may prove decisive in enabling clinicians to be prescriptive in their choice of treatment based on the likely outcome on an individual basis.

In the absence of such information, and in the light of the understandable desire to increase access and reduce stigma, professional groups tend to stress the likely positive impact of treatment. For example, according to the website of the Anxiety and Depression Association of America: 'Like other medical conditions, anxiety disorders tend to be chronic unless properly treated. Most kids find that they need professional guidance to successfully manage and overcome their anxiety' It has been argued that this may leave referrers, children and families alike with unrealistic and potentially unhelpful expectations as to what outcomes to expect from treatment.[34]

The overarching aim of this meta-analysis is to determine outcomes achieved following TAU by specialist mental health services among children and young people with depression and/ or anxiety. The specific objectives are to: 1) provide a synthesis of the available literature to investigate group-level and individual-level outcomes following routine care by specialist mental health services for children and young people with anxiety and/or depression; 2) consider the relationship between outcomes and potential moderators identified in the literature as potentially relevant: informant (i.e. self, parent, clinician),: target problem (i.e. anxiety, depression or both); study design (experimental, observational); mean age (<11 years, 11-16 years, >16 years); medication usage (0-100%): treatment dosage (mean number of sessions) and type of treatment offered (single modality, mixed); 3) to evaluate any secondary outcomes reported in the reviewed studies such as functioning and quality of life; and 4) to consider outcome trajectories over time and across outcome domains.

Method

Search strategy and selection criteria

A comprehensive and systematic search of electronic databases MEDLINE, Embase and PsycInfo was undertaken from their inception to January 2019. The search strategy consisted of four major components, containing key terms related to: 1) children and young people; 2) anxiety and depression; 3) routine care; and 4) recovery (Table S1, available online). We also searched reviews of child and adolescent psychotherapy research, followed reference trails in the identified publications, and obtained additional studies identified by manual searching.

Studies were systematically considered for inclusion if they satisfied the following criteria: i) participants were children and young people (mean age ≤ 21 years) accessing specialist mental health services for the treatment of depression and/ or anxiety symptoms or

equivalent i.e. emotional, internalising, affect and mood disorders (specialist mental health services are taken here to mean services which exist to provide mental health treatment, beyond primary care provision); ii) treatment was indicative of treatment as usual (TAU) for the treatment of mental health difficulties (TAU is defined as the care provided by specialist mental health services for the treatment of anxiety and/ or depression, which remain unaltered by the study design. TAU was expected to be diverse and involve individual and/or family-orientated and/ or group work, delivered by a range of practitioners with various theoretical orientations for varying levels of treatment intensity); iii) the primary outcome of interest was change in: anxiety and depressive symptoms, diagnosis, problem severity, problem improvement, recovery, remission or more general change in mental health across at least two time points (e.g. assessment and discharge); and iv) original research. Reasons for exclusion are provided in Supplement 1, available online.

A second review author independently assessed 10% of the full texts, and discrepancies were resolved through discussion.

Data Extraction and Management

A standardised, pre-piloted data extraction table was designed and used to collect the data from the remaining manuscripts. Data were extracted on study setting and provider, design, sample size, participant characteristics, presenting problem(s), recruitment source, intervention description, outcome measure(s), measurement timing, informant, moderator/predictor variables and results. Where possible all potential moderators were extracted from studies and coded in accordance with guidance from a recent review.[33] These included therapist training, dosage, therapist allegiance, sample representativeness and medication use.

Corresponding authors were contacted to resolve uncertainty, or in instances where additional information was required due to missing data.

Quality Assessment and Risk of Bias

Studies were scored for methodological quality using the validated 27-item Downs and Black checklist,[35] provided in Supplement 2, available online. The tool produces a composite score ranging from 0 to 28, where higher scores are indicative of better methodological quality. According to pre-defined criteria, scores of 26 to 28 demonstrate excellent quality, 20 to 25 good quality, 15 to 19 fair and scores of 14 or less are poor. The Downs and Black tool has been ranked in the top six quality assessment tools for use in systematic reviews [36] and exhibits sound psychometric properties including internal consistency, test–retest reliability, inter-rater reliability and criterion validity. [35]

For the purposes of the current review, the tool was adapted by removing 5 questions: number 14 which related to participant intervention blinding (generally not possible with psychological interventions) and numbers 21-24 as they related to group comparisons and randomisation, as this review was only concerned with within group pre-test-post-test data, reducing the total possible score to 23; the cut-offs were adjusted accordingly. Where the power was not stated, the question relating to power was scored 0. Intraclass correlation coefficients (ICCs) were used to assess the inter-rater agreement with a second researcher for 20% of included studies.

Quantitative Data Analysis

Data was analysed using Stata 14.1.

Group-level outcome change

In terms of group-level change, for studies reporting mean within-group change, the pre-treatment, post-treatment ES are reported using Hedges' g [37] and adjusted using Hedges' small sample correction.[38] ESs were calculated using the data provided in the publications with the formula:

$$g = \left(\frac{M_1 - M_2}{SD^*_{\text{pooled}}}\right) \times \left(\frac{N-3}{N-2.25}\right) \times \sqrt{\frac{N-2}{N}}$$

Where M_1 and M_2 are the within-group means at baseline and follow-up, respectively, SD^*_{pooled} is the pooled weighted standard deviation (SD) across both time points, and N is the sample size. The pooled weighted SD, rather than the SD of the within-person change score, is used as it allows estimated effects on a comparable metric to the standardised effect sizes reported in a meta-analysis of controlled trials with independent groups.[39] Hedges' g was chosen as an alternative to Cohen's *d*, as the latter is known to be upwardly biased for small samples. Standardised ESs were transformed to ensure directional uniformity across samples. In cases where studies yielded multiple ESs, thus violating the assumption of independence, each was separated by outcome then the ES of the primary outcome was chosen.

Meta-analyses for group-level outcomes were conducted on the full pooled sample using a random effects model using the first assessment post-baseline recorded by each study (>4 weeks post-baseline) and separately at the final assessment point recorded. Heterogeneity between the studies was assessed using I^2, where an I^2 value >50% was considered suggestive of heterogeneity.

Longitudinal meta-analyses were conducted using a mixed effects model with observations within samples weighted by the inverse of their variance (i.e. standard error) and a sample-level random effect explored the change in treatment effect over time.[40] These analyses were conducted for the primary outcome (clinical symptoms) and secondary outcomes (secondary clinical symptoms, functioning and quality of life (QoL).

Individual-level outcome change

Individual-level change rates were considered for a smaller pooled sub-sample, taken from the figures reported by the authors in each study.

Recovery was defined as scores which crossed from above to below a pre-defined threshold on a particular measure (i.e. 'recovered') or remained above the cut-off at follow-up (i.e. 'non-recovered') following treatment.

Reliable change determines whether change in scores was greater than could likely be solely attributed to measurement error and is based on the reliable change index (RCI). The RCI utilised was the pre-test score minus the post-test divided by the standard error of the difference. If the RCI > 1.96, change is regarded as statistically reliable at the 95% confidence level. Therefore, group membership can be defined as reliably improved (RCI≥1.96), reliably deteriorated (RCI≤−1.96), or no reliable change (RCI between −1.95 and 1.95).

Reliable recovery was defined as scores that crossed from above a cut-off at baseline to below a cut-off following treatment (i.e. 'recovered') and which demonstrated an improvement in score greater than the reliable change threshold on the same measure (i.e. 'reliably improved').

Meta-analyses were conducted on the smaller pooled sub-samples for self-reported recovery, reliable change and reliable recovery. These samples are highlighted in Table 3.

Moderator analysis

Potential moderating factors were examined using meta-analytic sub-group analyses on the full pooled sample.

Results

Search Flow

The database and manual searches yielded an initial pool of 6,350 records. After the removal of duplicates, 3880 publications were screened by title and abstract. Of those, 3,485 did not meet the inclusion criteria and were excluded. A total of 395 full texts were assessed for eligibility against the eligibility checklist, of which 357 were excluded prior to data

extraction (Table S2, available online). A total of 12 publications were excluded as they reported on the same study sample as a publication(s) already included. For example, six publications [41-46] were based on only one study sample, therefore, data from the original paper [41] was included in the final meta-analysis. Seven studies were excluded as additional information was not provided by corresponding authors on request.[28, 29, 47-51] A total of 38 studies remained, relating to 44 independent study samples (Figure 1).

[INSERT FIGURE 1]

Figure 1. Phases of Systematic Search Strategy

Quality of Included Studies

The scoring for each publication is available in Table S3, available online. Approximately one third of publications did not adequately describe the usual care intervention or arm within their study and only 9% recorded or reported on any harmful effects or 'adverse outcomes' of the interventions. In particular, when studies found individual-level deterioration in symptoms, this was not considered as a harmful or adverse effect.

Other potential sources of bias related to non-representative sampling, poor reporting of intervention fidelity and adherence, a failure to account for confounding factors in the analysis and underpowered analyses. Intraclass correlation coefficient (ICC) analyses determined good inter-rater agreement (ICC = 0.78, CI 0.69-0.84, p <.001).

This review aimed to code all potential intervention-level moderators in accordance with newly developed guidelines and coding scheme.[33] However, following data extraction the coding frame was incomplete given that many of the primary studies did not adequately report all potential moderators and few studies reported data across the same variables (see Table 2).

Population Characteristics

Table 1 provides information about the study and population characteristics of all 38 included studies, which were published between 1995 and 2018. Thirteen of the studies were conducted in the United States of America (USA), 5 in the United Kingdom (UK), 4 in Germany, 3 in Australia, Canada and Sweden and 1 each from Denmark, Norway, Finland, New Zealand, Italy, Ireland and India. The studies included 11,739 CYP, with sample sizes ranging from 6 to 4,659. Participants were aged from 7 to 20 years old with an average of 13.8. The majority of participants were female (mean = 61.65%, range 32-84%). Of the 38 studies, 33 employed a diagnostic or cut-off inclusion criterion.

[INSERT TABLE 1]

Intervention characteristics

The majority of the TAU described was eclectic or interdisciplinary in nature i.e. not aligned to a specific therapeutic modality (*n* = 24), 12 studies provided usual care interventions which were exclusively aligned to a single therapeutic modality (e.g. psychoanalytic, CBT, counselling, pharmacotherapy). Nine studies did not report the method of intervention delivery (i.e. individual or group), nine studies did not provide details of the intervention provider and twenty nine studies did not discuss intervention fidelity. Full details of the interventions are presented in Table 2.

