

The NICE guideline on the treatment of child and adolescent depression: a meta-review of the evidence for individual CBT

The National Institute for Clinical Excellence in the UK published its guideline on the treatment of children and adolescents who have been diagnosed with depression in 2005. Although the guideline has not been updated since, it is widely used in Child and Adolescent Mental Health Services in England and Wales to advocate that all children and young people who have been diagnosed with depression should have access to individual cognitive-behavioural therapy (CBT) through the Children and Adolescent Improving Access to Psychological Therapies (CYP-IAPT) project. This article critically reviews the guideline in terms of the evidence that NICE cites for the efficacy of individual CBT. In particular, it provides a meta-review of four randomised control trials where the effectiveness of individual CBT was compared to another psychological therapy intervention, a non-specific control intervention, and/or a wait-list. Each trial is considered separately, before looking at the overall evidence that they provide when the findings are considered as a whole. A trial comparing individual CBT to a non-psychological intervention (medication) is discussed separately. This review found that on present evidence, individual CBT cannot be viewed as an evidence-based psychological therapy for children and young people who have been diagnosed with depression.

The road of good intentions has gone dry as a bone.

- Bruce Springsteen (2012)

Introduction

The Department of Health (1996) states that clinical practice guidelines are developed to help clinicians and patients in making decisions about appropriate treatment for specific conditions. Since its inception in 1999, the National Institute of Clinical Excellence (NICE) developed several clinical guidelines to outline which drugs and treatments should be available on the NHS in England and Wales. In September 2005 NICE published its guideline on *Depression in Children and Young People: Identification and Management in Primary, Community and Secondary Care*. Full and summary outlines of this guideline can be found at <http://www.nice.org.uk>. NICE (2005) states that the guideline will be

of relevance to children and young people from five years until their 18th birthday. There has been no update to the guideline since it was published in 2005, so this remains the only official guidance from NICE.

NICE (2005a, 2005b) cautions that its guidelines, including this one, should not be viewed as a substitute for professional knowledge and clinical judgement. Moreover, it acknowledges that the usefulness and applicability of their guidelines can be limited by a number of factors, including a lack of high-quality research evidence, the quality of the methodology used in the development of the guideline, and the generalisability of the research findings. NICE also states that its clinical guidelines are intended to improve the process and outcomes of healthcare, and that these guidelines can form the basis for education and training of healthcare professionals.

Since the publication of the guideline, a dominant narrative has emerged that individual cognitive-behavioural therapy (CBT) is superior to other psychological therapies and that it should be provided for all children and young people in Child and Adolescent Mental Health Services (CAMHS) in the UK, including those who have been diagnosed with depression. One of the most important vehicles for driving this narrative forward is the recently conceived *Children and Young People's Improving Access to Psychological Therapies* (CYP-IAPT) project. The supporting argument for individual CBT's prominence in the CYP-IAPT project is usually based on the view that children and young people with mental health difficulties should all receive effective, evidence-based psychological treatments as recommended by NICE (see e.g., Fonagy, 2012; Pugh, 2012). As a result, individual clinicians and services are coming under increasing pressure to 'transform' themselves and to be 'NICE compliant' when treating children and young people with mental health difficulties, including those who have been diagnosed with depression.

Given this major development in CAMHS provision in England, it is important to critically examine the quality and interpretation of the evidence for individual CBT for the treatment of children and young people who have been diagnosed with depression as presented in the NICE guideline. This paper presents the outcome of a detailed, in-depth review of four "well-conducted randomised control trials (RCTs)" (NICE, 2005a, p.17) that the guidance for the use of individual CBT is based on. A fifth trial (TADS, 2004), also included in the NICE guidance, compared individual CBT to non-psychological interventions (medication and placebo) and will be discussed separately from the main review.

In this review both quantitative and qualitative frameworks were used to evaluate the evidence and its interpretation. The results of the review are presented in two ways. Firstly, each of the four studies is presented separately. The aim here is to provide a brief account of each study, together with an analysis of the quality and interpretation of the evidence for individual CBT for the treatment of children and young people who have been diagnosed with depression. Next, using these four randomised control trials as the evidence base, the studies are compared to and contrasted with each other to provide a meta-analysis of the quality and interpretation of the evidence for individual CBT for the treatment of children and young people who have been diagnosed with depression. The paper then goes on to consider the implications of the results of this meta-analysis for clinical practice, in particular for the training of healthcare professionals in the CYP-IAPT programme.

Studies considered and included

NICE (2005a, p.17) states that it used “well-conducted randomised control trials (RCTs)” to form the evidence base for this guideline (as with all others). It acknowledges in passing that “there are a number of difficulties with the use of RCT’s in the evaluation of interventions in mental health” but concludes without further elaboration that “the RCT remains the most important method for establishing treatment efficacy” (NICE, 2005a, p.17). Whilst this latter statement is obviously contentious and not necessarily supported by the evidence, it is beyond the scope of this article to rehearse all the arguments here. For a comprehensive discussion about evidence based practice in psychology the reader is referred to the *American Psychological Association Presidential Task Force on Evidence-Based Practice* (APA, 2006). Interestingly, NICE also acknowledges that it primarily used English-language systematic reviews in their search for their evidence base. Although beyond the scope of this paper to discuss this in detail, the issue here is to acknowledge the partial and incomplete nature of the evidence identified and presented, and how the dominant language of English can be used to colonise a territory like ‘evidence-based practice’. Be that as it may, for the purpose of this article, the studies included as support of the guidance will be judged in terms of the standards set by NICE, that is, whether or not they are well-conducted randomised control trials.

Following a systematic search for RCTs, NICE concluded that 18 trials assessing the efficacy of psychological therapies for children and young people who have been diagnosed with depression were eligible for inclusion. Fourteen of these studies were from the USA, three from the UK/Europe, and

one from Puerto Rico. In total, the data on 1520 participants were used in these trials. The participants' ages ranged from five to 18 years old. The trials lasted between four and 36 weeks and were conducted between 1986 and 2004.

