Fifteen years after the first substantive trial,¹ cognitive–behavioural therapy (CBT) has become the first form of psychotherapy to achieve widespread acceptance in schizophrenia. In the UK, the National Institute for Health and Care Excellence (NICE) have now recommended it twice,²,³ the second time for all people with the disorder. The American Psychiatric Association approved it ‘with moderate clinical confidence’ in 2004 for patients with stable schizophrenia,⁴ and more recently the US Schizophrenia Patient Outcomes Research Team (PORT) have endorsed it for patients who have persistent psychotic symptoms while receiving adequate pharmacotherapy.⁵,⁶ Other guideline development groups around the world have followed suit.⁷

But is this therapeutic optimism justified? Most large, methodologically rigorous trials of CBT have failed to demonstrate a significant advantage at trial end on either symptomatic or relapse-related measures,⁸–¹² with only two finding clear evidence of benefit on their primary outcomes.¹³,¹⁴ Instead, the judgement that CBT is effective rests on a series of meta-analyses, which have variously concluded that it is promising;¹⁵ that it produces higher rates of improvement in mental state;¹⁶ that it has beneficial effects on positive and negative symptoms, mood, functioning and social anxiety;¹⁷ and that it is effective in reducing readmissions to hospital, duration of admission and symptom severity.⁹

It is possible for meta-analyses to come to positive conclusions even when most of the individual studies included had negative findings (for example Lau et al²⁰). Nevertheless, basing judgements on meta-analyses requires that their findings should be reliable and valid, and there are clear indications that this may not always be the case. Thus, different meta-analyses of the same studies have had opposite findings, as has been found with mammography for breast cancer;²¹ or similar findings can be interpreted differently, as Steiner²² has argued is at the heart of the controversy over the effectiveness of selective serotonin reuptake inhibitors in depression.

A further problem is that meta-analyses have sometimes been found not to agree with the findings of subsequent ‘mega trials’ of the same treatment.²³,²⁴ One reason why meta-analyses can reach unreliable conclusions is failure to take into account study quality. In a review of this issue, Juni et al²⁵ concluded that there was ample evidence that the deficiencies of methodologically weak trials translated into biased findings in systematic reviews, and argued that the influence of study quality should be routinely assessed. Only one of the meta-analyses of CBT for schizophrenia has formally done this: Wykes et al²⁶ found that a quality score combining information from different aspects of the design and reporting of trials did not significantly moderate effect size in their main meta-analysis of end-of-study positive symptom scores. However, the use of quality scales is now no longer recommended, since these often rate aspects of a study that bear little relationship to known sources of bias, and also because different quality scales have been found to give different results.²⁷–²⁹ A range of aspects of study quality were assessed separately by NICE,² but low-quality studies continued to be included in all meta-analyses and potential moderating effects were not examined.

A further aspect of study quality relevant to trials of psychological treatment is whether the therapy is compared with treatment as usual (TAU) or to a control intervention. The position usually taken here is that evidence-based psychotherapies need to demonstrate benefits over and above what can be attributed to the so-called shared or non-specific effects of
psychological intervention, although an alternative view exists which considers that applying the logic of placebo-controlled trials to psychotherapy research is flawed. Dental, for example, has argued that whereas psychological factors such as warmth, kindness and the instilling of hope are unwanted complications that need to be removed from trials of medical treatments, they are intrinsic elements of all forms of psychotherapy without which nothing can be expected to happen. Some meta-analyses of CBT for schizophrenia carried out separate comparisons of CBT v. TAU and CBT v. other psychological interventions, however, these did not statistically compare the two sets of effect sizes. The potential importance of this issue has recently been highlighted by a Cochrane review that found no advantage for CBT compared with other psychosocial treatments – including those that were categorised as either active or inactive – on a range of measures including relapse, readmission to hospital and various measures of mental state and social functioning.

A final problem facing meta-analysis is publication bias, the fact that trials with positive findings are more likely to be published than those with negative findings. Publication bias is typically examined by means of funnel plots, which may show an absence of small studies with negative findings, and this can be supplemented by one or more statistical tests for asymmetry. To date, the only meta-analysis of CBT for schizophrenia to examine publication bias has been that of Wykes et al.; they found that a funnel plot of studies in their meta-analysis of positive symptoms was reasonably symmetrical, but they did not assess this further with statistical testing.

Five and four years, respectively, have passed since the two most recent comprehensive meta-analyses of CBT for schizophrenia by Wykes et al. and NICE. During this time a considerable number of further studies have been published, including two with samples of approximately 100 patients in each group. We therefore conducted an updated meta-analysis of CBT, specifically with respect to its effect on core schizophrenic symptoms. We used this data-set to examine the influence of three well-recognised sources of bias on effect size: randomisation, masking and completeness of outcome data. We also evaluated use of a control intervention as a potential moderating factor. Finally, we tested whether publication bias might be affecting the findings.

### Method

#### Identification and selection of studies

The review was conducted in accordance with PRISMA guidelines. Trials of CBT for schizophrenia were searched for using MEDLINE (1993 to Week 3, March 2013), PsycINFO (1993 until Week 4, March 2013), Embase (1993 until Week 4, March 2013) and the Cochrane central register of controlled trials (1993 until end of March 2013); 1993 being the year of the first published trial of CBT in schizophrenia. We used the following MeSH headings/keywords: (a) “cognitive therapy” OR “cognitive behavioural therapy” OR “CBT”; (b) “schizophrenia” OR “schizoaffective disorder” OR “psychosis” OR “non affective psychosis” OR “schizo”; (c) “randomised controlled trial” OR “clinical trial” (cognitive behavioural therapy or cognitive therapy or CBT) and (schizophrenia or schizo” or psychosis) and (randomised controlled trial or clinical trial). Studies in any language were considered. The search was supplemented by hand searching of meta-analyses and review articles. The reference lists of all obtained studies were also checked. This part of the search was also used to check for trials that potentially could be included that were carried out prior to 1993. A search for completed but not yet published trials was also conducted using metaRegister at Current Controlled Trials (www.controlled-trials.com/mrct) and authors of any such trials were contacted for details about prospective publication dates.

We employed broad inclusion criteria similar to those used by Wykes et al. and NICE and the Cochrane Collaboration. Thus, studies were included if a majority of the patients had a diagnosis of schizophrenia, schizoaffective or non-affective functional psychosis, either made clinically or according to diagnostic criteria. Studies had to include a parallel control group, but this could be of any type, i.e. waitlist, TAU or an intervention designed to control for the non-specific effects of psychotherapy (see below). We only included randomised trials, specifically excluding those which the authors stated were non-randomised or which used inappropriate randomisation methods (e.g. allocation by alternation or by availability of the intervention).

Since the outcome measures were schizophrenic symptoms, we required a statement that the type of CBT used was directed to at least one class of symptoms. Such studies were included in all of the three main meta-analyses of overall symptoms, positive symptoms and negative symptoms unless they specified that the CBT was specialised for negative symptoms, in which case they were only included in that meta-analysis. A small number of studies used CBT directed specifically to auditory hallucinations and these were only included in a supplementary meta-analysis of these and other studies reporting hallucination scores. Studies that indicated that the type of CBT used was not directed to schizophrenic symptoms, but was instead adapted for self-esteem, obsessive–compulsive symptoms, post-traumatic stress symptoms, anxiety or suicide were not included in any of the analyses.