[INSERT TABLE 2]

Results of the Included Studies

Group-level primary outcome at first assessment post-baseline

Group-level effects were considered with respect to the primary outcome, which was defined as a validated measure of clinical symptoms. Five publications contained more than one independent sub-group within their larger sample and were therefore entered as independent ESs within the meta-analysis, making a total of 44 samples and 11,739 participants.

Seven out of 38 publications reported more than one measure of symptoms. In these cases a decision was made about the most appropriate measure to use for the primary outcome analysis (those not chosen were included in the secondary outcome analysis of symptoms). Self-report measures were chosen above clinician-report measures. Global symptom measures (Children's Depression Inventory, CDI) were chosen above problem-specific measures (e.g. Social Phobia Inventory for Children, SPAIC), measures reported by on their subscale rather than total score, and measures of both symptoms and functioning (Strengths and Difficulties Questionnaire, SDQ). Depression measures (Beck Depression Inventory II, BDI-II) were chosen above anxiety measures (Multidimensional Anxiety Scale for Children 2nd Edition, MASC 2) if the aim of the study was to measure depression outcomes in depressed individuals.

The median period from baseline to first assessment post-baseline was 13 weeks (interquartile range = 12-27 weeks). As displayed in Figure 2, at first assessment post-baseline (>4 weeks post-baseline), the pooled standardised ES (Hedges' *g*) of the primary outcome was -0.74 (95% CI = -0.88, -0.59).

Significant statistical heterogeneity was indicated by an I^2 of 88.1% (p <. 001), were I^2 reflects the between study heterogeneity likely due to methodological disparity, in particular, relating to the wide range of assessment time points used. This was addressed in the longitudinal meta-regression.

[INSERT FIGURE 2]

Figure 2. Meta-Analysis of Primary Outcome at First Assessment Post-Baseline Recorded

Note: First assessment post-baseline is >4 weeks post-randomisation or post-baseline, k = 44 samples and n = 11,739 participants. Negative effect size (Hedges' g) values indicate a reduction in symptoms.

Group-level primary outcome at final assessment

The median period from baseline to final assessment was 26 weeks (interquartile range = 15–52 weeks), with a total of 44 samples and 11,745 observations. At the final assessment, the pooled standardised ES (Hedges' *g*) for the primary outcome was -0.87 (95% CI -1.01, -0.74), demonstrating a magnified large ES at the final assessment (Figure S1, available online). Significant statistical heterogeneity was indicated by an I^2 of 86.4% (p<.001). Again, this degree of statistical heterogeneity may be partially attributable to methodological disparity.

Individual-level primary outcome

A total of 25 included studies reported treatment response at an individual level, with 19 reporting recovery, 12 utilising the reliable change index (RCI) and 8 reporting reliable recovery, which encompasses both recovery and reliable change in a single metric (Table 3).

For self-report measures only, the weighted pooled percentage of individual-level change for the primary outcome of symptoms across eight studies and ten samples (*n* = 4,900) demonstrated that 33% recovered, as measured by movement across a predefined clinical threshold (average assessment time = 29.4 weeks, CI 22, 43%). For self-reported reliable change (*n* = 10 studies, *k* = 13, *n* = 5,331 participants, average assessment time = 30.3 weeks), 38% exhibited reliable improvement (CI, 29, 46%), 6% reliably deteriorated (CI 4, 9%) and 44% showed no reliable change (CI 36, 52%). Finally, 40% met the criteria for self-reported reliable recovery, which is defined as both recovery and reliable change combined (*n* = 6 studies, *k* = 9, *n* = 5,068, average assessment time = 32.5 weeks, CI 31, 48%).

It should be noted that the sub-samples used for each of the calculations above differed depending on the data provided by each study and every study did not report all individual-level change metrics.

It should also be noted that sub-samples were not always independent of each other. For most studies, the reliable improvement, reliable deterioration and no reliable change were mutually exclusive groups, and equalled 100%. Group membership of reliably changed with both recovered and reliably recovered were not mutually exclusive groupings. Thus, young people could be counted as both reliably changed and reliably recovered within the same study. The total weighted pooled percentage of individual-level change metrics do not total 100%.

[INSERT TABLE 3]

Moderator analyses for primary outcome

In order to determine if 'outcome informant' moderated the treatment effect on the primary outcome variable (clinical symptoms) the standardised ES was compared across informant groups. All observations across studies were coded as 'self', 'parent' (which included 'mother', father', 'guardian') and 'clinician/assessor' (which included blinded interviewers, observers, and research assistants). As there were only three teacher observations, these were excluded from analysis. It should be noted that not every sample had observations from each of the three informant types.

As displayed in Figure S2 (available online), self-reported symptoms demonstrated a standardised pre-post ES of -0.70 (CI -0.85, -0.55). Symptoms as rated by clinicians/assessors demonstrated the largest standardised change (Hedges' g = -1.30, 95% CI -1.64, -0.96) and were significantly larger than self-reports (mean difference = 0.6, $p < .001$). Parent reports exhibited the smallest standardised change in symptoms (Hedges' g = -0.59, CI -0.77, -0.41), with a difference of 0.11 with self-reports ($p = 0.37$).

In order to determine if 'presenting problem' moderated the effect on the primary outcome variable (clinical symptoms) the standardised ES was compared across problem groups. All observations across studies were coded as 'depression' (which included mood

disorders and depressive disorders), 'anxiety' and 'mixed' (which included internalising disorders, mixed anxiety and/or depression samples, emotional disorders).

As displayed in Figure S3 (available online), the standardised ES was largest for depression (Hedges' g = -0.89, 95% CI -1.04, -0.73) followed by anxiety (Hedges' g = -0.66, CI -0.86, -0.46) and then 'mixed' anxiety and depression (Hedges' g = -0.52, CI -0.82, -0.21). There was a significant difference between the standardised ES of those with depression and both anxiety and mixed anxiety and depression (difference = 0.23, p = 0.08 and 0.37, p = 0.35, respectively).

As displayed in Figure S4 (available online), There was a small significant moderating effect of treatment dosage (mean number of sessions) on the primary outcome variable (g = -0.03, $p < 0.001$). Across studies, the mean number of treatment sessions per study ranged from four to 196 (mean = 26.8). A total of 11 studies did not report dosage and were therefore excluded form analysis.

There was a non-significant moderating effect of study year on the primary outcome variable, which showed a slight trend for improvement in outcomes (g = -0.02, p =0.167, 95% CI -.04, 0.01). Similarly, when categorising the data as 'before 2010' or 'after 2010', the difference in pre-post ES was non-significant. (g = -0.21, p = 0.19, 95% CI -0.54, 0.11).

Additional analyses were conducted on variables, experimental design (experimental, observational), (<11 years, 11-16 years, >16 years), medication usage (0-100%) and therapeutic modality (single modality, mixed). None of these variables moderated the effect of treatment on symptom change, as there was no significant difference in ESs between sub-groups (Figures S5-8, available online).

Outcome trajectories

Longitudinal analysis of the standardised change (Hedges' g) in the primary outcome variable from baseline was conducted using mixed effects meta-regression (Figure 3). This

analysis included 38 studies, 70 observations from the 44 samples (range 1-5 observations per sample) including 11,739 participants. Time from baseline assessment (weeks) was entered into the model as a continuous variable. To determine whether the ES changed over time, either linearly or non-linearly, was assessed by running separate models including time as either a linear, quadratic and logarithmic function. The Bayesian Information Criterion (BIC), where a lower number indicates better fit after penalising for model complexity, indicates that the logarithmic trend provides the best fit to the data. This specifies that the symptoms improve rapidly initially but that improvement then slows until reaching a plateau; compared to the linear (monotonically increasing or decreasing effect) and quadratic (accelerating/decelerating effect) trends.

Figure 3 demonstrates that, for the logarithmic trend, there is an initial large reduction in symptoms within the first 3 months followed by a gradual deceleration in the rate of change thereafter. By 6 months, there is very little additional change. Of interest is the congruity between the 3 models; between 3 (-0.6) and 18 months (-1) there is relatively little difference in the estimated ESs.

[INSERT FIGURE 3]

Figure 3. Standardised Change in Primary Outcome from Baseline
Note: *n* = 70 Observations (range = 1-5 per sample). Primary outcome = symptoms. Circle size corresponds to the inverse of the standard error for each observation.

It should be noted that three samples, from three separate studies, exhibit mean increases in symptoms from baseline. For two studies the increase was negligible, however, for one study the effect was large and potentially clinically meaningful though this sample consisted of only four participants [52].

A funnel plot (Figure S9, available online) was utilised to assess potential small sample effects or publication bias. A lack of symmetry (particularly for smaller studies with

large standard errors) indicates potential publication bias or that there may be smaller studies missing. This can be seen in the plot with studies with a SE of approximately above 0.4.

Group-level secondary outcome analysis

Longitudinal meta-regressions were conducted to examine the standardised change over time for measures defined as secondary outcomes. Multiple outcome domains were reported across most studies, including family relationships, life interference, emotional adjustment, health outcomes and global functioning. For the purposes of the current analysis, outcomes were coded where appropriate under three broad outcome domains of 'symptoms' 'functioning' (inclusive of adjustment and impairment) and 'quality of life (QoL)'. Separate models were estimated for each outcome domain.

Figure S10 (available online) illustrates standardised change in functioning over time for 15 independent samples, 55 observations and 2,203 participants, which was not included in the primary longitudinal analysis. In this instance, lower scores are synonymous with increased functioning. Again, the logarithmic model provided the best fit to the data and demonstrates an initial large increase in functioning, followed by a plateau at around 15 months. The magnitude of the standardised ES for functioning appears to be larger than for symptoms and deceleration in the rate of change happens at a slower rate.

Figure S11 (available online) displays the standardised change in QoL over time and included 5 independent sub-samples and 27 observations, which was not included in the primary longitudinal analysis. In this instance, lower scores are synonymous with increased QoL. QoL exhibited substantially less change than symptomatic and functional outcomes. For the logarithmic trend, the ES is approximately 0.4 and exhibits relatively minimal change across time. Of interest are the 2 studies which show large reductions in quality of life ($g >$ 2).

Discussion

To date, no previous systematic review has aimed to address crucial gaps in the extant literature, which include: 1) a lack of an adequate and accurate synthesis of TAU outcome data considered both as pre-post effect sizes but also, crucially, in terms of individual-level rates of change following interventions in routine care; 2) identification of variables which may moderate the observed outcomes; 3) a consideration of change trajectories; and 4) consideration of change across primary and secondary outcomes.