Studies evaluating individual CBT

Of the 18 trials, only five (Brent et al., 1997; Rosselló & Bernal, 1999; TADS, 2004; Vostanis, Feehan, Grattan & Bickerton, 1996; Wood, Harrington & Moore, 1996) used a comparison group who received individual CBT. For the purpose of this review, the TADS (2004) RCT is discussed separately. This study compares individual CBT to medication, rather than another psychological therapy. Moreover, as will be seen later, the exclusion of this study does not make a significant difference to the overall picture for the effectiveness of individual CBT. The other four studies will now be considered in turn to evaluate the quality of the evidence each generated. In addition, the interpretation of the evidence for each trial will be critically reviewed.

Wood et al. (1996)

According to the NICE guideline this controlled trial indicated a “clinically important improvement” (NICE, 2005a, p.95) for brief individual CBT compared with relaxation therapy. The decision to describe relaxation here as a therapy, is an interesting one. The researchers themselves start off by describing it as “relaxation training” (p. 737), a comparison treatment that would be “somewhat analogous to giving a placebo” (p. 739) and that “would have little effect on the core symptoms of depression” (p. 738). They labelled their relaxation training program as a no treatment control condition, not as an active treatment. In fact, the researchers went as far as to say that, prior to the study, they had expected relaxation to be “much inferior” (p. 740) to the individual CBT treatment intervention. However, when it came to the discussion of the findings, they reported that the individual CBT intervention led to “a greater reduction in depressive symptoms than a control *psychological intervention*” (p. 743, italics added). They conceded that the differences reported were based mainly on child ratings of psychopathology and that the parental ratings did not reach statistical significance. The researchers state that the study was based on 109 consecutive outpatient referrals to a service that specialised in the treatment of depression. The trial included participants aged between nine and 17 years who were diagnosed with major depressive disorder (MDD) or minor depression (MD) using the *Schedule for Affective Disorders and Schizophrenia Child Version* (K-SADS) (Puig-Antich &

Chambers, 1978). In addition, they had to obtain a score of 15 or more on the *Mood and Feelings Questionnaire* (MFQ) (Angold, Costello, Pickles & Winder, 1987). The MFQ is a 32-item self-report questionnaire (completed separately by children and parents) and is based on DSM-III-R (APA, 1987) criteria for depression. Of the 109, a total of 53 participants entered the trial and 48 completed treatment. The individual CBT treatment consisted of 5 – 8 sessions and was based on a group-based intervention that had been used in non-clinical samples in school settings.

The MFQ was used as the main outcome measure. At the post-treatment assessment the researchers reported a significantly lower mean level of depressive symptoms based on data obtained from the MFQs completed by the children. However, the parental ratings did not show a similar statistically significant lowering of depressive symptoms. This significantly weakens the impact of the finding initially reported. The researchers also stated that the children who had the individual CBT intervention were significantly more satisfied with treatment. It is unclear why this was reported as a significant finding given that they intentionally used a control treatment that would have little effect on the participants' symptoms of depression. More importantly, the researchers reported that the differences between the two groups were reduced at the two follow-up points: on most measures the two groups did not differ at the 3-month follow-up, and by six months the differences were "trivial in a statistical sense, and therefore it seems reasonable to conclude that *specific* treatment effects...genuinely did not persist" (p. 744). However, they go on to state that due to the design, the intention of their study was to detect only the short-term effect of the individual CBT treatment. Another, more plausible, interpretation is that the study was designed to detect differences over a period of at least nine months, but that it failed to show lasting effects of a significant nature for the individual CBT intervention over this period. At the same time, the participants in the relaxation group continued to recover as time went on. Taken together, these findings might provide some support for the views that there are relatively high rates of spontaneous remission for untreated depression (e.g., Whiteford et al., 2013) and that different psychological therapy models are equally efficacious (e.g., Luborsky et al., 2002; Messer & Wampold, 2002; Rosenzweig, 1936; Wampold et al., 1997).

Vostanis et al. (1996)

In this RCT individual CBT was compared with a "non-focussed intervention" (p. 105). Again, in the NICE guideline it is stated that this study has indicated "clinically important improvement" for individual CBT compared with "non-directive supportive therapy" (2005a, p.95). However, the

researchers themselves make it clear that they decided “not to compare two active treatments” and that their non-focused intervention offered “no interpretations, solutions or suggestions” (2005a, p. 109) to the participants. Any claims that this trial compared individual CBT with another form of psychological therapy is clearly misleading.

In total, 117 potential participants were recruited from four Departments of Child and Adolescent Psychiatry for this study. Diagnosis of participants was based on DSM-III-R criteria for depression (APA, 1987). Although the researchers acknowledged that “these criteria are predominantly based on validation studies with adults without taking a developmental perspective” (p.106), it did not deter them from using these criteria for children as young as eight years old. Of the 117 children and adolescents screened for depression, 57 (49%) were included in the trial. Of these, 17 were diagnosed with MDD, 31 with minor depression, and nine with dysthymic disorder (DD). Twenty nine participants were allocated to the CBT group, and 28 to the control group. Of the 60 (51%) excluded, 12% were because they refused to attend regularly, requested another form of treatment such as family therapy, or dropped out after the initial assessment.

The aim was to offer each participant in the individual CBT group nine fortnightly sessions. The researchers did not specify the length of each session. Although they say that they controlled for therapists’ time (same frequency and number of sessions), it would appear that the participants in the control group were offered seven sessions and again the length of each session was not specified. The development of the CBT intervention was based on “research findings from studies on cognitive abnormalities in child and adolescent depression” (p. 108) (the authors do not specify which studies) and on similar treatment programmes designed by Stark (1990) for “non-clinical samples of youngsters” (p. 108) in a school context.