We included studies using both individual and group CBT. Given that CBT technique varied considerably across the studies, those that incorporated additional elements of therapy such as motivational interviewing, family engagement, behaviour therapy and social skills training were not excluded. Like other meta-analyses, however, we did not include studies that delivered CBT only as part of a prespecified, multicomponent package of care including several other interventions (sometimes referred to as integrated treatment or similar).

We included two studies that used acceptance and commitment therapy, since the authors considered this to be both related to CBT and directed to psychotic symptoms. On the same grounds we included one study where CBT took the form predominantly of coping skills enhancement. In recognition of the uncertainties about these therapies, however, we also calculated pooled effect sizes excluding these three studies.

#### Data extraction

Data were initially extracted by two of the authors working together, and were then independently re-extracted by another author, with differences being resolved. Effect sizes were calculated using Hedges’ g (i.e. the standardised difference between means, corrected for the tendency towards overestimation in small studies). When a study used two control groups the effect size for CBT was calculated against both of these combined. When data could not be extracted from information given in the article, sometimes it was available on the National Collaborating Centre for Mental Health website (www.nccmh.org.uk), the body which carries out meta-analyses on behalf of NICE. If the data still could not be found, authors were contacted. Data were pooled using Comprehensive Meta-analysis, version 2 for Windows (www.meta-analysis.com). The random effects option was used.
Overall symptoms

For the analysis of overall symptoms we included studies reporting total scores on general psychiatric scales that rated not just positive and negative symptoms but also other symptoms. Scales used included the Positive and Negative Syndrome Scale (PANSS), the Brief Psychiatric Rating Scale (BPRS), the Comprehensive Psychopathology Rating Scale (CPRS) and the Hopkins Psychiatric Rating Scale. Some studies separately reported PANSS positive, negative and general psychopathology subscale scores. In these cases we calculated the total score as the sum of the three subscale scores, taking into account the correlation coefficients between them as reported by Peralta & Cuesta in a sample of 100 patients with DSM-III-R schizophrenia (see online Data Supplement 1). We did not average scores from studies that only reported positive and negative symptom scores.

Positive symptoms

For the positive symptoms analysis we included studies that reported scores for delusions and hallucination subscales of published scales (i.e. the reality distortion syndrome) or for delusions, hallucinations and formal thought disorder subscales (i.e. the older, broader concept of positive symptoms). Scales used included positive symptom subscales of the PANSS, BPRS, the Krawiecka (Manchester) scale, the Schedule for the Assessment of Positive Symptoms (SAPS) and the Psychotic Symptom Rating Scales (PSYRATS). We did not include the change subscale of the CPRS, as this does not approximate very closely to positive psychotic symptoms. If a study reported separate measures of reality distortion and disorganisation, these were summed assuming a correlation of 0.40 between the two syndromes, as reported in a meta-analytic factor analysis of schizophrenic psychoses by Smith et al. Similarly, if a study provided separate delusion and hallucination subscale scores, these were averaged assuming a correlation of 0.34 from Smith et al.

Negative symptoms

Scales used by the studies in this analysis included negative symptom subscale of the PANSS, the SANS, the BPRS negative factor, a negative symptoms scale derived from the CPRS and from the Krawiecka (Manchester) scale. For one study that did not report global negative symptom scores we averaged scores from the four subscales of the SANS employed, using correlations reported in Smith et al. Another study used factor scores from the SANS, and these were also averaged using published data concerning the correlations between them.

Hallucinations

Most studies used the hallucinations scale from the PSYRATS; one study used a single item from the BPRS and another summed scores from four items from the CPRS. When studies reported individual PSYRATS hallucination subscale scores, these were averaged based on the correlations among them in a study of 276 patients with psychosis.

Examination of potential biasing factors in studies

We examined three sources of bias: randomisation, masking and incompleteness of outcome data. Bias from randomisation can be further divided into two distinct processes: (a) sequence generation, i.e. whether the method for allocating participants to interventions was based on some explicitly random process; and (b) allocation concealment, i.e. the demonstration that steps were taken to prevent the investigators gaining knowledge of forthcoming allocations. Studies were classified as being at low risk, at high risk or unclear using the Cochrane Risk of Bias Tool (see online Table DS1 for individual studies’ categorisations). Studies at high and low risk of bias were compared statistically if there were enough studies to do so. If not, low-risk studies were entered in a subanalysis of studies at low risk of bias from all three factors. Once again random effects models were used.

Randomisation (sequence generation)

We considered at low risk of bias studies that described use of random number tables, a random number generator, coin toss or drawing lots. Statements about block randomisation and/or stratification (within a centre), use of an independent statistician or independent service were also accepted, on the grounds that these strongly imply use of random numbers. Studies that merely stated that they used randomisation without further details were classified as unclear. Since we excluded a priori non-randomised trials, most studies that would have been classified as at high risk of bias were automatically removed from consideration. The only exception was an included study where a subset of the patients were assigned using inadequate randomisation.

Randomisation (allocation concealment)

We accepted as evidence of central allocation (one indicator of effective allocation concealment) any statement that indicated that randomisation was performed by an outside service or a person independent of the research team. If randomisation was carried out by a member of the research team, we required an explicit statement that he/she was independent or had no involvement in the baseline assessments. If studies only referred to use of envelopes, but did not state that they were sequentially numbered, opaque and sealed, they were categorised as unclear.

Masking

Since no studies of CBT have used double-blinding, only masking of the outcome assessment was examined. To be categorised as at low risk of bias, we required the study to state that the assessments were carried out by interviewers masked to treatment assignment. Studies that made no statement about masking were treated as at high risk of bias, on the grounds that it is unlikely that authors would fail to mention such a key methodological factor if they had employed it. Studies that referred to independent assessors without further elaboration were considered as unclear and authors were contacted. Four studies that indicated that the masking could have been compromised were also rated as unclear.

Incomplete outcome data

If no further details were given, we used a cut-off of >20% attrition in the whole sample as the threshold for considering a study to be at high risk of bias. Studies with attrition rates above this threshold were still considered as low risk if either (a) details of individuals who dropped out were given and were justifiable, or (b) if studies used intention-to-treat (ITT) analysis. Some studies used ITT, but data could only be extracted from tables that reported data for those who completed the study. In these cases, the study was categorised as being at low risk of bias if the
drop-out rate was $\geq 20\%$ (without reasons) and 'unclear' if the rate was $> 20\%$ (without reasons).

**Effect of the use of a control intervention**

The aim here was to examine the influence of the use of an intervention designed to control for the non-specific effects of psychotherapy. To this end we compared studies that employed control interventions that (a) were stated or implied to control for this (recreation and support, group support, befriending, supportive counselling/therapy, social activity therapy and goal-focused supportive contact); or (b) could be considered unlikely to have a specific effect on schizophrenic symptoms (psycho-education and cognitive remediaion therapy). We did not include the family therapy arm of one study (which the authors considered to be potentially therapeutic). Studies where we combined data from two control groups (i.e. control intervention and TAU)$^{69,72,73}$ were not included in this analysis.