The current systematic review identified 38 studies (44 independent samples) of TAU within specialist services for youth with anxiety and/ or depression. Only 25 (66%) of these included details of individual-level change (19 included information about recovery, 12 information about reliable change and 8 information about a combination of the recovery and reliable change). Given that depression and anxiety are the most prevalent mental health problems in youth,[1, 2] the paucity of information about outcomes from TAU in specialist mental health services, particularly in relation to individual-level change, is striking.

The quality of data in the identified studies was poor with much key information missing. In particular, there was a lack of detail provided by studies regarding the TAU intervention and what it involved, including who the provider was and their level of training, the delivery method, and intervention fidelity. There was also a large geographical skew towards North American studies, but this is largely in line with general research in this area.[26]

Within the studies identified, there was considerable heterogeneity in outcome metrics, time points of outcomes, and the informant reporting on the outcomes. There was also a vast array of measurement tools used, with multiple tools used per study. Such methodological diversity made comparisons across studies challenging and highlights the need for greater agreement on a core outcome set for this problem.[26]

There are clearly significant limitations in this study. As noted above, the quantity and quality of the data made analysis challenging. Not only were the interventions poorly

described but those that were described varied considerably across providers and settings and as noted above multiple and different outcome metrics and multiple different time points made comparisons difficult. However, in the face of these limitations it is striking that a consistent finding was that a high proportion of CYP within each sample did not meet the criteria for measurable improvement, such as recovery and reliable improvement, across measurement metrics and problem domains, irrespective of informant or time point of study.

In terms of individual-level change, for the primary outcome of self-reported symptoms, 33% clinically improved (29.4 weeks), 38% reliably improved (30.3 weeks), 6% reliably deteriorated (30.3 weeks), 44% showed no reliable change (30.3 weeks) and 40% reliably recovered (32.5 weeks). As previously noted, these metrics are weighted averages calculated based on different study samples. They are not necessarily mutually exclusive categories and are not intended to total 100%. Across different metrics of change, between 33% and 40% of young people showed improvement. These results are in line with earlier individual study findings and highlight that the majority of those seen by specialist mental health services for TAU do not show measurable improvement on any indicator of individual level-change.[24, 25]

It is important to note that our data cannot determine whether improvement is due to the treatment itself or whether young people might improve to a similar degree irrespective of the intervention received. The recurrent and episodic nature of anxiety and depression means they may spontaneously worsen, remain stable, or improve over a similar time frame with no treatment.[27, 53] Research suggests that in adult populations, 53% of prevalent cases of untreated major depression will remit spontaneously in a given year.[54] However, contact with mental health services reduces the future likelihood of depression compared to those without contact in adolescents.[55]

We would like to highlight that these results should not be taken to mean that the outcomes are suboptimal. These sorts of measurable improvement rates are better than individual-level outcomes for many other chronic conditions (e.g. paediatric diabetes, 24%)[56] and may reflect the current reality of what treatment is able to achieve currently. Moreover they may also reflect the lack of sensitivity of the measures used to measure change with some indication that more idiographic measures, such as movement towards achievement of goals, may be more sensitive to change.[57]

However it should also be noted that these results are somewhat lower than those achieved for early intervention for depression/anxiety in working-age adults where rates of reliable recovery of around 50% in the UK are currently reported.[58] This may reflect the greater challenge of treating young people or the greater complexity of problems of young people accessing youth services.

These results suggest the need to consider a) how best to address the needs of those who access specialist services and are not measurably improved at the end of treatment and b) how to improve outcomes from TAU for these groups of children and young people.

The pre-post ES (Hedges' *g*) at first outcome (average 13 weeks) and final outcome (average 26 weeks) was -0.74 and - 0.87, respectively. The longitudinal analysis revealed an initial large reduction in symptoms within the first 3 months followed by a gradual deceleration in the rate of change thereafter and by 6 months there was little additional change. This is in line with earlier studies which demonstrate a curvilinear outcome response, with initial rapid improvement rates followed by a decelerating curve over the course of treatment.[30, 59-61]

The moderator analysis suggested that, in line with previous research, informant perspective moderates outcomes, with clinician reports generally documenting greater improvements than self or parent reports.[62] This highlights the need to ensure the views of

patients themselves are considered wherever possible if patient shared decisions and empowerment are to be made a reality.[63]

Problem type also moderated outcomes with better outcomes for those studies focussed on depression than those for anxiety. This is not in line with earlier studies that generally report better outcomes following treatment for those with anxiety (e.g. Weisz et al. 2017 [7]). It is not clear why this might be. It might reflect that anxiety is less impacted by the generic treatments that formed most of TAU than depression is, and is better treated by evidence-based, manualised approaches provided by RCTs. Our data do not provide a decisive explanation for this result, however, it is possible there are other methodological explanations. For instance, there may be something unusual about the depression studies with a particularly large pre-post ES.

It is interesting to note that no other moderators were identified as being associated with differential outcomes. Given the uncertainty in the treatment effect estimates, the significance tests of the subgroup analyses could potentially be false negatives and thus drawing conclusions of no difference may not be robust. Of particular interest, publication year did not have a significant moderating effect on improvement in outcomes over time, however, there is insufficient power to say whether this result is reliable. This finding is consistent with a number of other reports that treatment outcomes are not moderated by study year.[7, 64, 65]

In terms of secondary outcomes measured, the most common were measures of functioning. This is in line with a recent large scale review of primary and secondary outcome measures.[26] Change in functioning showed a large ES within the initial 12 months, followed by a plateau at around 15 months. Quality of life, on the other hand, showed little change both initially and across time suggesting it may be more stable and independent of mental health outcomes.[66]

The use of effect size as a key pre-post measure has been critiqued as potentially misleading.[20, 22, 23] While clear benchmarks exist with regard to what constitutes a small, medium or large between-group ES, there is yet no agreed equivalent cut-offs for pre-post ES and are often inappropriately used as such. Without clear rules of thumb for the use of pre-post ES, interpreting these metrics is challenging. This highlights the need to consider alternative individual-level metrics so that we have a consistent way of discussing within-group treatment outcomes.

Despite limitations in the data, this study has important implications for both research and clinical practice. In terms of research it is imperative that future studies include more detailed descriptions of TAU interventions. This may be facilitated by the utilisation of the Template for Intervention Description and Replication or TIDieR framework.[67] Research protocols which aim to evaluate treatment effectiveness should endeavour to measure and report adverse outcomes and events, and in the future, it would be beneficial to standardise the way in which outcomes are measured and reported using both group and individual-level metrics.

In terms of clinical practice it may be important that clinicians discuss with their clients which outcomes are important to them and also consider outcomes from a variety of perspectives. It may also be beneficial for clinicians to consider how they support the majority of young people with anxiety and depression who leave their service without measurable change and how they manage expectations around this from the start of treatment.

Whilst the limited quantity and quality of data available means these results should be treated with caution, the majority of cases in receipt of TAU did not show positive outcomes on any individual level outcome metric. Improvement rates are somewhat lower than those achieved for working-age adults, suggesting the need for greater clinical and research focus on the needs of those young people who do not show improvements. The fact that outcomes

stabilise at around 6 months is consistent with earlier literature. The finding that changes for secondary outcomes, whilst smaller overall, continue to embed later than for primary outcomes indicate potential for further focus on non-symptom measures such as quality of life. It is less certain is which intervention components produced better outcomes, under what specific circumstances improved outcomes were achieved and for whom, and if similar outcomes would be archived with no treatment. We call for improved reporting of individual level metrics, greater attention to who reports and which outcomes are prioritised, greater focus on adverse events and greater detail as to what constitutes TAU to enable the field to move forward in understanding, and in time improving, outcomes from routine care for depression and anxiety, particularly when occurring together, in children and young people.

References

1. Bor W, Dean AJ, Najman J, Hayatbakhsh R. Are child and adolescent mental health problems increasing in the 21st century? A systematic review. Australian & New Zealand Journal of Psychiatry. 2014;48(7):606-16.
2. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual Research Review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. Journal of Child Psychology and Psychiatry. 2015;56(3):345-65.
3. Hardeveld F, Spijker J, De Graaf R, Nolen W, Beekman A. Recurrence of major depressive disorder and its predictors in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Psychological medicine. 2013;43(1):39-48.
4. Clayborne ZM, Varin M, Colman I. Systematic Review and Meta-Analysis: Adolescent Depression and Long-Term Psychosocial Outcomes. Journal of the American Academy of Child and Adolescent Psychiatry. 2019;58(1):72-9.
5. Bruce SE, Yonkers KA, Otto MW, Eisen JL, Weisberg RB, Pagano M, et al. Influence of psychiatric comorbidity on recovery and recurrence in generalized anxiety disorder, social phobia, and panic disorder: a 12-year prospective study. American Journal of psychiatry. 2005;162(6):1179-87.
6. Fombonne E, Wostear G, Cooper V, Harrington R, Rutter M. The Maudsley long-term follow-up of child and adolescent depression: I. Psychiatric outcomes in adulthood. The British Journal of Psychiatry. 2001;179(3):210-7.
7. Weisz JR, Kuppens S, Ng MY, Eckshtain D, Ugueto AM, Vaughn-Coaxum R, et al. What five decades of research tells us about the effects of youth psychological therapy: a multilevel meta-analysis and implications for science and practice. American Psychologist. 2017;72(2):79.
8. Ollendick TH, Jarrett MA, Grills-Taquechel AE, Hovey LD, Wolff JC. Comorbidity as a predictor and moderator of treatment outcome in youth with anxiety, affective, attention deficit/hyperactivity disorder, and oppositional/conduct disorders. Clinical Psychology Review. 2008;28(8):1447-71.
9. Fink E, Patalay P, Sharpe H, Holley S, Deighton J, Wolpert M. Mental health difficulties in early adolescence: a comparison of two cross-sectional studies in England from 2009 to 2014. Journal of Adolescent Health. 2015;56(5):502-7.
10. Avenevoli S, Baio J, Bitsko RH, Blumberg SJ, Brody DJ, Crosby A, et al. Mental health surveillance among children--United States, 2005-2011. 2013.
11. David-Ferdon C, Kaslow NJ. Evidence-based psychosocial treatments for child and adolescent depression. Journal of Clinical Child & Adolescent Psychology. 2008;37(1):62-104.
12. Silverman WK, Hinshaw SP. The second special issue on evidence-based psychosocial treatments for children and adolescents: A 10-year update. Journal of Clinical Child & Adolescent Psychology. 2008;37(1):1-7.
13. Chorpita BF, Daleiden EL. Mapping evidence-based treatments for children and adolescents: application of the distillation and matching model to 615 treatments from 322 randomized trials. Journal of consulting and clinical psychology. 2009;77(3):566.
14. Zhou X, Hetrick SE, Cuijpers P, Qin B, Barth J, Whittington CJ, et al. Comparative efficacy and acceptability of psychotherapies for depression in children and adolescents: A systematic review and network meta-analysis. World psychiatry. 2015;14(2):207-22.
15. Weisz JR, Kuppens S, Eckshtain D, Ugueto AM, Hawley KM, Jensen-Doss A. Performance of evidence-based youth psychotherapies compared with usual clinical care: a multilevel meta-analysis. JAMA psychiatry. 2013;70(7):750-61.
16. Weisz JR, Jensen-Doss A, Hawley KM. Evidence-based youth psychotherapies versus usual clinical care: a meta-analysis of direct comparisons. American Psychologist. 2006;61(7):671.
17. Spielmans GI, Gatlin ET, McFall JP. The efficacy of evidence-based psychotherapies versus usual care for youths: Controlling confounds in a meta-reanalysis. Psychotherapy Research. 2010;20(2):234-46.