Three therapists provided both the CBT and the control interventions. There is no discussion about the implications of this, in particular the fact that therapists would not have been blind to the intervention that they offered and what they might have carried over from one group to the other. The researchers also reported that there were “no between-therapist differences” (p. 110). However, it is impossible to verify this claim given the paucity of demographic information about the therapists. Participants were included in the trial if they attended at least two CBT treatment sessions. The paper did not mention Intention to treat analysis.

Importantly, the researchers reported that following the treatment there was *no* significant treatment effect on outcome overall. At the end of the treatment period, 25 CBT participants (87%) and 21 NFI participants (75%) no longer presented with a diagnosable depressive disorder. Moreover, at 9-month follow-up CBT and NFI did not differ significantly (chi-square test) on any of the outcome variables measured. When the researchers compared the two groups, they found that their pattern of psychosocial functioning was very similar. They therefore concluded that “no significant treatment effect was established” (p. 105).

Brent et al. (1997)

The NICE guideline states that this controlled trial indicated a “clinically important improvement” for individual CBT compared with non-directive supportive therapy (NST) and with systemic behaviour family therapy (NICE, 2005a, p.95). Participants between the ages of 13 and 18 years were recruited from a psychiatric clinic in Pittsburgh, USA. Those who met the DSM-III-R (APA, 1987) criteria for MDD, with an intake score of 13 or more on the Beck depression Inventory (BDI), were included in the study. Potential participants diagnosed with co-morbid psychiatric conditions (e.g., psychosis, bipolar disorder, obsessive-compulsive disorder, eating disorder etc.) were excluded. Of the 122 adolescents who were eligible, 107 agreed to randomization, and 78 completed the trial. Eight participants dropped out of the study. Of note is that they were found to be more hopeless than those who continued in treatment.

The individual CBT treatment was adapted from an adult model to take account of “the specific developmental needs of adolescents” (p.878). Although the treatment consisted of an ‘active’ phase (12 to 16 weekly sessions), and a ‘booster’ phase (2 to 4 monthly sessions), the researchers decided to report the results of the active phase only. Why they made this decision is not specified, but it is important to note that there were no significant differences in the clinical outcome variables (including depressive symptoms) across the three study arms at a two-year follow-up (Birmaher et al., 2000).

At the end of the treatment phase, the School Age Schedule for Affective Disorders and Schizophrenia, Present and Lifetime Versions (K-SADS-P/E) was used to determine the proportion of participants meeting the DSM-III-R (APA, 1987) criteria for MDD. Clinical remission was defined as “no longer fulfilling the criteria for MDD and a BDI<9 for at least 3 consecutive sessions and sustained until the

end of treatment” (p.879). It is unclear why this definition was used and what impact it had on the results that the researchers chose to report.

The researchers concluded that individual CBT emerged as more efficacious than SBFT or NST for the treatment of adolescent depression, including with respect to effect on clinical recovery and the rate of symptomatic improvement. They also stated that the findings of their study were consistent with those of Wood et al. (1996), who found “CBT superior to an alternative psychosocial intervention” (p.883). As discussed above, Wood et al. (1996) labelled their relaxation training program as a no treatment control condition, not as an active treatment. What the two studies have in common is the stated finding in both trials that individual CBT has a short-term effect on adolescents’ depressive symptoms. The researchers claim that their study showed that individual CBT is “more efficacious than SBFT or NST for adolescent MDD in clinical settings, resulting in more rapid and *complete* treatment response” (italics added, p. 877). Given that they did not report the results following the booster sessions or take any other follow-up measurements, this claim is clearly overstating the data. In the Wood et al. (1996) study this treatment effect dissipated at a 3-month follow-up. This study provides no evidence to the contrary.

Rosselló & Bernal (1999)

This study evaluated the efficacy of cognitive-behavioural therapy (CBT) and interpersonal psychotherapy (IPT). It is unusual for a psychological therapy trial insofar as it sets out to compare two active interventions against a waitlist. The researchers, both based in Puerto Rico, introduce the study by drawing attention to the fact that most treatment-outcome research studies are not generalizable to ethnic minority populations. In order to respond to “the question of the efficacy of treatment with ethnic minorities” they decided to focus their attention on “the efficacy of two therapeutic approaches for the treatment of Puerto Rican adolescent depression” (p.735). To ensure ecological validity, the researchers considered factors such as cultural, developmental, and socio-economic factors in adapting a CBT and an IPT treatment approach for Puerto Rican adolescents. In particular, the adapted models took into account “cultural aspects of the treatments that consider the ‘interpersonal’ aspects of the Latino culture” (p.724). This was not the only significant adaptation. Both CBT and IPT interventions were adapted from adult therapy models. Specifically, the individual CBT intervention

for adolescents used in the trials was adapted from a group intervention for depressed adults and that had been used with a Hispanic population in the San Francisco area.

Members of the research team visited local schools to recruit participants and eventually received 161 referrals. Of these, a total of 71 were included in the study and 90 excluded. The authors state that the reasons for exclusion varied from meeting the exclusion criteria (30%), not interested in participating (21%), not meeting the inclusion criteria (18%), not showing up for the initial appointment (14%), and moving out of Puerto Rico (3%). It is unclear what the reason(s) were for the other 14%. The participants were aged between 13 and 17 years. In the CBT intervention group, nine adolescents met the DSM-III-R (APA, 1987) criteria for the diagnosis for a MDD.