**Publication bias**

This was examined using three statistical techniques, Duval & Tweedie’s$^3$ trim and fill, Begg & Mazumdar’s$^{25}$ rank correlation test and Egger’s$^{26}$ test of the intercept.

**Results**

The search produced 1246 articles. Titles and, where relevant, abstracts were checked by two of the authors, leading to 1169 being eliminated. The full text of the 77 remaining studies plus 13 more added from further searching were examined. Fifty studies reporting 52 studies were finally included (two studies reported on two separate samples)$^{1,8–14,33–43,52–54,62,63,69–73,77–97}$

A flow chart of the selection process is shown in Fig. 1. Individual effect sizes extracted from the included studies plus a list of the excluded studies are given in online Table DS1 and Data Supplement 2.

**Pooled effect sizes**

The pooled effect size for 34 studies of overall symptoms was $-0.33$ (95% CI $-0.47$ to $-0.19$, $P < 0.001$) (negative sign favours CBT). The studies were heterogeneous ($Q = 102.71$, $P < 0.001$), with an $I^2$ value of 67.9 (95% CI 54.2–77.5), indicating that two-thirds of the variation among studies was as a result of heterogeneity rather than chance. The pooled effect size for 33 studies of positive symptoms was $-0.25$ (95% CI $-0.37$ to $-0.13$, $P < 0.001$). Once again the studies were heterogeneous ($Q = 63.12$, $P = 0.001$; $I^2 = 49.3$, 95% CI 24.1–66.1). The pooled effect size for 34 studies of negative symptoms was $-0.13$ (95% CI $-0.25$ to $-0.01$, $P = 0.03$). These studies were also heterogeneous ($Q = 63.11$, $P = 0.001$; $I^2 = 47.7$, 95% CI 21.9–65.0).

Forest plots of the studies in the meta-analysis of overall symptoms and positive symptoms are shown in Figs 2 and 3 (see online Figs DS1 and DS2 for those for negative symptoms and hallucinations).

Recalculating the pooled effect sizes excluding studies using coping strategy enhancement or acceptance and commitment therapy made little difference to the findings (overall symptoms: effect size $-0.33$ (95% CI $-0.48$ to $-0.19$, $P < 0.001$, 32 studies); positive symptoms: effect size $-0.24$ (95% CI $-0.36$ to $-0.11$, $P < 0.001$, 30 studies); negative symptoms: $-0.14$ (95% CI $-0.26$ to $-0.01$, $P = 0.04$, 31 studies).

There were 15 studies in the supplementary meta-analysis of hallucinations. The pooled effect size was $-0.34$ (95% CI $-0.61$ to $-0.06$, $P = 0.01$). These studies were heterogeneous ($Q = 46.02$, $P < 0.001$) with $I^2 = 69.6$ (95% CI 48.3–82.1).

**Examination of bias within studies**

The findings with respect to sequence generation, allocation concealment, masking and completeness of outcome data are summarised in Table 1. It can be seen that masking significantly moderated effect size for overall symptoms ($-0.62$ in 10 non-masked studies $v.$ $-0.15$ in 20 masked studies, $P = 0.001$) and positive symptoms ($-0.57$ in 8 non-masked studies $v.$ $-0.08$ in 21 masked studies, $P < 0.001$). The effect for negative symptoms was not significant ($-0.22$ in 8 non-masked studies $v.$ $-0.04$ in 22 masked studies, $P = 0.26$).

Only a few studies were considered to be at high risk of bias with respect to sequence generation, allocation concealment and incompleteness of outcome data (1–2 studies across all analyses) and so statistical analysis was not carried out. Instead, studies at low risk of bias from all three factors (i.e. high-risk studies plus those categorised as ‘unclear’) were entered into a subanalysis including only these studies. Pooled effect sizes were as follows: overall symptoms $-0.15$ (95% CI $-0.32$ to 0.01, $P = 0.07$, 8 studies); positive symptoms $-0.10$ (95% CI $-0.28$ to 0.09, $P = 0.30$, 9 studies); negative symptoms $-0.02$ (95% CI $-0.15$ to 0.11, $P = 0.76$, 11 studies).

In the supplementary meta-analysis of hallucinations, there was a large difference between masked and non-masked studies (effect size $-0.18$ (95% CI $-0.37$ to 0.01) in 12 masked studies $v.$ $-0.91$ (95% CI $-2.67$ to 0.85) in 2 non-masked studies), but statistical significance was not tested owing to the small number of non-masked studies. No studies were rated as being at high risk of bias for sequence generation or allocation concealment and only one study for incompleteness of outcome data. The pooled effect size for 10 studies at low risk of bias from all three variables was $-0.20$ (95% CI $-0.44$ to 0.04, $P = 0.10$).
Jauhar et al.

**Fig. 3** Forest plot of studies in the meta-analysis of positive symptoms.

CBT, cognitive–behavioural therapy.
Effect of use of a control intervention

Effect sizes for studies that did and did not use a control intervention in the three main analyses are shown in Table 2. This factor did not significantly moderate effect size in any of the analyses.

In the meta-analysis of hallucinations, the pooled effect size was smaller in five studies using a control intervention than in eight studies that did not, but once again the difference was not significant (effect size = 0.15 (95% CI = 0.54 to 0.24) v. = 0.55 (95% CI = 1.04 to 0.06), Q(B) = 1.58, P = 0.21).

Examination of publication bias

Funnel plots of the studies in the three main analyses are shown in Fig. 4, and results of the statistical analyses are shown in Table 3. Trim and fill only imputed studies in the meta-analysis of positive symptoms (one study), reducing the effect size minimally from = 0.25 to = 0.24. Begg & Mazumdar’s test was at trend level, but Egger’s test was not significant. Begg & Mazumdar’s test was significant in the meta-analyses of overall symptoms and negative symptoms, at P = 0.009 and 0.02 respectively. Egger’s test was at trend level in the meta-analysis of overall symptoms (P = 0.06) and was not significant in the meta-analysis of negative symptoms (P = 0.30).

Discussion

Main findings

This meta-analysis, which employed broad inclusion criteria similar to those used by Wykes et al.19 NICE3 and the Cochrane Collaboration,32 found that pooled effect sizes were in the ‘small’ range for all the classes of symptoms considered. Other recent meta-analyses have struggled to demonstrate levels of effectiveness against symptoms higher than this. Thus, Zimmerman et al.17 found a pooled effect size of = 0.37 for positive symptoms at the end of treatment, but they only included 15 studies. Wykes et al.19 also found a pooled effect size of = 0.37 for positive symptoms in a larger set of 32 studies, plus an effect size of = 0.44 for negative symptoms in 23 studies. However, they used Glass’s method of calculating effect size, which divides the difference in means by the standard deviation of the control group alone rather than the combined standard deviation of both groups; it is known that this can inflate the estimate of effect size.28 Although NICE3 concluded that ‘CBT was shown to be effective in reducing symptom severity as measured by total scores on items such as the PANSS and BPRS . . . at end of treatment’, the effect sizes for total symptom scores were = 0.27 in 13 studies comparing CBT with standard care and = 0.13 (a non-significant value) in 6 studies comparing it with ‘other active treatments’. The corresponding effect sizes for positive symptoms were = 0.17 (eight studies) and = 0.13 (six studies).

