Background
Cognitive–behavioural therapy (CBT) is considered to be effective for the symptoms of schizophrenia. However, this view is based mainly on meta-analysis, whose findings can be influenced by failure to consider sources of bias.

Aims
To conduct a systematic review and meta-analysis of the effectiveness of CBT for schizophrenic symptoms that includes an examination of potential sources of bias.

Method
Data were pooled from randomised trials providing end-of-study data on overall, positive and negative symptoms. The moderating effects of randomisation, masking of outcome assessments, incompleteness of outcome data and use of a control intervention were examined. Publication bias was also investigated.

Results
Pooled effect sizes were $-0.33$ (95% CI $-0.47$ to $-0.19$) in 34 studies of overall symptoms, $-0.25$ (95% CI $-0.37$ to $-0.13$) in 33 studies of positive symptoms and $-0.13$ (95% CI $-0.25$ to $-0.01$) in 34 studies of negative symptoms. Masking significantly moderated effect size in the meta-analyses of overall symptoms (effect sizes $-0.62$ (95% CI $-0.88$ to $-0.35$) v. $-0.15$ (95% CI $-0.27$ to $-0.03$), $P=0.001$) and positive symptoms (effect sizes $-0.57$ (95% CI $-0.76$ to $-0.39$) v. $-0.08$ (95% CI $-0.18$ to $0.03$), $P<0.001$). Use of a control intervention did not moderate effect size in any of the analyses. There was no consistent evidence of publication bias across different analyses.

Conclusions
Cognitive–behavioural therapy has a therapeutic effect on schizophrenic symptoms in the ‘small’ range. This reduces further when sources of bias, particularly masking, are controlled for.

Declaration of interest
None.
psychological intervention, although an alternative view exists which considers that applying the logic of placebo-controlled trials to psychotherapy research is flawed. 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 (for example), 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 included. When a study used two control groups the effect size was calculated by 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 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., 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 suicidality 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.
in all analyses. Heterogeneity was examined by means of $Q$ and $I^2$ statistics.

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\textsuperscript{65} 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 syndromes by Smith et al.\textsuperscript{66} Similarly, if a study provided separate delusion and hallucination subscale scores, these were averaged assuming a correlation of 0.34 from Smith et al.\textsuperscript{65}

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\textsuperscript{46} we averaged scores from the four subscales of the SANS employed, using correlations reported in Smith et al.\textsuperscript{65} Another study used factor scores from the SANS,\textsuperscript{42} and these were also averaged using published data concerning the correlations between them.\textsuperscript{66}

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.\textsuperscript{67}

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\textsuperscript{68} (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 \textit{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.\textsuperscript{69}

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\textsuperscript{14,36,70,73} 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 remediation 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) were not included in this analysis.

Publication bias
This was examined using three statistical techniques, Duval & Tweedie's trim and fill, Begg & Mazumdar's rank correlation test and Egger's 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 articles reporting 52 studies were finally included (two studies reported on two separate samples). The studies were heterogeneous, 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 $-0.33$ (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 $-0.09$ (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 $-0.10$ (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, $P = 0.10$).
Fig. 2 Forest plot of studies in the meta-analysis of overall symptoms.

CBT, cognitive–behavioural therapy.

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 $0.24$) $n$. $-0.55$ (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. and the Cochrane Collaboration, 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. 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. 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. Although NICE 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 influences of sources of bias

The importance of masking in trials of psychological treatments is recognised, even though less attention often seems to be paid to it than to other aspects of methodology. Nevertheless, its moderating effects have only previously been examined twice. Zimmerman et al. found effect sizes of $-0.29$ and $-0.54$ in studies with and without masked assessment in their meta-analysis of 15 studies of positive symptoms, but the difference was not significant. Wykes et al. found values of $-0.31$ in 14 masked studies compared with $-0.49$ in 10 non-masked studies; they did not state whether this represented a significant difference. Our meta-analysis of a larger set of studies found considerably

### Table 1: Comparisons between studies at high and low risk of bias from masking, allocation concealment and incomplete outcome data