18. Howard KI, Moras K, Brill PL, Martinovich Z, Lutz W. Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. American Psychologist. 1996;51(10):1059.
19. Ogles BM, Lunnen KM, Bonesteel K. Clinical significance: History, application, and current practice. Clinical psychology review. 2001;21(3):421-46.
20. Jensen SA, Corralejo SM. Measurement Issues: Large effect sizes do not mean most people get better–clinical significance and the importance of individual results. Child and Adolescent Mental Health. 2017;22(3):163-6.
21. Wolpert M. Commentary: Why measuring clinical change at the individual level is challenging but crucial–commentary on Jensen and Corralejo (2017). Child and Adolescent Mental Health. 2017;22(3):167-9.
22. Edbrooke-Childs J, Wolpert M, Zamperoni V, Napoleone E, Bear H. Evaluation of reliable improvement rates in depression and anxiety at the end of treatment in adolescents. BJPsych open. 2018;4(4):250-5.
23. Wolpert M, Görzig A, Deighton J, Fugard AJ, Newman R, Ford T. Comparison of indices of clinically meaningful change in child and adolescent mental health services: difference scores, reliable change, crossing clinical thresholds and 'added value'–an exploration using parent rated scores on the SDQ. Child and Adolescent Mental Health. 2015;20(2):94-101.
24. Smith AM, Jensen-Doss A. Youth psychotherapy outcomes in usual care and predictors of outcome group membership. Psychological services. 2017;14(1):66.
25. Murphy JM, Blais M, Baer L, McCarthy A, Kamin H, Masek B, et al. Measuring outcomes in outpatient child psychiatry: Reliable improvement, deterioration, and clinically significant improvement. Clinical child psychology and psychiatry. 2015;20(1):39-52.
26. Krause KR, Bear HA, Edbrooke-Childs J, Wolpert M. what outcomes count? A review of outcomes measured for adolescent depression between 2007 and 2017. Journal of the American Academy of Child & Adolescent Psychiatry. 2018.
27. Kovacs M, Obrosky S, George C. The course of major depressive disorder from childhood to young adulthood: Recovery and recurrence in a longitudinal observational study. Journal of affective disorders. 2016;203:374-81.
28. Nilsen TS, Handegård BH, Eisemann M, Kvernmo S. Predictors of rate of change for children and youth with emotional disorders: a naturalistic observational study. Child and adolescent psychiatry and mental health. 2016;10(1):11.
29. Gordon M, Antshel KM, Lewandowski L. Predictors of treatment outcome in a child and adolescent psychiatry clinic: A naturalistic exploration. Children and Youth Services Review. 2012;34(1):213-7.
30. Jackson DS, Keir SS, Sender M, Mueller CW. Reliable change and outcome trajectories across levels of care in a mental health system for youth. Administration and Policy in Mental Health and Mental Health Services Research. 2017;44(1):141-54.
31. Brookman-Frazee L, Haine RA, Garland AF. Innovations: child and adolescent psychiatry: measuring outcomes of real-world youth psychotherapy: whom to ask and what to ask? Psychiatric Services. 2006;57(10):1373-5.
32. Hudson JL, Keers R, Roberts S, Coleman JR, Breen G, Arendt K, et al. Clinical predictors of response to cognitive-behavioral therapy in pediatric anxiety disorders: the Genes for Treatment (GxT) study. Journal of the American Academy of Child & Adolescent Psychiatry. 2015;54(6):454-63.
33. Spielmans GI, Flückiger C. Moderators in psychotherapy meta-analysis. Psychotherapy Research. 2018;28(3):333-46.
34. Wolpert M, Rousmaniere T. Talking failure in therapy and beyond. PSYCHOLOGIST. 2017;30:40-3.
35. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. Journal of Epidemiology & Community Health. 1998;52(6):377-84.

36. Deeks J, Dinnes J, D'amico R, Sowden A, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. Health technology assessment (Winchester, England). 2003;7(27):iii-x, 1-173.
37. Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. journal of Educational Statistics. 1981;6(2):107-28.
38. Hedges L, Olkin I. Statistical models for meta-analysis. New York: Academic Press; 1985.
39. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychological methods. 2002;7(1):105.
40. Maas CJ, Hox JJ, Lensvelt-Mulders GJ. Longitudinal meta-analysis. Quality and Quantity. 2004;38(4):381-9.
41. Mufson L, Dorta KP, Wickramaratne P, Nomura Y, Olfson M, Weissman MM. A randomized effectiveness trial of interpersonal psychotherapy fordepressed adolescents. Archives of general psychiatry. 2004;61(6):577-84.
42. Young JF, Mufson L, Davies M. Impact of comorbid anxiety in an effectiveness study of interpersonal psychotherapy for depressed adolescents. Journal of the American Academy of Child & Adolescent Psychiatry. 2006;45(8):904-12.
43. Gunlicks-Stoessel M, Mufson L. Early patterns of symptom change signal remission with interpersonal psychotherapy for depressed adolescents. Depression and anxiety. 2011;28(7):525-31.
44. Gunlicks-Stoessel M, Mufson L, Jekal A, Turner JB. The impact of perceived interpersonal functioning on treatment for adolescent depression: IPT-A versus treatment as usual in school-based health clinics. Journal of consulting and clinical psychology. 2010;78(2):260.
45. McGlinchey EL, Reyes-Portillo JA, Turner JB, Mufson L. Innovations in Practice: The relationship between sleep disturbances, depression, and interpersonal functioning in treatment for adolescent depression. Child and Adolescent Mental Health. 2017;22(2):96-9.
46. Reyes-Portillo JA, McGlinchey EL, Yanes-Lukin PK, Turner JB, Mufson L. Mediators of interpersonal psychotherapy for depressed adolescents on outcomes in Latinos: The role of peer and family interpersonal functioning. Journal of Latina/o Psychology. 2017;5(4):248.
47. Asarnow JR, Jaycox LH, Duan N, LaBorde AP, Rea MM, Murray P, et al. Effectiveness of a quality improvement intervention for adolescent depression in primary care clinics: a randomized controlled trial. Jama. 2005;293(3):311-9.
48. Trask EV, Garland AF. Are children improving? Results from outcome measurement in a large mental health system. Administration and Policy in Mental Health and Mental Health Services Research. 2012;39(3):210-20.
49. Chorpita BF, Weisz JR, Daleiden EL, Schoenwald SK, Palinkas LA, Miranda J, et al. Long-term outcomes for the Child STEPs randomized effectiveness trial: A comparison of modular and standard treatment designs with usual care. Journal of consulting and clinical psychology. 2013;81(6):999.
50. Porter S, McConnell T, McLaughlin K, Lynn F, Cardwell C, Braiden Hj, et al. Music therapy for children and adolescents with behavioural and emotional problems: a randomised controlled trial. Journal of Child Psychology and Psychiatry. 2017;58(5):586-94.
51. Mantzouranis G, Baier V, Holzer L, Urben S, Villard E. Clinical significance of assertive community treatment among adolescents. Social psychiatry and psychiatric epidemiology. 2018:1-9.
52. Hayes L, Boyd CP, Sewell J. Acceptance and commitment therapy for the treatment of adolescent depression: A pilot study in a psychiatric outpatient setting. Mindfulness. 2011;2(2):86-94.
53. Burcusa SL, Iacono WG. Risk for recurrence in depression. Clinical psychology review. 2007;27(8):959-85.
54. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. The Lancet. 2013;382(9904):1575-86.
55. Neufeld SA, Dunn VJ, Jones PB, Croudace TJ, Goodyer IM. Reduction in adolescent depression after contact with mental health services: a longitudinal cohort study in the UK. The Lancet Psychiatry. 2017;4(2):120-7.

56. RCPCH. Royal College of Paediatrics and Child Health. National Paediatric Diabetes Audit. . 2015.
57. Edbrooke-Childs J, Jacob J, Law D, Deighton J, Wolpert M. Interpreting standardized and idiographic outcome measures in CAMHS: what does change mean and how does it relate to functioning and experience? Child and Adolescent Mental Health. 2015;20(3):142-8.
58. Clark DM, Canvin L, Green J, Layard R, Pilling S, Janecka M. Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. The Lancet. 2018;391(10121):679-86.
59. Cannon JA, Warren JS, Nelson PL, Burlingame GM. Change trajectories for the Youth Outcome Questionnaire Self-Report: Identifying youth at risk for treatment failure. Journal of Clinical Child & Adolescent Psychology. 2010;39(3):289-301.
60. Warren JS, Nelson PL, Mondragon SA, Baldwin SA, Burlingame GM. Youth psychotherapy change trajectories and outcomes in usual care: Community mental health versus managed care settings. Journal of Consulting and Clinical Psychology. 2010;78(2):144.
61. Napoleone E, Evans C, Patalay P, Edbrooke-Childs J, Wolpert M. Trajectories of change of youth depressive symptoms in routine care: shape, predictors, and service-use implications. European child & adolescent psychiatry. 2019:1-10.
62. Peselow ED, Karamians R, Lord M, Tobia G, IsHak WW. The discrepancy between patients and informants on clinician-rated measures in major depressive disorder: implications for clinical trials and clinical practice. International clinical psychopharmacology. 2014;29(2):111-5.
63. Mulley A, Coulter A, Wolpert M, Richards T, Abbasi K. New approaches to measurement and management for high integrity health systems. Bmj. 2017;356:j1401.
64. Johnsen TJ, Friborg O. The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. Psychological Bulletin. 2015;141(4):747.
65. Cristea IA, Stefan S, Karyotaki E, David D, Hollon SD, Cuijpers P. The effects of cognitive behavioral therapy are not systematically falling: A revision of Johnsen and Friborg (2015). 2017.
66. Patalay P, Fitzsimons E. Correlates of mental illness and wellbeing in children: are they the same? Results from the UK Millennium Cohort Study. Journal of the American Academy of Child & Adolescent Psychiatry. 2016;55(9):771-83.
67. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. Bmj. 2014;348:g1687.
68. Bachmann M, Bachmann CJ, John K, HEINZEL-GUTENBRUNNER M, Remschmidt H, Mattejat F. The effectiveness of child and adolescent psychiatric treatments in a naturalistic outpatient setting. World psychiatry. 2010;9(2):111-7.
69. Barrington J, Prior M, Richardson M, Allen K. Effectiveness of CBT versus standard treatment for childhood anxiety disorders in a community clinic setting. Behaviour Change. 2005;22(1):29-43.
70. Baruch G. Evaluating the outcome of a community-based psychoanalytic psychotherapy service for young people between 12 and 25 years old: Work in progress. Psychoanalytic Psychotherapy. 1995;9(3):243-67.
71. Baruch G, Fearon P. The evaluation of mental health outcome at a community-based psychodynamic psychotherapy service for young people: A 12-month follow-up based on self-report data. Psychology and Psychotherapy: Theory, research and practice. 2002;75(3):261-78.
72. Barwick M, Urajnik D, Sumner L, Cohen S, Reid G, Engel K, et al. Profiles and service utilization for children accessing a mental health walk-in clinic versus usual care. Journal of evidence-based social work. 2013;10(4):338-52.
73. Biegel GM, Brown KW, Shapiro SL, Schubert CM. Mindfulness-based stress reduction for the treatment of adolescent psychiatric outpatients: A randomized clinical trial. Journal of consulting and clinical psychology. 2009;77(5):855.