The influence of use of a control intervention

Effect sizes for studies that did and did not use a control intervention in the three main analyses are shown in Table 2. This factor did not significantly moderate effect size in any of the analyses.

In the meta-analysis of hallucinations, the pooled effect size was smaller in five studies using a control intervention than in eight studies that did not, but once again the difference was not significant (effect size = 0.15 (95% CI = 0.54 to 0.24) v. = 0.55 (95% CI = 1.04 to 0.06), Q(B) = 1.58, P = 0.21).

Table 2 Comparison between studies not using and using a control intervention

<table>
<thead>
<tr>
<th></th>
<th>CBT v. TAU</th>
<th>CBT v. control intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Effect size (95% CI)</td>
<td>Studies, n</td>
</tr>
<tr>
<td>Overall symptoms</td>
<td>= 0.33 (0.45 to 2.1)</td>
<td>21</td>
</tr>
<tr>
<td>Positive symptoms</td>
<td>= 0.31 (0.45 to 0.17)</td>
<td>19</td>
</tr>
<tr>
<td>Negative symptoms</td>
<td>= 0.17 (0.33 to 0.02)</td>
<td>20</td>
</tr>
</tbody>
</table>

CBT, cognitive–behavioural therapy; TAU, treatment as usual.
larger differences – four to seven times across the three main meta-analyses – suggesting an exaggeration of the treatment effect that is at least as great as the 17–36% found in trials of medical treatments,99-104 In this respect it is worth noting that, in the forest plots shown in Figs 2 and 3, only two studies published since 2008 (the cut-off year in the NICE3 meta-analysis) have found a significant advantage for CBT against overall symptoms,94,96 and only one for positive symptoms.14 The assessments in the first two of these studies were presumptively made non-masked (the authors did not comment on masking), and in the third the masking was found to have become progressively more compromised as the trial went on.

Statistical examination of the effects of bias from inadequate randomisation and incompleteness of outcome data was not possible because of the small numbers of studies classified as being at high risk. Nevertheless, restriction of the sample to studies considered to be at low risk of bias from all three sources reduced the pooled effect sizes to non-significance for all classes of symptom. These pooled effect sizes may not be reliable, since the numbers of studies that survived this procedure were quite small (this was due to large numbers of studies being classified as ‘unclear’). However, such a finding does arguably place an onus on advocates of CBT for schizophrenia to demonstrate its effectiveness in at least one large trial free of the above methodological weaknesses.

Surprisingly, given the universal agreement on the importance of using a placebo in trials of medical treatments, we found that use of a control for the non-specific effects of psychological intervention did not moderate effect size in any of the analyses. One explanation of this finding could simply be that there were not enough studies using control interventions (9–12 in the three main analyses) to detect an effect. There is a possible hint that this may be the case, in that the pooled effect sizes for studies using a control intervention were found to be non-significant in all the three main analyses. But clearly, more studies will be needed to decide this issue.

Other explanations of our null finding here are also possible, notably that it supports the argument of authors like Kirsch30 and Bentall,31 that pill placebo and control for the non-specific effects of psychotherapy are not equatable. As described in the introduction, this position is based mainly on theoretical arguments and neither author attempted to tackle an important practical consideration, that of the Hawthorne effect.102 This is the well-established finding that people singled out for almost any kind of intervention tend to improve their performance or behaviour simply by virtue of the special attention they receive; it seems unlikely that this would not happen in psychotherapy trials. Apart from this, variations in the degree of ‘therapeuticness’ among different control interventions almost certainly needs to be considered. We were quite restrictive in our approach, including only control interventions that would not be expected to have specific effects on schizophrenic symptoms. In contrast, NICE3 compared CBT with ‘other active treatments’ in one of their two main sets of meta-analyses, including under this heading not only supportive counselling, befriending and the like, but also family therapy. Clearly, such an approach blurs the boundary between controlling for psychological confounding factors and examining whether CBT is more effective than other forms of psychotherapy.

Our findings with respect to publication bias did not suggest that this factor was exerting a significant influence on effect size. Some statistical evidence of bias was found in the meta-analysis of overall symptoms, and to a lesser extent in that of negative symptoms, but this could hardly be regarded as convergent. Our findings accord with those of Niemeyer et al,103 who imputed studies in ten data-sets selected from five published meta-analyses of CBT for schizophrenia and found that this resulted in no or relatively minor reductions in pooled effect sizes. However, it goes against findings in depression: Cuijpers et al104 found that an initial pooled effect size of 0.69 in 89 studies reduced to 0.49 after Duval & Tweedie’s trim and fill imputed 26 studies, and both Begg & Mazumbar’s and Egger’s tests were highly significant.

### Table 3 Results of tests for publication bias in the analyses of overall symptoms, positive symptoms and negative symptoms

<table>
<thead>
<tr>
<th>Effect size (95% CI)</th>
<th>Begg &amp; Mazumdar’s test*</th>
<th>Egger’s testa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall symptoms</td>
<td>2.37 0.009 1.56 0.06</td>
<td></td>
</tr>
<tr>
<td>Positive symptoms</td>
<td>1.49 0.07 0.15 0.44</td>
<td></td>
</tr>
<tr>
<td>Negative symptoms</td>
<td>2.12 0.02 0.54 0.30</td>
<td></td>
</tr>
</tbody>
</table>

* P-values are one-tailed, as recommended.

a. One study imputed.

---

**Fig. 4** Funnel plots of studies in the meta-analyses of overall symptoms (a), positive symptoms (b) and negative symptoms (c).

Studies with positive findings are to the left; white circles are imputed studies.
Implications

Should CBT for schizophrenia continue to be recommended in clinical practice? Given that we, and others including NICE,9 have found evidence for only small effects on overall symptoms, plus the fact that a large, methodologically rigorous 2008 trial failed to demonstrate any effectiveness against relapse,10 the UK government’s continued vigorous advocacy of this form of treatment (for example see The All Party Parliamentary Group on Mental Health11) might be considered puzzling. Our finding of non-significant effects on positive symptoms in a relatively large set of 21 masked studies also suggests that claims that CBT is effective against these symptoms of the disorder are no longer tenable. The same appears to apply to negative symptoms, although here the possibility that specially adapted forms of therapy will have an effect cannot be excluded (there have been only two such studies to date). We did not examine the effect of CBT on depression, anxiety or distress as a result of psychotic symptoms, so no judgements on its effects in these areas can be made.

References


52 Bradshaw W. Integrating cognitive-behavioral psychotherapy for persons with schizophrenia into a psychiatric rehabilitation program: results of a three year trial. Community Ment Health J 2000, 36: 491–500.


Cognitive–behavioural therapy for the symptoms of schizophrenia


86 Startup M, Jackson MC, Bendix S. North Wales randomized controlled trial of cognitive behaviour therapy for acute schizophrenia spectrum disorders: outcomes at 6 and 12 months. Psychol Med 2004; 34: 413–22.