Treatment comprised 12 one-hour individual therapy sessions once a week for 12 weeks. In delivering the interventions, parents were interviewed before and after the therapy and “if needed, therapists were allowed to discuss issues related to treatment with the parents individually or together with the adolescent” because “Puerto Rican adolescents depend on their parents for solutions, alternatives, and even attendance of therapy sessions” (p.737). Of the 25 adolescents assigned to the CBT group, four dropped out during the 12 week treatment period. However, at the 3-month follow-up point only 14 of the original 25 (56%) remained in the study. Moreover, in light of ethical considerations the wait-list group was not available at follow-up as they were eligible to receive therapy after the post-evaluation. To summarise: 161 adolescents were originally referred, 71 were included, and at follow-up 25 (CBT = 14; IPT = 11) remained in the study (16% of those originally referred; 35% of those included in the study).

Only self-report measures were used to determine the outcome of the treatment. At 3-month follow-up, no difference was found between the two treatments, and the researchers concluded that “both IPT and CBT are efficacious treatments for depressed Puerto Rican adolescents” (p.734). Nevertheless, the researchers also go on to say that “*although not statistically significant*, the CBT group appeared to continue to make gains in reduced symptoms of depression at follow-up” (italics added, p.739) and that the results show “a significant advantage of treatment over no treatment” (p.742). With no statistical significance and no data available for the wait-list group at 3-month follow-up, these statements need to be taken as the subjective views of the researchers, rather than any objective, empirical evidence.

Quality and interpretation of the evidence across the four studies

In this section the quality and interpretation of the evidence for individual CBT across the four studies will be reviewed. By definition, evidence-based practice emphasises interventions with strong support (APA, 2008). If we take account of factors such as strength of the evidence, sample sizes, participant characteristics, inclusion and exclusion criteria, treatment approaches and conditions, outcome measures and ecological validity, what conclusions can we draw from these four trials about the support for individual CBT for children and young people who have been diagnosed with depression?

Limited and weak evidence overall

In relation to the specific guideline under review here, NICE concluded that the overall evidence for the effectiveness of individual CBT was “inconclusive” (p.95). In a subsequent review of the guideline, Murray and Cartwright-Hatton (2006), writing from a CBT perspective, point to the guideline’s “weak evidence base” (p.135) and concludes that “the evidence base that underpins the guideline is very sparse” (p.129). Although they go on to describe the paucity of research in this area as “shocking” (p.135), they nevertheless conclude that “...the evidence for CBT is stronger than for other therapies...” (p.133). But what does this actually mean and what value can be placed on it in the context of evidence-based practice?

The Rosselló and Bernal (1999) study reported that both individual CBT and IPT were equally efficacious at a three month follow-up, and that there were no statistically significant reduced symptoms of depression at this time for those participants treated with individual CBT. However, there was no control group at this stage of the study, so it is impossible to say if any specific treatment effect for individual CBT (or IPT, for that matter) actually persisted at follow-up. The Vostanis et al. (1996) study was unable to indicate a statistically significant treatment effect between CBT and a non-focussed intervention at 9-month follow-up. In fact, the results indicated that the pattern of psychosocial functioning of the two groups of participants was very similar at this time. In the Wood et al. (1996) study individual CBT was not compared to another active treatment, but rather to a no treatment control condition (a ‘placebo’, if you like). Like the Rosselló and Bernal (1999) study, this trial showed that no statistically significant treatment effects were found at 6-month follow-up. In effect, the only conclusion that can be drawn from this study is that at 6-month follow-up individual CBT was no more effective than no treatment for adolescents who have been diagnosed with

depression. Although the Brent et al. (1997) study found some evidence to indicate that individual CBT might be superior to either SBFT or NST, the authors overstate this finding altogether. The study lacks precision or power as a result of a small sample size. Interestingly, there is no acknowledgement of this limitation in the paper. Although there is a statistically reliable difference in the remission percentage between the study arms, the differences have very wide confidence intervals (for CBT vs SBFT 8%-54%, and CBT vs NST 0% - 47%). This suggests that although differences may exist, the differences could be so small that they are not of any clinical significance. This is also the case for the BDI scores reported, where clearly the range is so wide as to be meaningless.

To get around the problem of the lack of statistical significance of the results in the four RCTs reported, the authors of the NICE guideline choose to talk of “clinically important improvement” (NICE, 2005a, p.95). What does this concept mean? What constitutes a clinically important improvement? How is it established? If we do not use statistical significance as our barometer, how would we know that a finding indicating a clinically important improvement is not simply due to chance? Differences reported as clinically *important* are so small in magnitude that they typically lack practical relevance and are unlikely to be clinically *significant*.

A final caution relates to the overall paucity of data at follow-up generated by these trials. The studies with follow-up data did not report any significant difference between individual CBT and another intervention / control group. If anything, the data supports evidence that no single model of psychological therapy is clearly superior to all others (e.g., Luborsky et al., 2002; Messer & Wampold, 2002; Rosenzweig, 1936; Wampold et al., 1997).

No evidence for younger children

In this NICE guideline no evidence is presented that individual CBT is effective for younger children who have been diagnosed with depression. In fact, the four RCT studies on which the guideline is based did not include a single child under the age of eight years, with the majority of participants at least 13 years old. In the guideline it is acknowledged that there are important developmental differences between ‘adult depression’ and ‘depression in younger people’ and it is made clear that “our conclusions **should not be assumed to apply to younger children**, without further investigation” (NICE, 2005a, p.75). However, as we shall see later, this caution is omitted from their Quick Reference Guide (NICE, 2005b) and actively contradicted in the CYP-IAPT training. For example, the first phase of the CYP-IAPT project focuses specifically on training therapists to use CBT for 3-10 year olds.

Small sample sizes

The NICE guideline being reviewed here included only four studies where individual CBT was evaluated in a RCT. Although the sample sizes in all four studies were small and therefore suffered from a lack of power, this was acknowledged in only two of the studies. One example of the small numbers on which the evidence is based can be seen in the Rosselló and Bernal (1999) study where at three month follow-up only 14 participants were left in the CBT group, 11 in the IPT group, and none in the control group (wait-list). To be fair to these authors, they did acknowledge the small sample size and suggested that their findings may be considered “preliminary” (Rosselló & Bernal, 1999, p.742).