74. Carter T, Guo B, Turner D, Morres I, Khalil E, Brighton E, et al. Preferred intensity exercise for adolescents receiving treatment for depression: a pragmatic randomised controlled trial. BMC psychiatry. 2015;15(1):247.
75. Clarke G, Debar L, Lynch F, Powell J, Gale J, O'connor E, et al. A randomized effectiveness trial of brief cognitive-behavioral therapy for depressed adolescents receiving antidepressant medication. Journal of the American Academy of Child & Adolescent Psychiatry. 2005;44(9):888-98.
76. Clarke G, DeBar LL, Pearson JA, Dickerson JF, Lynch FL, Gullion CM, et al. Cognitive behavioral therapy in primary care for youth declining antidepressants: a randomized trial. Pediatrics. 2016:e20151851.
77. Deighton J, Argent R, De Francesco D, Edbrooke-Childs J, Jacob J, Fleming I, et al. Associations between evidence-based practice and mental health outcomes in child and adolescent mental health services. Clinical child psychology and psychiatry. 2016;21(2):287-96.
78. Nemirovski Edlund J, Thorén A, Carlberg G. Outcome of psychodynamic child psychotherapy in routine practice. European Journal of Psychotherapy & Counselling. 2014;16(3):228-44.
79. Edlund JN, Carlberg G. Psychodynamic psychotherapy with adolescents and young adults: Outcome in routine practice. Clinical child psychology and psychiatry. 2016;21(1):66-80.
80. Goldbeck L, Ellerkamp T. A randomized controlled trial of multimodal music therapy for children with anxiety disorders. Journal of Music Therapy. 2012;49(4):395-413.
81. Jónsson H, Thastum M, Arendt K, Juul-Sørensen M. Group cognitive behavioural treatment of youth anxiety in community based clinical practice: Clinical significance and benchmarking against efficacy. Journal of anxiety disorders. 2015;35:9-18.
82. Kamin HS, McCarthy AE, Abel MR, Jellinek MS, Baer L, Murphy JM. Using a brief parent-report measure to track outcomes for children and teens with internalizing disorders. Child Psychiatry & Human Development. 2015;46(6):851-62.
83. Kenaley BLD, Williams NJ. A preliminary evaluation: Demographic and clinical profiles and changes in functioning in children receiving psychosocial rehabilitation. Children and Youth Services Review. 2011;33(2):301-7.
84. Kobak KA, Mundt JC, Kennard B. Integrating technology into cognitive behavior therapy for adolescent depression: a pilot study. Annals of general psychiatry. 2015;14(1):37.
85. Lundh A, Forsman M, Serlachius E, Lichtenstein P, Landén M. Outcomes of child psychiatric treatment. Acta Psychiatrica Scandinavica. 2013;128(1):34-44.
86. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MF. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. Bmj. 2012;344:e2598.
87. Muratori F, Picchi L, Bruni G, Patarnello M, Romagnoli G. A two-year follow-up of psychodynamic psychotherapy for internalizing disorders in children. Journal of the American Academy of Child & Adolescent Psychiatry. 2003;42(3):331-9.
88. Nilsen TS, Handegård B-H, Eisemann M, Kvernmo S. Evaluating change in symptomatic and functional level of children and youth with emotional disorders: a naturalistic observation study. European child & adolescent psychiatry. 2015;24(10):1219-31.
89. O'Brien F, Olden N, Migone M, Dooley B, Atkins L, Ganter K, et al. Group cognitive behavioural therapy for children with anxiety disorder–an evaluation of the 'Friends for Youth'programme. Irish journal of psychological medicine. 2007;24(1):5-12.
90. Richardson LP, Ludman E, McCauley E, Lindenbaum J, Larison C, Zhou C, et al. Collaborative care for adolescents with depression in primary care: a randomized clinical trial. Jama. 2014;312(8):809-16.
91. Sanford M, Boyle M, McCleary L, Miller J, Steele M, Duku E, et al. A pilot study of adjunctive family psychoeducation in adolescent major depression: feasibility and treatment effect. Journal of the American Academy of Child & Adolescent Psychiatry. 2006;45(4):386-95.
92. Sharma P, Mehta M, Sagar R. Efficacy of transdiagnostic cognitive-behavioral group therapy for anxiety disorders and headache in adolescents. Journal of anxiety disorders. 2017;46:78-84.

93. Shirk SR, DePrince AP, Crisostomo PS, Labus J. Cognitive behavioral therapy for depressed adolescents exposed to interpersonal trauma: An initial effectiveness trial. Psychotherapy. 2014;51(1):167.
94. Southam-Gerow MA, Weisz JR, Chu BC, McLeod BD, Gordis EB, Connor-Smith JK. Does cognitive behavioral therapy for youth anxiety outperform usual care in community clinics? An initial effectiveness test. Journal of the American Academy of Child & Adolescent Psychiatry. 2010;49(10):1043-52.
95. Strandholm T, Karlsson L, Kiviruusu O, Pelkonen M, Marttunen M. Treatment characteristics and outcome of depression among depressed adolescent outpatients with and without comorbid axis II disorders. Journal of personality disorders. 2014;28(6):853-63.
96. Sultan MA, Courtney DB. Adjunctive trazodone and depression outcome in adolescents treated with serotonin re-uptake inhibitors. Journal of the Canadian Academy of Child and Adolescent Psychiatry. 2017;26(3):233-40.
97. Walter D, Dachs L, Faber M, Goletz H, Goertz-Dorten A, Hautmann C, et al. Effectiveness of outpatient cognitive-behavioral therapy for adolescents under routine care conditions on behavioral and emotional problems rated by parents and patients: an observational study. European child & adolescent psychiatry. 2018;27(1):65-77.
98. Weersing VR, Weisz JR. Community clinic treatment of depressed youth: benchmarking usual care against CBT clinical trials. Journal of consulting and clinical psychology. 2002;70(2):299.
99. Weersing VR, Iyengar S, Kolko DJ, Birmaher B, Brent DA. Effectiveness of cognitive-behavioral therapy for adolescent depression: A benchmarking investigation. Behavior Therapy. 2006;37(1):36-48.
100. Weisz JR, Southam-Gerow MA, Gordis EB, Connor-Smith JK, Chu BC, Langer DA, et al. Cognitive–behavioral therapy versus usual clinical care for youth depression: An initial test of transportability to community clinics and clinicians. Journal of consulting and clinical psychology. 2009;77(3):383.
101. Weitkamp K, Daniels JK, Baumeister-Duru A, Wulf A, Romer G, Wiegand-Grefe S. Effectiveness trial of psychoanalytic psychotherapy for children and adolescents with severe anxiety symptoms in a naturalistic treatment setting. British journal of psychotherapy. 2018;34(2):300-18.
102. Wiggins A, Oakley Browne M, Bearsley-Smith C, Villanueva E. Depressive disorders among adolescents managed in a child and adolescent mental health service. Australasian Psychiatry. 2010;18(2):134-41.

Table 1. Characteristics of Included Studies

Study identifier	Setting	Design	Recruited/ referred	TAU sample, *n*[a]	Age, mean ± SD (range)	% Female participants	% Ethnic minority	Presenting problem	Diagnosis/ cut-off requirement [b]
Bachmann et al. (2010)[68]	Germany; 9 child and adolescent psychiatric practices	Observational	Referred	Depressive disorder, 38; anxiety disorder, 53 (total sample 306)	8.8±3.3[c]	40.2 [c]	Not reported	Anxiety disorder and depressive disorder	Yes (diagnostic interview)
Barrington, Prior, Richardson, and Allen (2005)[69]	Australia; 1 CAMHS	RCT	Referred, recruited	26 [c]	10.2±1.75	69.2	Not reported	Anxiety disorder	Yes (DSM-IV)
Baruch (1995)[70]	UK; 1 community based psychodynamic psychotherapy centre	Observational	Referred	106 [c]	18.7±3.2 (12-25) [c]	73.8 [c]	22	Internalising problems (84% mood disorder)	Yes (ICD-10)
Baruch and Fearon (2002)[71]	UK; 1 community based psychodynamic psychotherapy centre	Observational	Referred	151 [c]	19.7±3.2 [c]	67.5[c]	Not reported	Internalising problems (52.7% depression)	Yes (ICD-10)
Barwick et al. (2013)[72]	Canada; 1 child and family centre	Quasi-experimental	Referred, recruited	56 [c]	10.4±3.64 [c]	35.7 [c]	48.2	Internalising behaviours	No
Biegel, Brown, Shapiro and Schubert (2009)[73]	USA; 1 outpatient child and adolescent psychiatry department	RCT	Referred, recruited	Mood disorder, 23; anxiety disorder, 14 (total sample 52)	15.0±1.19 (14-18) [c]	76.9 [c]	48.1	Mood disorder, anxiety disorder	No
Carter et al. (2015)[74]	UK; academic setting	RCT	Recruited	43	15.4±0.9	81	3	Depression	Yes (CDI-2 > 14)
Clarke et al. (2005)[75]	USA; paediatric primary care	RCT	Recruited	75	15.32±1.6	77.33	15.07	Depression	Yes (DSM-IV)
Clarke et al. (2016)[76]	USA; primary care	RCT	Recruited	106	14.6±1.7(12-18)[c]	68.4 [c]	11.8	Depression	Yes
Deighton et al. (2016)[77]	UK; CAMHS	Observational	Referred	490	13.9 (11-18)	59.5	19.4	Emotional disorders	No