Data supplement 1

**Formula for calculating PANSS total scores, from PANSS positive, negative and general psychopathology subscale scores**

The total score as the sum of the three subscale scores, and its standard deviation, were calculated from:

\[
\sigma_T = \sqrt{\sigma_P^2 + \sigma_N^2 + \sigma_G^2 + 2\sigma_P \sigma_N \rho_{PN} + 2\sigma_P \sigma_G \rho_{PG} + 2\sigma_N \sigma_G \rho_{NG}}
\]

where \(\sigma_T\), \(\sigma_P\), \(\sigma_N\) and \(\sigma_G\) are the standard deviations for the total, positive, negative and general psychopathology scores respectively, and \(\rho_{PN}\), \(\rho_{PG}\) and \(\rho_{NG}\) are the correlation coefficients between the subscale scores as reported by Peralta (Peralta, 1994 #150) in a sample of 100 patients with DSM-III-R schizophrenia.
<table>
<thead>
<tr>
<th></th>
<th>CBT N</th>
<th>Control N</th>
<th>Sequence generation</th>
<th>Allocation concealment</th>
<th>Blinding</th>
<th>Completeness of outcome data</th>
<th>Use of control intervention</th>
<th>Effect sizes (CI)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drury et al (1996)</td>
<td>20</td>
<td>20</td>
<td>Low risk</td>
<td>Unclear</td>
<td>High risk</td>
<td>Low risk</td>
<td>Recreation and support</td>
<td>Positive -0.94 (-1.58/-0.30) Negative -0.18 (-0.79/-0.42)</td>
<td>-</td>
</tr>
<tr>
<td>Kuipers et al (1997)</td>
<td>23</td>
<td>24</td>
<td>Low risk</td>
<td>Unclear</td>
<td>High risk</td>
<td>Low risk</td>
<td>No</td>
<td>Overall -0.34 (-0.91/+0.22)</td>
<td>-</td>
</tr>
<tr>
<td>Daniels et al (1998)</td>
<td>20</td>
<td>20</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Unclear</td>
<td>No</td>
<td>Negative -0.67 (-1.30/-0.05)</td>
<td>-</td>
</tr>
<tr>
<td>Kuipers et al (1997)</td>
<td>1</td>
<td>23</td>
<td>Low risk</td>
<td>Unclear</td>
<td>High risk</td>
<td>Low risk</td>
<td>No</td>
<td>Overall -3.97 (-5.87/-2.07) Positive -1.60 (-2.83/-0.38) Negative -2.51 (-3.96/-1.06)</td>
<td>Total symptom score ES combined from PANSS subscale scores</td>
</tr>
<tr>
<td>Levine et al (1998)</td>
<td>77</td>
<td>20</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Group support</td>
<td>Overall +0.57 (-0.34/+1.47)</td>
<td>-</td>
</tr>
<tr>
<td>Daniels et al (1998)</td>
<td>8</td>
<td>10</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Supportive counselling/ psychoeducation</td>
<td>Overall -0.72 (-1.37/-0.07) Positive -0.79 (-1.45/+0.14) Negative -0.33 (-0.97/+0.30)</td>
<td>-</td>
</tr>
<tr>
<td>Tarrier et al (1999)</td>
<td>23</td>
<td>47</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Two control groups used</td>
<td>Overall -0.43 (-0.92/+0.07) Negative -0.14 (-0.63/+0.35)</td>
<td>-</td>
</tr>
<tr>
<td>Haddock et al (1999)</td>
<td>8</td>
<td>10</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Supportive counselling/ psychoeducation</td>
<td>Overall -1.45 (-2.54/-0.37)</td>
<td>-</td>
</tr>
<tr>
<td>Pinto et al (1999)</td>
<td>19</td>
<td>18</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High risk</td>
<td>Low risk</td>
<td>Supportive therapy</td>
<td>Overall -0.09 (-0.48/+0.31) Positive -0.08 (-0.48/+0.31) Negative +0.01 (-0.38/+0.40)</td>
<td>Total symptom score ES combined from PANSS subscale scores</td>
</tr>
<tr>
<td>Turkington et al (2000)</td>
<td>10</td>
<td>5</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Befriending</td>
<td>Overall -0.08 (-0.49/+0.33) Negative +0.07 (-0.35/+0.48)</td>
<td>-</td>
</tr>
<tr>
<td>Turkington et al (2000)</td>
<td>10</td>
<td>5</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Befriending</td>
<td>Overall -1.14 (-2.22/-0.05)</td>
<td>-</td>
</tr>
<tr>
<td>Granholm et al (2002)</td>
<td>8</td>
<td>7</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High risk</td>
<td>Unclear</td>
<td>No</td>
<td>Overall -0.74 (-1.73/+0.25) Positive -0.76 (-1.75/+0.23) Negative -0.03 (-0.98/+0.93)</td>
<td>Not a subgroup of Granholm 2005</td>
</tr>
<tr>
<td>Lewis et al (2002)</td>
<td>78*</td>
<td>131*</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Two control groups used</td>
<td>Overall -0.01 (-0.29/+0.26) Positive -0.01 (-0.29/+0.27)</td>
<td>-</td>
</tr>
</tbody>
</table>