Diverse populations and ecological validity

Two relatively small studies were based in the UK (Vostanis et al., 1996; Wood et al., 1996), one was based in the USA (Brent et al., 1997) and one in Puerto Rico (Rosselló & Bernal, 1999). These diverse subjects raise questions about the ecological validity of any claim that the findings of these studies put together are generalisable to the general population. It would be like saying that a young person or adolescent from Manchester or Birmingham in the UK is the same as, or sufficiently similar to, one from Pittsburgh in the USA and one from San Juan in Porto Rico. Rosselló and Bernal (1999) make the point that treatment outcome research studies with western ethnic *majority* populations are not generalisable to ethnic *minority* populations. But surely the reverse is also true, that the findings of treatment outcome research studies such as theirs with an ethnic minority population are not necessarily generalisable to a wider population.

Varying inclusion and exclusion criteria

All four studies used the DSM-III-R (APA, 1987) to decide who could be included in their studies. However, what qualified as ‘depression’ and how it could be determined varied from study to study. Two studies used “major depressive disorder (MDD)” (Wood et al., 1996, p.737; Brent et al., 1997, p.878), another “major depressive disorder, dysthymia, or both” (Rosselló & Bernal, 1999, p.735), and another simply “depression” (Vostanis et al., 1996, p.106). This variation seems to indicate that somewhat different versions of the construct of ‘depression’ were used across the four studies to diagnose potential participants. Also of note is that the criteria in this version of the DSM are predominantly based on validation studies with adults without taking a developmental perspective. Another important issue is the lack of consistency across the studies in terms of the instruments that were used to assess clinical depression at the pre-treatment stage. One study (Rosselló & Bernal, 1999) used the depression section of the Diagnostic Interview Schedule for Children (Shaffer et al., 1993),

another (Vostanis et al., 1996) the Mood and Feelings Questionnaire (MFQ) (Angold & Costello, 1987) and the child version of the Schedule for Affective Disorders and Schizophrenia (K-SADS) (Puig-Antich & Chambers, 1978), another (Wood et al., 1996) the MFQ only, and yet another the Beck Depression Inventory (BDI) (Beck, Steer & Garbin, 1988).

The four studies all used different exclusion criteria, thus increasing the inconsistency of selecting participants across the four studies. Moreover, and perhaps more importantly, all four trials excluded children and young people with other co-morbid problems. Rosselló and Bernal (1999) excluded participants for the following reasons: “serious imminent suicidal risk; psychotic features; bipolar disorders; alcoholism; conduct disorder, or drug use disorder; organic brain syndrome; marked hyperaggression; need for immediate treatment or hospitalization; currently receiving psychotropic medication or psychotherapy; and legal involvement” (p.735). A number of conditions resulted in exclusion from the Brent et al. (1997) study: “psychosis, bipolar I or II disorder, obsessive-compulsive disorder, eating disorder, substance abuse within the last 6 months, ongoing physical or sexual abuse, pregnancy, and chronic medical illness” (p.879). In addition, participants who made a suicide attempt and some who were still seriously symptomatic at midpoint were removed from their protocol. Wood, Harrington and Moore’s (1996) exclusion criteria were: “psychotic disorder, inpatients, taking or likely to require antidepressants, unable to complete the questionnaires, autism, attending a special school because of learning problems, major physical illness or epilepsy” (p.738). Vostanis et al. (1996) do not explicitly state their exclusion criteria, but a quick calculation shows that up to 51% of their original sample was excluded from the study for various reasons. Twelve per cent of potential participants were “excluded prior to randomisation on methodological, clinical or ethical grounds” (p.107). In the real world of CAMH services children and young people’s presentations are often complicated: they might have many different, interacting biological, psychological and social problems. The participants in these trials are therefore likely to be unrepresentative of the children and young people seen by mental health professionals in everyday clinical practice.

Inconsistent treatment approaches and conditions

When researchers perform the same experiment, with similar instruments, under similar conditions, and achieve the same results we regard these results as reliable. The four studies under review here all claimed to use individual CBT as the intervention for which efficacy was being evaluated. But a closer inspection reveals that the ‘CBT’ used varied from study to study. In the Rosselló and Bernal (1999) study the version of individual CBT that was used was adapted from a group intervention for depressed

Hispanic adults in the San Francisco area (Munoz et al., 1995). The authors state that their form of CBT included elements from cognitive behavioural therapy, cognitive therapy and rational-emotive therapy (Rosselló & Bernal, 1999). The Wood et al. (1996) version of CBT included cognitive and social problem solving components, as well as a component designed to help with the symptoms of depression (such as sleep hygiene measures). The authors said that their treatment programme was similar to a group-based treatment that has been used previously with non-clinical samples. In the Brent et al. (1997) study the researchers adapted the cognitive therapy approach of Beck et al. (1979) to use with the adolescents in their study. In the Vostanis et al. (1996) study the researchers developed yet another version of CBT which was also based on treatment programmes designed for non-clinical samples.

Each version of CBT also had its own unique protocol. For example, in one study the treatment consisted of 12 weekly sessions, in another it was 9 fortnightly sessions, in another it was 12 to 16 sessions in 12 to 16 weeks, plus 2 to 4 booster sessions in as many months, and in another it was 5 to 8 sessions in a period of approximately two months. Moreover, as Dubicka et al. (2010) noted “unlike the prescriptions of antidepressants in medication trials, administration of psychological treatment is difficult to homogenise, even with manualised therapy within a single study” (p.439). In these four studies, not only was there variance across the versions of CBT used, there was also variance across the studies in terms of *who* delivered the intervention on variables such as professional background, training and supervision received, gender, age, etcetera. In addition, the involvement of the family system (parents) varied significantly across the four studies, ranging from no apparent involvement, to psycho-education to pharmacological interventions for some parents.