<table>
<thead>
<tr>
<th>High risk of bias</th>
<th>Low risk of bias</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effect size (95% CI)</strong></td>
<td><strong>Studies, n</strong></td>
</tr>
<tr>
<td>Sequence generation</td>
<td>Overall symptoms</td>
</tr>
<tr>
<td>Positive symptoms</td>
<td>$-0.07$ (95% CI $0.49$ to $0.35$)</td>
</tr>
<tr>
<td>Negative symptoms</td>
<td>$0.14$ (95% CI $0.28$ to $0.56$)</td>
</tr>
<tr>
<td>Allocation concealment</td>
<td>Overall symptoms</td>
</tr>
<tr>
<td>Positive symptoms</td>
<td>$-0.96$ (95% CI $1.61$ to $0.32$)</td>
</tr>
<tr>
<td>Negative symptoms</td>
<td>$0.09$ (95% CI $0.52$ to $0.69$)</td>
</tr>
<tr>
<td>Masking</td>
<td>Overall symptoms</td>
</tr>
<tr>
<td>Positive symptoms</td>
<td>$-0.57$ (95% CI $0.76$ to $0.39$)</td>
</tr>
<tr>
<td>Negative symptoms</td>
<td>$-0.22$ (95% CI $0.51$ to $0.08$)</td>
</tr>
<tr>
<td>Incomplete outcome data</td>
<td>Overall symptoms</td>
</tr>
<tr>
<td>Positive symptoms</td>
<td>$-0.18$ (95% CI $0.60$ to $0.25$)</td>
</tr>
<tr>
<td>Negative symptoms</td>
<td>$-0.05$ (95% CI $0.56$ to $0.46$)</td>
</tr>
</tbody>
</table>

n/a, not applicable.

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

<table>
<thead>
<tr>
<th>Effect size (95% CI)</th>
<th>Studies, n</th>
<th>Effect size (95% CI)</th>
<th>Studies, n</th>
<th>Q(B)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBT v. TAU</td>
<td>Overall symptoms</td>
<td>$-0.33$ (95% CI $-0.45$ to $-0.21$)</td>
<td>21</td>
<td>$-0.32$ (95% CI $-0.74$ to $0.09$)</td>
<td>9</td>
</tr>
<tr>
<td>Positive symptoms</td>
<td>$-0.31$ (95% CI $-0.45$ to $-0.17$)</td>
<td>19</td>
<td>$-0.24$ (95% CI $-0.54$ to $0.06$)</td>
<td>10</td>
<td>0.17</td>
</tr>
<tr>
<td>Negative symptoms</td>
<td>$-0.17$ (95% CI $-0.33$ to $-0.02$)</td>
<td>20</td>
<td>$-0.08$ (95% CI $-0.29$ to $0.13$)</td>
<td>12</td>
<td>0.49</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.\textsuperscript{99–101} 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 NICE\textsuperscript{3} meta-analysis) have found a significant advantage for CBT against overall symptoms,\textsuperscript{94,96} and only one for positive symptoms.\textsuperscript{11} 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 Kirsch\textsuperscript{30} and Bentall,\textsuperscript{31} that pill placebo and control for the non-specific effects of psychotherapy are not equitable. 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.\textsuperscript{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, NICE\textsuperscript{3} 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 \textit{et al.}\textsuperscript{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: Cuipers \textit{et al.}\textsuperscript{104} 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.
Implications

Should CBT for schizophrenia continue to be recommended in clinical practice? Given that we, and others including NICE, 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, the UK government’s continued vigorous advocacy of this form of treatment (for example see The All Party Parliamentary Group on Mental Health) 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 speciality 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.

S. Jaanah, MB, ChB, BSc(Hons), MRCPsych, Department of Psychiatry, Institute of Mental Health, Singapore; J. Rudda, MD, PhD, FXIMAG Germanes Hospitals Research Foundation, Barcelona and CIBERSAM, Spain; E. Fung, MD, Department of Psychiatry, Faculty of Medicine, University of Calgary, Alberta, Canada; R. Salvador, PhD, FXIMAG Germanes Hospitals Research Foundation, Barcelona and CIBERSAM, Spain; K. R. Laws, PhD, School of Life and Medical Sciences, University of Herfordshire, Hatfield, UK

Correspondence: P. J. McKenna, Benito Menni CASM, Germanes Hospitalàries del Sagrat Cor de Jesús, C/ Doctor Antoni Pujades 38-43, 08030 - Sant Boi de Llobregat (BARCELONA), Spain. Email: mckenna.peter10@gmail.com

First received 13 Jun 2012, final revision 16 May 2013, accepted 10 Jun 2013

Acknowledgements

This work was supported in part by (a) Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), (b) several grants from the Instituto de Salud Carlos III including to demonstrate any effectiveness against relapse, the UK National Institute of Health and Clinical Excellence.

References