Table DS1 Summary table of study characteristics and effect sizes
(Note: a version of this table giving more detail/justification about the bias classifications is available at [www.cbtinschizophrenia.com/](http://www.cbtinschizophrenia.com/))
<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Size</th>
<th>Dropout</th>
<th>Risk</th>
<th>Blinding</th>
<th>Risk</th>
<th>Outcome</th>
<th>Weight</th>
<th>Overall</th>
<th>Positive</th>
<th>Negative</th>
<th>Risk</th>
<th>Hallucinations</th>
<th>Risk</th>
<th>Overall</th>
<th>Positive</th>
<th>Negative</th>
<th>Total symptom score ES combined from PANSS subscale scores</th>
<th>Total symptom score ES combined from PANSS subscale scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turkington et al (2002)</td>
<td>257</td>
<td>165</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Low risk</td>
<td>No</td>
<td>Overall</td>
<td>-0.22 (-0.42/-0.03)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Hallucinations -0.11 (-0.46/+0.25)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Durham et al (2003)</td>
<td>22</td>
<td>38</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Two control</td>
<td>Two control</td>
<td>Overall</td>
<td>Positive</td>
<td>Negative</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gumley et al (2003)</td>
<td>72</td>
<td>72</td>
<td>Unclear</td>
<td>Low risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>No</td>
<td>Overall</td>
<td>-0.42 (-0.75/-0.09)</td>
<td>Positive</td>
<td>-0.35 (-0.67/-0.02)</td>
<td>Negative</td>
<td>-0.35 (-0.69/-0.01)</td>
<td>Overall</td>
<td>+0.06 (-0.90/+1.01)</td>
<td>Positive</td>
<td>+0.11 (-0.84/+1.07)</td>
<td>Negative</td>
</tr>
<tr>
<td>Jolley et al (2003)</td>
<td>7</td>
<td>8</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Low risk</td>
<td>No</td>
<td>Overall</td>
<td>-0.70 (-1.32/-0.08)</td>
<td>Positive</td>
<td>-0.35 (-0.95/+0.25)</td>
<td>Negative</td>
<td>-0.49 (-1.10/+0.12)</td>
<td>Overall</td>
<td>-0.52 (-0.77/-0.27)</td>
<td>Positive</td>
<td>-0.69 (-0.95/-0.44)</td>
<td>Negative</td>
</tr>
<tr>
<td>Rector et al (2003)</td>
<td>24</td>
<td>18</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>No</td>
<td>Overall</td>
<td>-0.52 (-0.77/-0.27)</td>
<td>Positive</td>
<td>-0.69 (-0.95/-0.44)</td>
<td>Negative</td>
<td>-0.71 (-0.97/-0.46)</td>
<td>Overall</td>
<td>-0.60 (-1.09/-0.11)</td>
<td>Positive</td>
<td>-0.52 (-1.00/-0.03)</td>
<td>Negative</td>
</tr>
<tr>
<td>Wang et al (2003)</td>
<td>126</td>
<td>125</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High risk</td>
<td>Low risk</td>
<td>No</td>
<td>Overall</td>
<td>-0.52 (-0.77/-0.27)</td>
<td>Positive</td>
<td>-0.69 (-0.95/-0.44)</td>
<td>Negative</td>
<td>-0.71 (-0.97/-0.46)</td>
<td>Overall</td>
<td>-0.60 (-1.09/-0.11)</td>
<td>Positive</td>
<td>-0.52 (-1.00/-0.03)</td>
<td>Negative</td>
</tr>
<tr>
<td>Bechdolf et al (2004)</td>
<td>40</td>
<td>48</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Psychoeducation</td>
<td>Psychoeducation</td>
<td>Overall</td>
<td>Positive</td>
<td>-0.52 (-0.44/+0.39)</td>
<td>Negative</td>
<td>+0.16 (-0.25/+0.58)</td>
<td>Overall</td>
<td>+0.29 (-0.13/+0.71)</td>
<td>Positive</td>
<td>-0.02 (-0.44/+0.39)</td>
<td>Negative</td>
</tr>
<tr>
<td>Startup et al (2004)</td>
<td>34</td>
<td>32</td>
<td>Low risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>No</td>
<td>Overall</td>
<td>-0.60 (-1.09/-0.11)</td>
<td>Positive</td>
<td>-0.52 (-1.00/-0.03)</td>
<td>Negative</td>
<td>-0.32 (-0.80/+0.16)</td>
<td>Overall</td>
<td>+0.29 (-0.13/+0.71)</td>
<td>Positive</td>
<td>-0.02 (-0.44/+0.39)</td>
<td>Negative</td>
</tr>
<tr>
<td>Trower et al (2004)</td>
<td>15</td>
<td>17</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>No</td>
<td>Hallucinations -0.75 (-1.45/-0.05)</td>
<td>Hallucinations ES combined from PSYRATS subscale scores</td>
<td>Hallucinations ES combined from PSYRATS subscale scores</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cather et al</td>
<td>15</td>
<td>13</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Psychoeducation</td>
<td>Psychoeducation</td>
<td>Positive</td>
<td>-0.05 (-0.77/+0.67)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Study</td>
<td>N</td>
<td>N</td>
<td>Risk Level</td>
<td>Intervention</td>
<td>ES</td>
<td>HPC</td>
<td>Low Risk</td>
<td>Overall</td>
<td>Supportive Counseling</td>
<td>Positive</td>
<td>Negative</td>
<td>ES Combined</td>
<td>Total Symptom Score ES Combined from PANSS Subscale Scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----</td>
<td>-----</td>
<td>------------</td>
<td>--------------</td>
<td>-----</td>
<td>--------</td>
<td>----------</td>
<td>---------</td>
<td>-----------------------</td>
<td>----------</td>
<td>----------</td>
<td>-------------</td>
<td>----------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Granholm et al (2005)</td>
<td>32/37**</td>
<td>33/39**</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Low risk</td>
<td>No</td>
<td>Overall -0.05 (-0.53/0.43)</td>
<td>Positive +0.35 (-0.10/0.80)</td>
<td>Negative +0.15 (-0.30/0.59)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valmaggia et al (2005)</td>
<td>35</td>
<td>23</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Supportive counselling</td>
<td>Overall +0.08 (-0.44/0.60)</td>
<td>Positive -0.30 (-0.83/0.22)</td>
<td>Negative +0.36 (-0.16/0.89)</td>
<td>Hallucinations -0.74 (-1.28/0.21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wykes et al (2005)</td>
<td>37</td>
<td>36</td>
<td>Low risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>No</td>
<td>Hallucinations -0.03 (-0.48/0.43)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barrowclough et al (2006)</td>
<td>54</td>
<td>45</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>No</td>
<td>Overall -0.25 (-0.65/0.14)</td>
<td>Positive -0.03 (-0.43/0.36)</td>
<td>Negative -0.06 (-0.45/0.33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaudiano et al (2006)</td>
<td>21</td>
<td>19</td>
<td>Low risk</td>
<td>High risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>No</td>
<td>Overall -0.48 (-1.10/0.14)</td>
<td>Positive -0.96 (-1.61/-0.32)</td>
<td>Negative +0.09 (-0.52/0.69)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penadés et al (2006)</td>
<td>20</td>
<td>20</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Cognitive remediation</td>
<td>Overall -0.55 (-1.17/0.07)</td>
<td>Positive +0.75 (+0.12/+1.38)</td>
<td>Negative +0.11 (+0.50/0.72)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McLeod et al (2007)</td>
<td>10</td>
<td>10</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Hallucinations -1.06 (-1.96/-0.16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deng et al (2008)</td>
<td>64</td>
<td>64</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High risk</td>
<td>Unclear</td>
<td>No</td>
<td>Overall -0.92 (-1.28/-0.56)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>England et al (2007, 2008)</td>
<td>44</td>
<td>21</td>
<td>Low risk</td>
<td>Unclear</td>
<td>High risk</td>
<td>Low risk</td>
<td>No</td>
<td>Hallucinations -1.82 (-2.42/-1.22)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garety et al (2008) (no carer)</td>
<td>90*</td>
<td>90*</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>No</td>
<td>Overall -0.12 (-0.41/0.17)</td>
<td>Positive -0.17 (-0.46/0.12)</td>
<td>Negative -0.10 (-0.39/-0.19)</td>
<td>Hallucinations +0.07 (-0.34/0.49)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garety et al (2008) (carer)</td>
<td>21*</td>
<td>23*</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>No</td>
<td>Overall -0.31 (-0.90/0.27)</td>
<td>Positive -0.20 (-0.79/0.38)</td>
<td>Negative -0.17 (-0.75/+0.41)</td>
<td>Hallucinations -0.73 (-1.60/+0.14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jackson et al (2008)</td>
<td>31</td>
<td>31</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Befriending</td>
<td>Positive -0.05 (-0.54/0.44)</td>
<td>Negative -0.44 (-0.94/+0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LeComte et al (2006)</td>
<td>36</td>
<td>51</td>
<td>High risk</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Two control groups used</td>
<td>Overall -0.15 (-0.57/+0.27)</td>
<td>Positive -0.07 (-0.49/+0.35)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Sample Size</td>
<td>Low Risk</td>
<td>Unclear</td>
<td>High Risk</td>
<td>Low Risk</td>
<td>No</td>
<td>Overall</td>
<td>Positive</td>
<td>Negative</td>
<td>Measure of positive symptoms not used, since based on scores on only 2 PANSS items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------</td>
<td>----------</td>
<td>---------</td>
<td>-----------</td>
<td>----------</td>
<td>----</td>
<td>---------</td>
<td>----------</td>
<td>----------</td>
<td>-----------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wu et al (2008)</td>
<td>48</td>
<td>52</td>
<td>Unclear</td>
<td>High risk</td>
<td>Unclear</td>
<td></td>
<td>Overall</td>
<td>-1.52</td>
<td>(-1.96/-1.07)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farhall et al (2009)</td>
<td>45</td>
<td>47</td>
<td>Low risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>No</td>
<td>Overall</td>
<td>+0.01</td>
<td>(-0.39/+0.42) Positive -0.20 (-0.61/+0.20) Negative +0.36 (-0.05/+0.77)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fowler et al (2009)</td>
<td>20</td>
<td>23</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Low risk</td>
<td>No</td>
<td>Overall</td>
<td>-0.34</td>
<td>(-0.94/+0.25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haddock et al (2009)</td>
<td>approx 28*</td>
<td>approx 29*</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Social activity therapy</td>
<td>Overall</td>
<td>-0.23 (-0.75/+0.28) Positive -0.16 (-0.68/+0.35) Negative -0.38 (-0.90/+0.13) Hallucinations -0.11 (-0.67/+0.45) Numbers in each group estimated from total study numbers.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penn et al (2009)</td>
<td>32</td>
<td>33</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Supportive therapy</td>
<td>Hallucinations -0.12 (-0.61/+0.36)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pinninti et al (2010)</td>
<td>14</td>
<td>11</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>No</td>
<td>Positive</td>
<td>-0.34</td>
<td>(-1.11/+0.43) Hallucinations -0.13 (-0.89/+0.64)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edwards et al (2011) (thioridazine)</td>
<td>12</td>
<td>11</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Low risk</td>
<td>No</td>
<td>Positive</td>
<td>-0.64</td>
<td>(-1.45/+0.17) Negative -0.62 (-1.43/+0.19)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edwards et al (2011) (clozapine)</td>
<td>11</td>
<td>14</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Low risk</td>
<td>No</td>
<td>Positive</td>
<td>+0.19</td>
<td>(-0.58/+0.95) Negative -0.30 (-1.07/+0.47)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grant et al (2011)</td>
<td>31</td>
<td>29</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>No</td>
<td>Negative</td>
<td>-0.06</td>
<td>(-0.56/+0.44) Negative symptom ES combined from SANS subscale scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Klingberg et al (2011)</td>
<td>99</td>
<td>99</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Cognitive remediation</td>
<td>Negative +0.12 (-0.16/+0.40)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van der Gaag et al (2011)</td>
<td>109</td>
<td>97</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Low risk</td>
<td>No</td>
<td>Overall</td>
<td>-0.09</td>
<td>(-0.37/+0.18) Positive -0.46 (-0.73/-0.18)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White et al (2011)</td>
<td>14</td>
<td>10</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Low risk</td>
<td>No</td>
<td>Positive</td>
<td>-0.46</td>
<td>(-1.25/+0.33) Negative -0.40 (-1.20/+0.39)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lincoln et al (2012)</td>
<td>40</td>
<td>40</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Low risk</td>
<td>No</td>
<td>Overall</td>
<td>-0.29</td>
<td>(-0.72/+0.15) Positive -0.41 (-0.85/+0.03) Negative +0.10 (-0.33/+0.54) Blindness compromised</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Sample Size</td>
<td>Risk</td>
<td>Risk</td>
<td>Risk</td>
<td>Risk</td>
<td>Interaction</td>
<td>Hallucinations</td>
<td>Hallucinations ES combined from PSYRATS subscale scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>-------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------------</td>
<td>----------------</td>
<td>------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shawyer et al (2012)</td>
<td>19</td>
<td>18</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Befriending</td>
<td>+0.53 (-0.011/+1.17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Granholm et al (2012)</td>
<td>26</td>
<td>32</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>Goal focused supportive contact</td>
<td>Positive -0.11 (-0.62/+0.40) Negative -0.05 (-0.56/+0.46)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rathod et al (2012)</td>
<td>13</td>
<td>14</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>No</td>
<td>Overall -0.18 (-0.91/+0.56) Negative +0.01 (-0.72/+0.74) Hallucinations -0.15 (-0.89/-0.58)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Numbers different in different analyses.
**Numbers varied depending on whether ITT or completer data could be extracted.