Outcome measures

In the four studies a variety of instruments and protocols were used in an attempt to measure the effectiveness of individual CBT, but there were no consistency across them. The Beck Depression Inventory (BDI) (Beck, Steer & Garbin, 1988), the most commonly used outcome measure for assessing symptoms of depression, was apparently employed in only one study (Brent et al., 1997). Moreover, although it is common occurrence to use self-report questionnaires in these types of study, this type of measure suffers from a number of problems, including that it relies on the subjective impressions of the reporters. It is obviously not unimportant if someone reports a change in their mood, but it does not establish whether or not their symptoms have objectively changed. It also raises the prospect of ‘demand characteristics’ (Orne, 1962), that is, the possibility that the participants formed an

interpretation of the trial's purpose and unconsciously changed their behaviour to fit that interpretation.

In conclusion, the variation in sampling and methodology between these four studies, as well as the small sample sizes and inadequate number of RCTs, definitely limit the generalisability of the findings and any conclusions that can be drawn from them.

What about TADS?

One potential criticism of the above review is the exclusion of the *Treatment for Adolescents with Depression Study (TADS)* (2004). Ten years after it was conducted, this study is still widely regarded as the 'gold standard' trial for adolescents who have been diagnosed with depression. This RCT was carried out in the USA and involved a volunteer sample of adolescents between the ages of 12 to 17 years with a primary diagnosis of MDD. It comprised four comparison arms: fluoxetine alone, CBT alone, CBT with fluoxetine, and placebo. It was a large trial with a robust design, but its generalisability is nevertheless regarded as limited (Cox, Callahan, Churchill, Hunot, Merry, Parker and Hetrick, 2012). Although the intention is not to discuss the TADS study in detail here, a number of issues are worth noting in the context of this paper. Firstly, a significant number (56%) of participants were volunteers recruited through advertisements rather than from actual clinical settings. It raises the issue of how comparable the participants were with adolescents routinely seen in clinical practice. Secondly, the CBT was not blinded. It means that many participants would have been aware that they were receiving CBT. It can be argued that psychological treatment interventions are rarely, if ever, blinded in trials such as these. However, when CBT was compared to a blinded pill-placebo in this study, there was no effect at all (Effect size -0.03). Thirdly, there was a large dropout rate (28%) from the CBT group in the trial. It raises a question of how those who completed the CBT may have differed from those participants who dropped out. Fourthly, no child under the age of 12 was included in the trial. The trial therefore provides no evidence that individual CBT is effective for younger children who have been diagnosed with depression.

In terms of findings, when compared with placebo, the combination of fluoxetine with CBT was statistically significant ($P = .001$) after 12 weeks of intervention. However, no follow-up measures were conducted and the endurance of this effect is therefore unclear beyond the 12-week intervention stage. Also, the results of the combined treatment were not straightforward as they were significant on only one primary outcome measure (Children's Depression Rating Scale-Revised) and not the other (Clinical Global Impressions), when compared to fluoxetine alone. More importantly, however, is that

no statistical difference was found when CBT on its own was compared to placebo at the same 12-week interval. Nevertheless, the authors recommended CBT as a treatment of choice for adolescents who have been diagnosed with depression. It is also worth noting that after the publication of the NICE guideline there has been numerous secondary analyses and follow-up data published that has been included in subsequent meta-analyses (e.g., Cox et al., 2012; Dubicka et al., 2010), but the conclusions remain limited. Most notably, the comprehensive Cochrane review¹ recently carried out by Cox et al. (2012) concluded that

There is very limited evidence upon which to base conclusions about the relative effectiveness of psychological interventions, antidepressant medication and a combination of these interventions. On the basis of the available evidence, the effectiveness of these interventions for treating depressive disorders in children and adolescents cannot be established. Further appropriately powered RCTs are required (Cox et al., 2012, p.2).

Given that research findings since 2005 have not radically altered the evidence base used by the NICE guidance group, what are the implications for clinical practice, training and research? It goes without saying that more research (including replication studies) is urgently needed to establish a more credible evidence base, so the rest of this paper will focus on the implications for clinical practice and training. In particular, these two aspects will be reviewed in light of the recent CYP-IAPT project that is being implemented up and down the country to ‘transform’ all NHS Child and Adolescent Mental Health Services in the UK.

Implications for clinical practice

Recent years have seen an initiative to improve access to psychological therapies in the NHS. This initiative was led by an economist, Professor Lord Richard Layard (Layard, 2006), and became known as the Improving Access to Psychological Therapies (IAPT) programme. It was initially launched in adult mental health services with a stated aim of “rolling out services across England offering interventions approved by the National Institute of Health and Clinical Excellence (NICE) for treating people with depression and anxiety disorders” (IAPT, 2013). Since then Lord Layard lobbied strongly that “within 5

¹ This review included ten studies (one of which was the TADS (2004) RCT), with participants aged from 8 to 19 years (Cox et al., 2012).

years the NHS will offer evidence-based psychological therapy to all children and adults who need it” (Layard, 2010, p. 1). He claimed that conditions such as childhood depression “can be treated by psychological therapies” and that these treatments “have been tested in rigorous clinical trials” (Layard, 2010, p.3). He stated that CBT had a good success rate for children with depression, (Layard, 2010). Following powerful lobbying by Lord Layard, the IAPT programme was extended to children and young people. The key aim of the CYP-IAPT project is to “transform existing services for children and young people...by providing treatment which is based on best evidence” (CYP-IAPT, 2013, p.4).