Edbrooke-Childs, Wolpert, Zamperoni, Napoleone and Bear (2018)[22]	UK; CAMHS	Observational	Referred	Anxiety, 1208; depression, 621; comorbid anxiety and depression, 2635 (4464 total sample)	14.5±1.9 (8-18)	74.9	8.4	Anxiety, depression, comorbid anxiety and depression	Yes (above threshold on RCADS)
Edlund, Thorén and Carlberg (2014)[78]	Sweden; public child and adolescent mental health clinic	Observational	Referred	Anxiety disorders, 66; mood disorders, 9 (207 total sample)	7.1±2.0 (4-12)[c]	41.28 [c]	Not reported	Mood disorder, anxiety disorder	Yes (DSM-IV)
Edlund and Carlberg (2016)[79]	Sweden; public child and adolescent mental health clinic	Observational	Referred	Anxiety disorders, 59; mood disorders, 69 (218 total sample)	19.17 ± 2.45 (14-24) [c]	76 [c]	Not reported	Mood disorder, anxiety disorder	Yes (DSM-IV)
Goldbeck and Ellerkamp (2012)[80]	Germany; outpatient clinic in the Department of Child and Adolescent Psychiatry and Psychotherapy at a University Medical Centre	RCT	Recruited	18	9.94±1.47 (8-12)	33.3	5.6	Anxiety disorder	Yes (DSM-IV)
Mufson et al. (2004)[41]	USA; 5 school-based health clinics	RCT	Referred, recruited	29	14.9±1.7 (12-18)	75.9	65.5	Depression	Yes (DSM-IV) + HDRS of 10 or higher + CGAS of 65 or lower
Hayes, Boyd and Sewell (2011)[52]	Australia; 1 public child and adolescent psychiatric service	RCT	Referred, recruited	16	15.49±1.35 (12-18)	56	Not reported	Depression	100% Met clinical criteria on the SDQ, 75% on RADS-2
Jónsson, Thastum, Arendt and Juul-Sørensen (2015)[81]	Denmark; 3 child and adolescent psychiatric centres and 4 school counselling centres	Observational	Referred, recruited	87	11.18±1.6 (7-16)	52.9	Not reported	Anxiety	Yes (DSM-IV)
Kamin et al. (2015)[82]	USA; 1 outpatient paediatric psychiatry service	Observational	Referred	900 (total sample 1692)	12.09 ± 3.81	49.8	Not reported	Internalising disorders	Yes (DSM-IV)
Kenaley and Williams (2011)[83]	USA; 1 for-profit outpatient child and	Observational	Referred	34 (total sample 53)	10.45 ± 4.0[c](4-18)	32.1 [c]	19	Internalising disorders	Yes (DSM-IV)

	adolescent mental health clinic								
Kobak, Mundt and Kennard (2015)[84]	USA; mixed	RCT	Referred, recruited	37	15.4± 1.52 (12-17) [c]	66 [c]	64.5	Depression	Yes (DSM-V)
Lundh, Forsman, Serlachius, Lichtenstein and Landén (2013)[85]	Sweden; multicentre CAMHS	Observational	Referred	Mood disorders, 2213; anxiety disorders, 2446 (12613 total sample)	anxiety disorders, 13.4 ±3.2; mood disorders, 14.4 ±2.5	Anxiety disorders, 66; mood disorders, 66	Not reported	Mood disorder, anxiety disorder	Yes (DSM-IV and/or ICD-10)
Merry et al. (2012)[86]	New Zealand; 15 school-based counselling services, 7 youth clinics, 2 general practices	RCT	Recruited	93	15.58±1.66 (12-19)	68.8	39.8	Depression	No
Muratori, Picchi, Bruni, Patarnello and Romagnoli (2003)[87]	Italy; Division of Child and Adolescent Neuropsychiatry	Quasi-experimental	Referred, recruited	29	8.7 ± 1.2	41.4	Not reported	Depression or anxiety	Yes (DSM-IV)
Nilsen, Handegård, Eisemann and Kvernmo (2015)[88]	Norway; 2 outpatient CAMHS clinics	Observational	Referred	84	12.49	66.7	9.8	Anxiety and/or depression	Yes (DSM-IV)
O'Brien et al. (2007)[89]	Ireland; 1 child and adolescent guidance centre	RCT	Referred, recruited	6	12.5±1.0 (7-15)	33.3	Not reported	Anxiety disorder	Yes (DSM-IV)
Richardson et al. (2014)[90]	USA; 9 paediatric and family medicine clinics	RCT	Recruited	51	15.5±1.3 (13-17)	72	33	Depression	Yes (PHQ-9 >10 and CDRS-R >42 or KIDDIE-SADS)
Sanford et al. (2006)[91]	Canada; 1 child and adolescent mood disorder clinic and 4 psychiatry programmes	RCT	Referred, recruited	15	16.1±1.6	73.3	Not reported	Major depression	Yes (DSM-IV)
Sharma, Mehta and Sagar (2017)[92]	India; 1 Child Guidance Clinic at a Department of Psychiatry	RCT	Recruited	31	13.87±2.2	48.4	Not reported	Anxiety	Yes (ICD-10)
Shirk, DePrince, Crisostomo and Labus (2014)[93]	USA; 1 outpatient community mental health centre	RCT	Referred, recruited	23	15.69±1.55 (13-17)	82.6	47.8	Depression	Yes (DSM-IV)

Southam-Gerow et al. (2010)[94]	USA; 6 public community mental health clinics	RCT	Referred, recruited	24	10.9±2.1 (8-15) [c]	56.2[c]	61.5	Anxiety	Yes (DSM-IV)
Strandholm, Karlsson, Kiviruusu, Pelkonen and Marttunen (2014)[95]	Finland; psychiatric outpatient clinics	Observational	Referred, recruited	98 (151 total sample)	13-19 [c]	83.7	Not reported	Depression	Yes (≥ 10 on BDI-21 and ≥ 5 on GHQ-36)
Sultan and Courtney (2017)[96]	Canada; mental health centre	Observational	Referred	Depressive disorder, 27; anxiety disorder, 19 (36 total sample)	15-18	75	Not reported	Depressive disorder, anxiety disorder	Yes (BDI-II > 20)
Walter et al. (2018)[97]	Germany; university outpatient clinic	Observational	Referred, recruited	Anxiety disorders, 115; depressive disorders, 106; emotional disorders, 74 (677 total sample)	14.3±2.2 (11-20.4) [c]	44.76 [c]	Not reported	Anxiety disorder, depressive disorder, emotional disorders	Yes (ICD-10)
Weersing and Weisz (2002)[98]	USA; 6 community mental health centres	Observational	Referred, recruited	67	12.9±2.6 (7-17)	55.2	52	Depression	Yes (DSM-III-R)
Weersing et al. (2006)[99]	USA; 1 psychiatric outpatient clinic	Observational	Referred	80	15.56±1.4	77	15	Depression	Yes (DSM-III-R or DSM-IV)
Weisz et al. (2009)[100]	USA; 7 community mental health clinics	RCT	Referred, recruited	25	11.48±2.37	56 [c]	67	Depression	Yes (DSM-IV)
Weitkamp et al. (2018)[101]	Germany; outpatient psychoanalytic child and adolescent psychotherapy clinic	Observational	Referred, recruited	86	13.51±4.42 (4-21)	69.8	Not reported	Anxiety	Yes (SCARE>15)
Wiggins et al. (2010)[102]	Australia; 1 regional CAMHS	Observational	Referred	17 (total sample 76)	15.7 (12-18)	65	Not reported	Depression	No (22% or 17/76) had a clinician diagnosis)

Note: CAMHS = child and adolescent mental health services; DSM-IV = Diagnostic and Statistical Manual of Mental Disorders, 4th Edition; ICD-10 = 10th revision of the International Statistical Classification of Diseases and Related Health Problems; RCADS = The Revised Children's Anxiety and Depression Scale; RCT = randomised-controlled trial; SCARED = Screen for Child Anxiety Related Emotional Disorders; SD = standard deviation; TAU = treatment as usual; UC = usual care; UK = United Kingdom; USA = United States of America.

[a] Study sample for routine care arm, not inclusive of other trial arms.

[b] Diagnosis required/ not required for eligibility in study.

[c] Data provided is averaged over the total study sample, rather than anxiety/depression sub-sample.

Table 2. Characteristics of TAU Interventions

Study identifier	Data collection period	Intervention description	Single therapeutic modality	Provider(s)	Delivery method	Dosage, mean sessions [a]	Medication usage, %	Intervention fidelity/ quality reported
Bachmann et al. (2010)[68]	2004-2006	Interdisciplinary (psychotherapy, CBT, psychodynamic psychotherapy, systemic and family therapy, pharmacotherapy)	No	Child and adolescent psychiatrists and psychotherapists	Not reported	7.65	26	Discussed
Barrington, Prior, Richardson, and Allen (2005)[69]	Not reported	Interdisciplinary non-CBT approaches (family therapy, play therapy, psychodynamic psychotherapy, supportive psychotherapy)	No	Therapists (*n*=9): doctoral level psychologists, social workers, senior psychiatric nurses, a psychiatrist and an occupational therapist.	Not reported	12.00	Not reported	Discussed
Baruch (1995)[70]	1993-1995	Open-ended psychoanalytic psychotherapy	Yes	Psychotherapists	Individual	17.00	Not reported	Not reported
Baruch and Fearon (2002)[71]	1993-2001	Open-ended psychoanalytic psychotherapy	Yes	Psychotherapists	Individual	Not reported	Not reported	Not reported
Barwick et al. (2013)[72]	2006-2008	CORE counselling programme	Yes	Not reported	Individual	Not reported	Not reported	Not reported
Biegel, Brown, Shapiro and Schubert (2009)[73]	2005-2006	Individual or group psychotherapy and/or psychotropic medication	No	Clinical staff	Mixed	Not reported	38.5 pre-test, 47.8 post-test	Not reported
Carter et al. (2015)[74]	Not reported	Counselling alone (41.8%) talking therapy in CAMHS (23.2%), CBT alone (0%), waiting list (11.6%), no treatment (11.6%)	No	Clinical staff	Individual	Not reported	19	Not reported
Clarke et al. (2005)[75]	2000-2001	Any non-study healthcare services or medications	No	Not reported	Not reported	5.00	100	Not reported
Clarke et al. (2016)[76]	2006-2010	Self-selected TAU: Year 1 outpatient mental health (48.1%), antidepressants (7.6%), other medication (11.3%), school counselling (25.5%). Year 2 outpatient mental health (58.5%), antidepressants (17.9%), other meds (23.6%), school counselling (35.9%)	No	Not reported	Not reported	Not reported	Year 1 7.6, year 2 17.9	Not reported
Deighton et al. (2016)[77]	Not reported	Usual CAMHS care: CBT (36.9%), child psychotherapy (1.8%), family therapy (8%), creative therapy (0.82%), drug treatment (4.9%), parent training (3.06%), parent other intervention (6.12%) counselling	No	Not reported	Mixed	13.40	4.90	Not reported