**Additional references cited in table**
### Fig. DS1 Forest plot of studies of negative symptoms

<table>
<thead>
<tr>
<th>Study name</th>
<th>Statistics for each study</th>
<th>Sample size</th>
<th>Hedge's g and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drury 1996</td>
<td>-0.19</td>
<td>20</td>
<td>0.42</td>
</tr>
<tr>
<td>Daniels 1996</td>
<td>-0.67</td>
<td>20</td>
<td>0.05</td>
</tr>
<tr>
<td>Levine 1998</td>
<td>-2.51</td>
<td>6</td>
<td>-1.06</td>
</tr>
<tr>
<td>Pitt 1999</td>
<td>-0.34</td>
<td>19</td>
<td>0.30</td>
</tr>
<tr>
<td>Tarrier 1999</td>
<td>-0.14</td>
<td>23</td>
<td>0.35</td>
</tr>
<tr>
<td>Lader 2000</td>
<td>0.01</td>
<td>55</td>
<td>0.41</td>
</tr>
<tr>
<td>Sensky 2000</td>
<td>0.07</td>
<td>44</td>
<td>0.48</td>
</tr>
<tr>
<td>Granholm 2002</td>
<td>-0.03</td>
<td>8</td>
<td>0.93</td>
</tr>
<tr>
<td>Gunley 2003</td>
<td>-0.35</td>
<td>66</td>
<td>-0.01</td>
</tr>
<tr>
<td>Jolley 2003</td>
<td>-0.08</td>
<td>7</td>
<td>0.88</td>
</tr>
<tr>
<td>Rector 2003</td>
<td>-0.49</td>
<td>24</td>
<td>0.12</td>
</tr>
<tr>
<td>Wang 2003</td>
<td>-0.71</td>
<td>125</td>
<td>0.46</td>
</tr>
<tr>
<td>Bechdolff 2004</td>
<td>0.16</td>
<td>40</td>
<td>0.55</td>
</tr>
<tr>
<td>Startup 2004</td>
<td>-0.32</td>
<td>34</td>
<td>0.16</td>
</tr>
<tr>
<td>Cellier 2005</td>
<td>-0.01</td>
<td>15</td>
<td>0.71</td>
</tr>
<tr>
<td>Granholm 2005</td>
<td>0.15</td>
<td>37</td>
<td>0.59</td>
</tr>
<tr>
<td>Valmaggia 2005</td>
<td>0.06</td>
<td>36</td>
<td>0.89</td>
</tr>
<tr>
<td>Barroclough 2006</td>
<td>-0.05</td>
<td>54</td>
<td>0.33</td>
</tr>
<tr>
<td>Gauliano 2006</td>
<td>0.09</td>
<td>21</td>
<td>0.52</td>
</tr>
<tr>
<td>Penades 2006</td>
<td>0.11</td>
<td>20</td>
<td>0.59</td>
</tr>
<tr>
<td>Garety 2008 no carer</td>
<td>-0.10</td>
<td>50</td>
<td>0.19</td>
</tr>
<tr>
<td>Garety 2008 carer</td>
<td>-0.17</td>
<td>21</td>
<td>0.41</td>
</tr>
<tr>
<td>Jackson 2003</td>
<td>-0.44</td>
<td>31</td>
<td>0.05</td>
</tr>
<tr>
<td>LeConte 2008</td>
<td>0.14</td>
<td>56</td>
<td>0.23</td>
</tr>
<tr>
<td>Forth 2003</td>
<td>0.36</td>
<td>45</td>
<td>0.77</td>
</tr>
<tr>
<td>Haddock 2009</td>
<td>-0.33</td>
<td>28</td>
<td>0.13</td>
</tr>
<tr>
<td>Grant 2011</td>
<td>-0.03</td>
<td>31</td>
<td>0.44</td>
</tr>
<tr>
<td>Edwards 2011 TRZ</td>
<td>-0.62</td>
<td>12</td>
<td>0.19</td>
</tr>
<tr>
<td>Edwards 2011 CLOZ</td>
<td>-0.30</td>
<td>11</td>
<td>0.47</td>
</tr>
<tr>
<td>Klingberg 2011</td>
<td>0.12</td>
<td>14</td>
<td>0.16</td>
</tr>
<tr>
<td>White 2011</td>
<td>-0.41</td>
<td>14</td>
<td>0.39</td>
</tr>
<tr>
<td>Granholm 2012</td>
<td>-0.05</td>
<td>26</td>
<td>0.46</td>
</tr>
<tr>
<td>Lincoln 2012</td>
<td>0.10</td>
<td>40</td>
<td>0.33</td>
</tr>
<tr>
<td>Rathod 2012</td>
<td>0.01</td>
<td>13</td>
<td>0.72</td>
</tr>
<tr>
<td>-0.13</td>
<td>-0.25</td>
<td>14</td>
<td>-0.01</td>
</tr>
</tbody>
</table>
**Fig. DS2** Forest plot of studies of hallucinations