The results of the meta-analysis above shows that there is little, if any empirical evidence that would support the use of individual CBT as a treatment of superior efficacy for young people who have been diagnosed with depression, and no empirical evidence whatsoever for children eight years and younger. Surprisingly, the weak evidence base for individual CBT is not explicitly mentioned in the Quick Reference Guide (NICE, 2005). This is important because, if anything, this shorter version of the guideline is more likely to be read by clinicians. Instead, individual CBT is recommended as a first-line treatment in the management of moderate to severe depression for children and young people. Importantly, there is no caution on the use of individual CBT for children of a younger age (as is the case with some other treatments in the same guideline). It seems that this lack of evidence is also largely ignored when it comes to the training of the therapists to work in the CYP-IAPT programme. References to the effectiveness of CBT (often referred to as ‘the best evidence’) are pervasive in the CYP-IAPT literature, including the curriculum for training mental health workers delivering treatments to children and young people who have been diagnosed with depression (IAPT, 2013).

CYP-IAPT training for depression

In its first year of training CYP-IAPT has supported training in two modalities, CBT and parenting. In choosing CBT as the only psychological therapy intervention in year one, those in charge of the training curriculum repeatedly state that this choice is based on ‘best evidence’ as indicated by the 2005 NICE guidelines (see CYP-IAPT website and training curriculum, 2012). From the present meta-analysis of the studies included in the NICE guideline, we know that the evidence base for individual CBT in this area is very weak and that to talk of ‘best evidence’ lacks any significance or substance.

In terms of training healthcare professionals who will be working with children and young people who have been diagnosed with depression, the training manual (IAPT, 2013) clearly states that the “course

will have a cognitive behavioural theoretical base with preference for approaches with the soundest evidence and where cognitive and behavioural techniques are integrated in therapy” (p.39). Its aim is to “provide a post-qualification training in evidence based CBT for children and young people with depression” (p.40). Although it acknowledges that, from an overall perspective, “there is not sufficient evidence to suggest that CBT is clearly superior to a number of other psychological treatments”, it nevertheless concludes that CYP-IAPT workers “will be trained in the NICE guidance (2005)” to “deliver CBT for children and young people” (p.40). It states that it will teach healthcare professionals to “adapt CBT interventions for younger children presenting with depression” (p.39) and that “IAPT workers will need to be able to adapt the approach to younger children” (p.40). The CYP-IAPT training manual concludes that successful completion of the course “should ensure that all trainees reach a level of competence that would enable them to obtain the outcomes reported in the relevant NICE Guideline for depression” (p.40).

Given that the ‘best evidence’ for individual CBT is so weak, it is perplexing that CYP-IAPT advocates so strongly for its use to treat children and young people who have been diagnosed with depression. Of particular concern, however, is their stated aim to train people to use individual CBT with younger children. It is perhaps not completely surprising given that they, like many others, probably rely on NICE guidance which, as indicated earlier, provides no caution against this approach in their Quick Reference Guide (NICE, 2005). This is nevertheless worrying, given the fact that none of the four RCTs included by NICE as ‘best evidence’ included a single child under the age of nine.

Conclusions

Although individual CBT is recommended as a first-line treatment for children and young people who have been diagnosed with depression (NICE, 2005b), the present meta-analysis of the best evidence available does not support this recommendation. Overall, the empirical evidence for the effectiveness of individual CBT for young people who have been diagnosed with depression is weak and inconclusive, and for children under the age of nine it is non-existent. Trainees and mental health professionals who are participating in the CYP-IAPT project now find themselves in the unusual position of having to learn and apply a psychological therapy approach for which very little empirical evidence exists. We need to challenge the idea that individual CBT for children and adolescents who have been diagnosed with depression should be prioritised in the NHS – using evidence is a good place to start. Our modest efforts

might serve as some protection against attempts to limit care to children and young people in England and Wales who have been diagnosed with depression.

Finally, this paper also draws attention to NICE's process for evaluating the evidence that is subsequently used for producing national guidance. It is beyond the scope of this paper to discuss this in detail, but a comprehensive critique of NICE's approach to setting guidelines for psychotherapy provision has been published by the UK Council for Psychotherapy (Guy, Thomas, Stephenson & Loewenthal, 2011). Nevertheless, given that this paper highlights the inadequate evaluation of evidence in one particular area, it raises the important question of what safeguards there are within the NICE process to ensure that evidence has been appropriately evaluated, and how these can be improved. This is becoming even more important since the recent judicial ruling in the UK which states that it is unlawful to deny treatment on the basis of a disagreement with NICE's evaluation of the evidence (Dyer, 2014).

Declaration of Conflicting Interests

None declared

Acknowledgements

I would like to thank Tamsyn Noble, Lizette Nolte, David Winter and the two reviewers for their helpful feedback and comments on earlier drafts of this paper.

References

- American Psychiatric Association. (1987) *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.
- American Psychological Association on Evidence-Based Practice with Children and Adolescents. (2008) *Disseminating evidence-based practice for children & adolescents: A systems approach to enhancing care*. Washington, DC: APA.
- American Psychological Association Presidential Task Force on Evidence-Based Practice. (2006) Evidence-based practice in psychology. *American Psychologist*, 61: 271-185.
- Angold, A. & Costello, E.J. (1987) *The moods and feelings questionnaire*. Unpublished document. Duke University Developmental Program.