		(9.8%), other therapeutic intervention (21.43%)						
Edbrooke-Childs, Wolpert, Zamperoni, Napoleone and Bear (2018)[22]	2011-2015	Services were part of publicly funded provision. Emphasised evidence-based interventions for given problems. The majority of treatment provided drew on a range of interventions.	No	MDT (Child and adolescent psychiatrists, psychotherapists)	Not reported	10.24	Not reported	Not reported
Edlund, Thorén and Carlberg (2014)[78]	Not reported	Open-ended psychodynamic psychotherapy	Yes	Psychodynamically orientated child psychotherapists	Individual	65.90	Not reported	Not reported
Edlund and Carlberg (2016)[79]	2002-2009	Open-ended psychodynamic psychotherapy	Yes	Psychodynamically orientated child psychotherapists	Individual	43.00	Not reported	Not reported
Goldbeck and Ellerkamp (2012)[80]	2006-2008	Included brief behavioural interventions, psychodynamic psychotherapy, non-specific group therapy, 14.3% of TAU group received nothing due to waiting lists	No	Child psychiatrists, paediatricians, psychologists, child psychotherapists, social workers	Mixed	4.00	0	Not reported
Mufson et al. (2004)[41]	1999-2002	Standard psychological treatment provided by the school-based clinic. Varied psychotherapy resembling supportive counselling.	No	School mental health clinicians: social workers (*n*=11), doctoral level clinical psychologists (*n*=2) Predominantly dynamically trained	Mixed	7.90	9.5	Not reported
Hayes, Boyd and Sewell (2011)[52]	Not reported	Manualised CBT that includes psychoeducation, early warning signs planning, coping with unpleasant thoughts, increasing pleasant activities, problem analysis, problem solving, goal setting, crisis management	Yes	Therapists	Individual	15.58	Not reported	Discussed
Jónsson, Thastum, Arendt and Juul-Sørensen (2015)[81]	2011-2012	Cool Kids CBT programme	Yes	16 therapists: psychologists (*n*=14), occupational therapist (*n*=1) and child and adolescent psychiatrist (*n*=1), specialized in psychotherapy. 4 psychologists specialised in clinical psychology. 44% had no prior experience of treatment of youth anxiety, 56% had	Group	10.00	9	Not reported

				between 3 and 15 years of experience				
Kamin et al. (2015)[82]	2007-2013	Therapy only (32.1%), psychopharmacology (49.2%), combination treatment (18.7%)	No	Not reported	Individual	Not reported	49.2	Not reported
Kenaley and Williams (2011)[83]	2004-2007	Child psychosocial rehabilitation based on empowerment and cognitive-behavioural frameworks. Based on a biopsychosocial model of psychopathology	Yes	Bachelor-level CPSR Specialist	Individual	196.00	Not reported	Discussed
Kobak, Mundt and Kennard (2015)[84]	Not reported	Not reported	No	Clinical psychologists, educational psychologists, counsellors, behavioural mental health. 15/18 total sample had MSc or doctoral degrees. Mean number of years working with adolescents was 12.2 years.	Not reported	12.00	Not reported	Not reported
Lundh, Forsman, Serlachius, Lichtenstein and Landén (2013)[85]	2006-2010	Counselling and psychotherapy with different time-frames and settings. Individual, group and therapy. No data on specific psychotherapeutic method used	No	Professional psychotherapists and counsellors.	Mixed	13.10	Not reported	Not reported
Merry et al. (2012)[86]	2009-2010	Data on the nature of TAU available for 89% of participants: counselling (89.2%), waiting list for active treatment (13.3%), drugs (2.4%)	No	Not reported	Individual	4.80	2.40	Not reported
Muratori, Picchi, Bruni, Patarnello and Romagnoli (2003)[87]	Not reported	51.7% participants didn't follow any treatment and the other 48.3% followed various kinds of treatment (*n*=7 individual therapy, *n*=3 supportive therapy for parents, *n*=3 school tutoring)	No	Not reported	Mixed	12.00	0	Not reported
Nilsen, Handegård, Eisemann and Kvernmo (2015)[88]	2002-2005	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported
O'Brien et al. (2007)[89]	Not reported	Sessions consisted of a review of progress, review of medication and a generally supportive approach to managing difficulties	No	Psychiatrist or psychologist	Individual	4.00	50 pre-test, 0 post-test	Not reported

Richardson et al. (2014)[90]	2010-2013	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Discussed
Sanford et al. (2006)[91]	1999-2000	Counselling and/or drug therapy with supportive case management. Treatments not standardised with respect to frequency of follow-up or fidelity to a specific treatment model	No	Not reported	Mixed	Not reported	86.7	Discussed
Sharma, Mehta and Sagar (2017)[92]	Not reported	Pharmacotherapy. During visits only medication was reviewed and no psychotherapy sessions were held	Yes	Psychiatrists	Individual	6.00	100	Not reported
Shirk, DePrince, Crisostomo and Labus (2014)[93]	Not reported	Eclectic with client-centred, family and psychodynamic interventions favoured. Treatment did not follow a specific manual and was based on therapists' case formulations. Therapists agreed to use treatment strategies and procedures that they regularly used and believed to be effective in their clinical practice	No	Clinic based therapists (*n*=4) doctoral-level psychologists (*n*=2) with 3 and 4 years of clinical experience respectively. Therapists were eclectic, with client-cantered, psychodynamic, and family interventions favoured.	Mixed	7.22	22.2	Discussed
Southam-Gerow et al. (2010)[94]	Not reported	Treatment procedures normally used and believed to be effective. In general, the TAU therapists used a range of treatment procedures consistent with multiple theoretical orientations	No	Therapists (*n*=21): social workers (27.3%), doctoral level psychologists (9.1%), master's level psychologists (51.5%), and other (12.1%). Averaged 4.49 years of training and 4.9 years of additional professional experience.	Mixed	Not reported	Not reported	Not reported
Strandholm, Karlsson, Kiviruusu, Pelkonen and Marttunen (2014) [95]	1998-2001	Regular, best available treatment at clinics. Psychosocial treatment consists of individual sessions (i.e. supportive therapy, psychotherapy) as the basis of treatment with family and network meetings. 48.4% of individual appointments were counselling, of the family/ network appointments, 43.3% was family counselling, 3.1% was family therapy and 53.6% was classified as other	No	Treatment teams consist of an adolescent psychiatrist in charge of the treatment, one or more psychologists, a social worker, one or more psychiatric nurses.	Mixed	20.48	56.1	Not reported
Sultan and Courtney (2017)[96]	2009-2014	Youth intensive outpatient program. Runs 4 days per week including 2.5 hours of therapeutic programming including	No	MDT	Mixed	48	100	Not reported

		psychotherapy groups and recreational groups.						
Walter et al. (2018)[97]	2008-2014	Treatment was delivered at the publicly funded University Outpatient Clinic for child and adolescent CBT. CBT based on currently recommended methods.	Yes	Therapists were post-graduate students with a Master's degree and in the second half of their training in child and adolescent CBT.	Individual	43.70	24.4	Information on the specific treatment components used was provided.
Weersing and Weisz (2002)[98]	Not reported	Range of psychodynamic, cognitive and behavioural techniques but endorsed significantly more psychodynamic techniques than either cognitive or behavioural	No	Therapists	Not provided	11.00	Not reported	Not reported
Weersing et al. (2006)[99]	1995-2000	Manualised CBT. Manual did not include session-by-session scripts for therapist behaviour or workbooks or homework.	Yes	Clinicians	Individual	19.50	65	Discussed
Weisz et al. (2009)[100]	1998-2003	Therapists used the treatment procedures they used regularly and believed to be effective in their clinical practice. Utilised client-centred most frequently, followed by family, psychodynamic and CBT.	No	Community clinic therapists (*n*=28): social workers (22%), 1doctoral-level psychologists (14%), master's level psychologists (56%) and other (8%). Averaged 4.3 years training and 2.4 years additional professional experience prior to the study	Mixed	20.52	33.3	Not reported
Weitkamp et al. (2018)[101]	2007-2010	Outpatient psychoanalytic child and adolescent psychotherapy. Predominantly child-focused, complemented by parent sessions. Usually on a ratio of 4:1. No therapy manual imposed. Therapy conducted twice per week.	Yes	26 therapists (81% female), all had a university degree in social pedagogy, education science or psychology and had completed board-certified degrees on psychoanalytical child and adolescent psychotherapy. On average 12 years of working experience.	Individual	94.04	7.0	Adherence to code of practice was checked with a retrospective treatment fidelity checklist filled out by the therapists at the end of treatment for each patient.
Wiggins et al. (2010)[102]	2008	Treatment consisted of 41 psychological interventions, medication, inpatient treatment, and liaison with the client's	No	Psychologists (*n*=3), social workers (*n*=2), mental health nurses	Mixed	Not reported	17	Discussed

	schools and other health professionals (both internal and external to the CAMHS team). Treatment often involved not only the adolescent, but also their parents, family, school and other individuals or organisations involved with the young person. T1: supportive intervention (87%), behavioural (73%), psychodynamic (80%), family therapy (20%), cognitive (13%). T4: supportive (50%), behavioural (63%), psychodynamic (50%), family (0%), cognitive (25%).	(*n*=7), occupational therapists (*n*=2), consultant psychiatrists (*n*=2), psychiatric registrar (*n*=1).

Note: CAMHS = child and adolescent mental health services; CBT = Cognitive behavioural therapy; MDT = multidisciplinary team; TAU = treatment as usual.