<table>
<thead>
<tr>
<th>Study name</th>
<th>Statistics for each study</th>
<th>Sample size</th>
<th>Hedges’s g and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hedges's g</td>
<td>Lower limit</td>
<td>Upper limit</td>
</tr>
<tr>
<td>Lewis 2002</td>
<td>-0.11</td>
<td>-0.46</td>
<td>0.25</td>
</tr>
<tr>
<td>Durham 2003</td>
<td>-0.16</td>
<td>-0.69</td>
<td>0.36</td>
</tr>
<tr>
<td>Trouer 2004</td>
<td>-0.75</td>
<td>-1.45</td>
<td>-0.05</td>
</tr>
<tr>
<td>Vallmaggia 2005</td>
<td>-0.74</td>
<td>-1.28</td>
<td>-0.21</td>
</tr>
<tr>
<td>Wykes 2005</td>
<td>-0.03</td>
<td>-0.48</td>
<td>0.43</td>
</tr>
<tr>
<td>Cather 2005</td>
<td>-0.20</td>
<td>-0.92</td>
<td>0.53</td>
</tr>
<tr>
<td>McLenn 2007</td>
<td>-1.06</td>
<td>-1.96</td>
<td>-0.16</td>
</tr>
<tr>
<td>England 2008</td>
<td>-1.82</td>
<td>-2.42</td>
<td>-1.22</td>
</tr>
<tr>
<td>Garety 2008 no carer</td>
<td>0.08</td>
<td>-0.34</td>
<td>0.49</td>
</tr>
<tr>
<td>Garety 2008 carer</td>
<td>-0.73</td>
<td>-1.60</td>
<td>0.14</td>
</tr>
<tr>
<td>Haddock 2009</td>
<td>-0.11</td>
<td>-0.67</td>
<td>0.45</td>
</tr>
<tr>
<td>Penn 2009</td>
<td>-0.13</td>
<td>-0.61</td>
<td>0.36</td>
</tr>
<tr>
<td>Pinninti 2010</td>
<td>-0.13</td>
<td>-0.89</td>
<td>0.64</td>
</tr>
<tr>
<td>Rathod 2012</td>
<td>-0.15</td>
<td>-0.89</td>
<td>0.58</td>
</tr>
<tr>
<td>Shawyer 2012</td>
<td>0.53</td>
<td>-0.11</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>-0.34</td>
<td>-0.61</td>
<td>-0.07</td>
</tr>
</tbody>
</table>
Data supplement 2

Excluded studies
(Note: only studies employing a control group are listed)

Milton et al, 1993
- Effect sizes not extractable.
Tarrier et al, 1993
- Waitlist control was not a parallel group.
- Used an unpublished rating scale (for hallucinations).
Garety et al, 1994
- Not randomized.
Kemp et al, 1996
- Intervention now universally regarded as compliance therapy.
LeCompte & Pelc, 1996
- Only relapse data reported.
Buchkremer et al, 1997/Hornung et al 1995
- CBT part of a multimodal intervention, not directed to symptoms
Jackson et al, 1998
- Not randomized.
Barrowclough et al, 1999
- Multimodal intervention given to pairs of patients and family members. Only half the pts received the CBT/psychotic symptoms part of package.
Bailer et al, 2001
- Not randomized.
Barrowclough et al, 2001
- Study carried out on dual diagnosis patients.
Klingberg et al, 2001
- Intervention not CBT
Bach and Hayes, 2002
- No symptom measures reported.
Hall and Tarrier, 2003
- CBT directed only to self-esteem.
Power et al, 2003
- CBT directed to suicidality. No diagnostic specification of patients.
Jenner et al, 2004
- Multimodal intervention.
Kuipers et al, 2004
- Multimodal intervention.
Morrison et al, 2004
- Not randomized.
Jackson et al, 2005
- Not adequately randomized (randomization by alternation).
Temple et al, 2005
- Not randomized.
Baker et al, 2006
- Study carried out on dual-diagnosis patients.
Garety et al, 2006
- Multimodal intervention.
Grawe et al, 2006
- Multimodal intervention.
Barton et al, 2009
- Same study as Fowler et al (2009).
Barretto et al, 2009
- Not randomized.
Chadwick et al, 2009
- Effect sizes not extractable. Trial of mindfulness therapy (related to acceptance and commitment therapy).
Gleeson et al, 2009
- Multimodal intervention (CBT, family therapy, other optional modules) provided to both patients and families.
Jackson et al, 2009
- CBT directed to PTSD symptoms.
Barrowclough et al, 2010
- Study carried out on dual diagnosis patients.
Peters et al, 2010
- Study carried out on patients with psychotic symptoms but with no further diagnostic specification.
Moritz et al, 2011
- Multimodal intervention with overall small component of CBT.
Mortan et al, 2011
- Not adequately randomized.
Premkumar et al, 2011
- Not randomized.
Palma-Sevillano et al, 2011
- Multimodal intervention.
Tundo et al, 2012
- CBT directed to OCD symptoms.

References for excluded studies


Cognitive-behavioural therapy for the symptoms of schizophrenia: systematic review and meta-analysis with examination of potential bias
S. Jauhar, P. J. McKenna, J. Radua, E. Fung, R. Salvador and K. R. Laws
BJP 2014, 204:20-29.
Access the most recent version at DOI: 10.1192/bjp.bp.112.116285