- Angold, A., Costello, E.J., Pickles, A., & Winder, F. (1987) *The development of a questionnaire for use in epidemiological studies of depression in children and adolescents*. London: Medical Research Council Child Psychiatry Unit.
- Beck, A.T., Rush, A.J., Shaw, B.F., & Emery, G. (1979) *Cognitive therapy of depression*. New York: Guilford Press.
- Beck, A. T., Steer, R.A., & Garbin, M.G. (1988) Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8*: 77-100.
- Birmaher, B., Brent, D.A., Kolko, D., Baugher, M., Bridge, J., Holder, D., Iyengar, S., & Ulloa, R.E. (2000) Clinical outcome after short-term psychotherapy for adolescents with major depressive disorder. *Archives of General Psychiatry, 57*: 29-36.
- Brent, D.A., Holder, D., Kolko, D., Birmaher, B., Baugher, M., Roth, C., Iyengar, S., & Johnson, B.A. (1997) A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive therapy. *Archives of General Psychiatry, 54*: 877-885.
- Cox,, G.R., Callahan, P., Churchill, R., Hunot, V., Merry, S.N., Parker, A.G., & Hetrick, S.E. (2012) Psychological therapies versus antidepressant medication, alone and in combination for depression in children and adolescents. *The Cochrane Library, Issue 11*.
- Department of Health (1996) *Primary Care: Delivering the future?* London: HMSO.
- Dubicka, B., Elvins, R., Roberts, C., Chick, G., Wilkinson, P., & Goodyer, I.M. (2010) Combined treatment with cognitive-behavioural therapy in adolescent depression: meta-analysis. *The British Journal of Psychiatry, 197*: 433-440.
- Dyer, C. (2014) Disagreeing with NICE's evaluation of evidence is not grounds for a CCG to deny recommended treatment, judge rules. *BMJ, 348*: 2847.
- Fonagy, P. (2012) Welcome and introduction. *Children and Young People's Improving Access to Psychological Therapies Newsletter, June 2012*: p.1.
- Guy, A., Thomas, R., Stephenson, S., & Loewenthal, D. (2011) *NICE under scrutiny: the impact of the National Institute for Health and Clinical Excellence guidelines on the provision of psychotherapy in the UK*. UKCP Research Unit, Roehampton University, UK.
- Improving Access to Psychological Therapies (IAPT). (2013) Available at: <http://www.iapt.nhs.uk/> (Accessed: 1 August 2013).

- Layard, R. (2010) A campaign for psychological therapies: the case. Available at: http://www.mind.org.uk/assets/0000/7445/lord_layard_document_new.pdf (Accessed: 1 August 2013).
- Layard, R. (2006) The case for psychological treatment centres. *British Medical Journal*, 332: 1030-1032.
- Luborsky, L., Rosenthal, R., Diguier, L., Andrusyna, T.P., Berman, J.S., Levitt, J.T., Seligman, D.A., & Krause, E.D. (2002) The Dodo bird verdict is alive and well – mostly. *Clinical Psychology: Science and Practice*, 9: 2-12.
- Messer, S.B., & Wampold, B.E. (2002) Let's face facts: Common factors are more potent than specific therapy ingredients. *Clinical Psychology: Science and Practice*, 9: 21-25.
- Muñoz, R.F., Ying, Y.W., Bernal, G., Pérez-Stable, E.J., Sorensen, J.L., Hargreaves, W.A., Miranda, J., & Miller, L.S. (1995) Prevention of depression with primary care patients: A randomized controlled trial. *American Journal of Community Psychology*, 23: 199-213.
- Murray, J. & Cartwright-Hatton, S. (2006) NICE Guidelines on treatment of depression in childhood and adolescence: implications from a CBT perspective. *Behavioural and Cognitive Psychotherapy*, 34: 129-137.
- National Institute for Health and Clinical Excellence (NICE). (2005a) *Depression in Children and young people: identification and management in primary, community and secondary care. Clinical Practice Guideline No. 28*. London: British Psychological Society and Royal College of Psychiatrists.
- National Institute for Health and Clinical Excellence (NICE). (2005b) *Depression in Children and young people: identification and management in primary, community and secondary care. Quick reference guide*. London: British Psychological Society and Royal College of Psychiatrists.
- Orne, M.T. (1962) On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications. *American Psychologist*, 17: 776–783.
- Pugh, K. (2012) Year one and year two therapies. *Children and Young People's Improving Access to Psychological Therapies Newsletter, June 2012*: p.2.
- Puig-Antich J. & Chambers, W. (1978) *The schedule for affective disorders and schizophrenia for school-age children (Kiddie-SADS)*. New York: New York State Psychiatric Institute.

- Rosenzweig, S. (1936) Some implicit common factors in diverse methods of psychotherapy: "At last the Dodo bird said, 'Everybody has won and all must have prizes.'" *American Journal of Orthopsychiatry*, 6: 412-415.
- Rosselló, J. & Bernal, G. (1999) The efficacy of cognitive-behavioural and interpersonal treatments for depression in Puerto Rican adolescents. *Journal of Consulting and Clinical Psychology*, 67: 734-745.
- Shaffer, D., Gould, M.S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., & Aluwahlia, S. (1983) A children's global assessment scale (CGAS). *Archives of General Psychiatry*, 40, 1228-1231.
- Springsteen, B. (2012) We take care of our own. On *Wrecking Ball* [CD]. Columbia/Sony Music.
- Stark, K. (1990) *Childhood depression: School-based intervention*. New York: Guilford Press.
- Treatment for adolescents with depression study team (TADS). (2004) Fluoxetine, cognitive-behavioural therapy, and their combination for adolescents with depression: treatment for adolescents with depressions study (TADS) randomized controlled trial. *JAMA*, 292: 807-820.
- Vostanis, P., Feehan, C., Grattan, E., & Bickerton, W. (1996) A randomised controlled trial of cognitive-behavioural treatment for children and adolescents with depression: 9-month follow-up. *Journal of Affective Disorders*, 40: 105-116.
- Wampold, B.E., Mondin, G.W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997) A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes." *Psychological Bulletin*, 122: 203-225.
- Whiteford, H.A., Harris, M.G., McKeon, G., Baxter, A., Pennell, C., Barendregt, J.J., & Wang, J. (2013) Estimating remission from untreated major depression: a systematic review and meta-analysis. *Psychological Medicine*, 43: 1569-1585.
- Wood, A., Harrington, R., & Moore, A. (1996) Controlled trial of a brief cognitive-behavioural intervention in adolescent patients with depressive disorders. *Journal of Child Psychology and Psychiatry*, 37: 737-746.