[a] In instances where the mean was not reported, the median was used instead.

Table 3. Individual-Level Change Data

Study	Sample	Sample, *n*	Outcome measure	Outcome domain	Informant	Assessment time (weeks)	Recovery (%)[a]	Reliable improvement (%)[b]	Reliable deterioration (%)[b]	No reliable change (%)[b]	Reliable recovery (%)[c]
Bachmann et al. (2010)[68]	Anxiety	53	CBCL	Problems	Caretaker	52	20	-	-	-	-
Bachmann et al. (2010)[68]	Depression	38	CBCL	Problems	Caretaker	52	25	-	-	-	-
Barrington et al. (2005)[69]	Anxiety	26	ADIS (DSM-IV)	Diagnosis	Clinician	12.9	50	-	-	-	-
Barrington et al. (2005)[69]	Anxiety	26	ADIS (DSM-IV)	Diagnosis	Clinician	25.8	69	-	-	-	-
Barrington et al. (2005)[69]	Anxiety	26	ADIS (DSM-IV)	Diagnosis	Clinician	52	68	-	-	-	-
Baruch (1995)[70e]	Internalising	49	YSR - Internalising	Problems	Self	12.9	12	24.5	6.1	69.4	-
Baruch (2002)[71e]	Internalising	151	YSR/YASR	Problems	Self	52	31.8	46.4	2.6	51	-
Biegel et al. (2009)[73]	Entire sample	52	SCL-90 (anxiety)	Anxiety Symptoms	Self	20.9	-	20	2.5	77.5	-
Biegel et al. (2009)[73e]	Entire sample	52	SCL-90 (depression)	Depression Symptoms	Self	20.9	-	27.5	10	62.5	-
Biegel et al. (2009)[73]	Entire sample	52	STAI-present	Anxiety Symptoms	Self	20.9	-	20	12.5	67.5	-
Biegel et al. (2009)[73]	Entire sample	52	STAI-past	Anxiety Symptoms	Self	20.9	-	40	12.5	47.5	-
Biegel et al. (2009)[73]	Mood disorder	21	DSM-IV	Diagnosis	Clinician	20.9	-10[d]	-	-	-	-
Biegel et al. (2009)[73]	Anxiety disorder	21	DSM-IV	Diagnosis	Clinician	20.9	-10[d]	-	-	-	-
Clarke et al. (2005)[75e]	Depression	58	CES-D	Symptoms	Self	52	31	-	-	-	-
Clarke et al. (2016)[76]	Depression	99	K-SADS	Diagnosis	Assessor	6	0	-	-	-	-
Clarke et al. (2016)[76]	Depression	95	K-SADS	Diagnosis	Assessor	12	12.1	-	-	-	-

Clarke et al. (2016)[76]	Depression	90	K-SADS	Diagnosis	Assessor	26	43.4	-	-	-	-
Clarke et al. (2016)[76]	Depression	87	K-SADS	Diagnosis	Assessor	52	68.7	-	-	-	-
Clarke et al. (2016)[76]	Depression	82	K-SADS	Diagnosis	Assessor	78	75.8	-	-	-	-
Clarke et al. (2016)[76]	Depression	91	K-SADS	Diagnosis	Assessor	104	78.8	-	-	-	-
Edbrooke-Childs et al. (2018)[22e]	Anxiety	1208	RCADS	Symptoms	Self	26.39	60.9	52.8	7.7	33.1	45.9
Edbrooke-Childs et al. (2018)[22e]	Depression	621	RCADS	Symptoms	Self	26.39	56.4	44.3	3.7	52	41.5
Edbrooke-Childs et al. (2018)[22e]	Comorbid Anxiety and Depression	2635	RCADS	Symptoms	Self	26.39	38.7	34.6	9.7	28.5	25.5
Edlund et al. (2014)[78]	Anxiety Disorder	66	CGAS	Global Functioning	Clinician	43.9	-	82	1	17	49
Edlund et al. (2014)[78]	Mood Disorder	9	CGAS	Global Functioning	Clinician	43.9	-	44	0	56	44
Edlund and Carlberg (2016)[79e]	Anxiety Disorder	53	SCL-90	Symptoms	Self	43	-	71.7	11.3	17	51
Edlund and Carlberg (2016)[79e]	Mood Disorder	51	SCL-90	Symptoms	Self	43	-	74	10	16	55
Edlund and Carlberg (2016)[79]	Anxiety Disorder (>20 years)	28	CGAS	Global Functioning	Clinician	43	-	61	0	39	36
Edlund and Carlberg (2016)[79]	Mood Disorder(>20 years)	36	CGAS	Global Functioning	Clinician	43	-	64	0	36	42
Edlund and Carlberg (2016)[79]	Anxiety Disorder (>20 years)	30	GAF	Global Functioning	Clinician	43	-	77	0	23	60
Edlund and Carlberg (2016)[79]	Mood Disorder(>20 years)	27	GAF	Global Functioning	Clinician	43	-	58	0	42	39
Goldbeck and Ellerkamp (2012)[80]	Anxiety	18	K-SADS	Diagnosis	Clinician	25.9	33	-	-	-	-

Hayes et al. (2011)[52e]	Depression	11	RADS-2	Symptoms	Self	15.58	-	36	0	64	-
Hayes et al. (2011)[52]	Depression	4	RADS-2	Symptoms	Self	28.48	-	25	50	25	-
Hayes et al. (2011)[52]	Depression	11	SDQ	Symptoms	Self	15.58	-	0	0	100	-
Hayes et al. (2011)[52]	Depression	4	SDQ	Symptoms	Self	28.48	-	0	0	100	-
Jónsson et al. (2015)[81]	Anxiety	87	ADIS	Diagnosis	Clinician	12	46	-	-	-	-
Jónsson et al. (2015)[81]	Anxiety	87	ADIS	Diagnosis	Clinician	24.9	59.8	-	-	-	-
Jónsson et al. (2015)[81e]	Anxiety	85	SCAS	Symptoms	Self	12	15.3	28.2	5.9	65.9	37.1
Jónsson et al. (2015)[81]	Anxiety	85	SCAS	Symptoms	Mother	12	40	58.8	2.4	38.8	50
Jónsson et al. (2015)[81]	Anxiety	82	SCAS	Symptoms	Father	12	21.4	45.1	3.7	51.2	29
Kamin et al. (2015)[82]	Internalising	372	PSC-IS	Psychosocial Functioning	Parent	12.9	-	18.4	8.1	51.4	17.3
Merry et al. (2012)[86]	Depression	93	CDRS-R	Symptoms	Observer	8.4	35.5	-	-	-	-
Merry et al. (2012)[86]	Depression	93	CDRS-R	Symptoms	Observer	21.5	52.7	-	-	-	-
Mufson et al. (2004)[41]	Depression	29	HRSD	Symptoms	Clinician	12	34	-	-	-	-
Mufson et al. (2004)[41e]	Depression	29	BDI	Symptoms	Self	12	52	-	-	-	-
Muratori et al. (2003)[87]	Internalising	29	CGAS	Global Functioning	Clinician	25.8	44.83	-	-	-	-
Muratori et al. (2003)[87]	Internalising	29	CGAS	Global Functioning	Clinician	104	44.83	-	-	-	-
Nilsen et al. (2015)[88]	Anxiety and/or depression	39	CGAS	Global Functioning	Clinician	21.52	16.28	2.33	2.33	60.46	9.3
Nilsen et al. (2015)[88e]	Anxiety and/or depression	13	SDQ	Symptoms	Self	21.52	7.69	15.38	0	84.61	7.69

Richardson et al. (2014)[90]	Depression	51	CDRS-R	Symptoms	Observer	52	38.6	-	-	-	-
Richardson et al. (2014)[90e]	Depression	51	PHQ-9	Symptoms	Self	52	20.7	-	-	-	-
Sanford et al. (2006)[91]	Depression	15	K-SADS	Diagnosis	Clinician	24	17	-	-	-	-
Sanford et al. (2006)[91]	Depression	15	K-SADS	Diagnosis	Clinician	36	14	-	-	-	-
Shirk et al. (2014)	Depression	23	K-SADS	Diagnosis	Clinician	16	48	-	-	-	-
Southam-Gerow et al. (2010)[94]	Anxiety	24	DISC	Diagnosis	Clinician	28	58.3	-	-	-	-
Walter et al. (2018)[97e]	Internalising	344	YSR	Symptoms	Self	70	-	12.8	2.3	40.4	44.5
Walter et al. (2018)[97]	Internalising	512	CBCL	Symptoms	Parent	70	-	16.5	2.9	52.3	28.3
Weisz et al. (2009)[100]	Depression	24	DISC	Diagnosis	Clinician	39	77.3	-	-	-	-
Weitkamp et al. (2018)[101e]	Anxiety	58	SCARED	Symptoms	Self	24	-	17.2	10.3	20.7	51.7

Note: ADIS = Anxiety Disorders Interview Schedule; BDI = Beck Depression Inventory; CBCL = Child Behaviour Checklist; CDRS-R = Children's Depression Rating Scale-Revised; CES-D = Center for Epidemiologic Studies Depression Scale; CGAS = Children's Global Assessment Scale; DISC = Diagnostic Interview Schedule for Children; DSM-IV = Diagnostic and Statistical Manual of Mental Disorders, 4th Edition; GAF = Global Assessment of Functioning; HoNOSCA = Health of the Nation Outcome Scales for children and Adolescents; HRSD = Hamilton Rating Scale for Depression; K-SADS = Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version; PHQ-9 = Patient Health Questionnaire; PSC-IS = Paediatric Symptom Checklist – Internalising Subscale; RADS-2 = Reynolds Adolescent Depression Scale; RCADS = The Revised Children's Anxiety and Depression Scale; SCARED = Screen for Child Anxiety Related Emotional Disorders; SCAS = The Spence Children's Anxiety Scale; SCL-90 = Symptom Checklist-90; SDQ = Strengths and Difficulties Questionnaire; STAI-past = State-Trait Anxiety Inventory, trait; STAI-present = State-Trait Anxiety Inventory, state; YASR = Young Adult Self Report Form; YSR = Youth Self Report.

[a] Recovery, represents proportion of participants who moved from above a clinical threshold to below a clinical threshold.
[b] Reliable improvement, reliable deterioration and no reliable change metrics consider whether improvement, deterioration or no change in symptoms is greater than could likely be solely attributed to measurement error.
[c] Reliable recovery is the proportion of participants who experienced both reliable improvement and recovery.
[d] The -10 in this instance refers to 0% clinical improvement and 10% clinical deterioration.
[e] Indicates samples included in meta-analysis.