Schizophrenia Bulletin vol. 34 no. 3 pp. 523–537, 2008
doi:10.1093/schbul/sbm114
Advance Access publication on October 25, 2007

Cognitive Behavior Therapy for Schizophrenia: Effect Sizes, Clinical Models, and Methodological Rigor

Til Wykes[1,2], Craig Steel[2], Brian Everitt[3], and Nicholas Tarrier[4]

[2]Department of Psychology and; [3]Department of Biostatistics and Computing, Institute of Psychiatry, King's College London, De Crespigny Park, London SE5 8AF, UK; [4]Division of Clinical Psychology, School of Psychological Sciences, University of Manchester

Background: **Guidance in the United States and United Kingdom has included cognitive behavior therapy for psychosis (CBTp) as a preferred therapy. But recent advances have widened the CBTp targets to other symptoms and have different methods of provision, eg, in groups.** ***Aim:*** **To explore the effect sizes of current CBTp trials including targeted and nontargeted symptoms, modes of action, and effect of methodological rigor.** ***Method:*** **Thirty-four CBTp trials with data in the public domain were used as source data for a meta-analysis and investigation of the effects of trial methodology using the Clinical Trial Assessment Measure (CTAM).** ***Results:*** **There were overall beneficial effects for the target symptom (33 studies; effect size = 0.400 [95% confidence interval {CI} = 0.252, 0.548]) as well as significant effects for positive symptoms (32 studies), negative symptoms (23 studies), functioning (15 studies), mood (13 studies), and social anxiety (2 studies) with effects ranging from 0.35 to 0.44. However, there was no effect on hopelessness. Improvements in one domain were correlated with improvements in others. Trials in which raters were aware of group allocation had an inflated effect size of approximately 50%–100%. But rigorous CBTp studies showed benefit (estimated effect size = 0.223; 95% CI = 0.017, 0.428) although the lower end of the CI should be noted. Secondary outcomes (eg, negative symptoms) were also affected such that in the group of methodologically adequate studies the effect sizes were not significant.** ***Conclusions:*** **As in other meta-analyses, CBTp had beneficial effect on positive symptoms. However, psychological treatment trials that make no attempt to mask the group allocation are likely to have inflated effect sizes. Evidence considered for psychological treatment guidance should take into account specific methodological detail.**

Key words: CBTp/schizophrenia/meta-analysis/trials/symptoms/functioning

Introduction

Cognitive behavior therapy (CBT) has been accepted as a treatment for affective disorders for a number of years and has been fully integrated into services since the 1980s. However, despite the case studies by Beck[1] and Shapiro and Ravenette[2] in the 1950s,' specific symptom interventions for schizophrenia did not appear until much later. Psychotherapy for schizophrenia in the form of psychodynamic therapies had been discredited and during the period of deinstitutionalization symptoms were treated merely as behaviors to be extinguished (eg, Liberman et al[3] and Meichenbaum and Cameron[4]). However, despite earlier optimism neither medication nor behavioral treatments successfully extinguished symptoms which were either present sporadically or remained continuously despite adequate treatment. Theoretical underpinnings such as the stress-vulnerability models were developed to understand not only the development of the disorder but also its maintenance. These also began to be informed by research on expressed emotion (eg, Brown et al[5] and Butzlaff and Hooley[6]) and so began to include social and psychological markers as well as biological ones. The difficulty in identifying rigorous and unambiguous psychosocial markers may have hampered further development of this area.

CBT for affective disorders became accepted in the health services through government guidelines (eg, UK National Institute for Health and Clinical Excellence), but it also increased its theoretical research base. It was inevitable that eventually some of the developed techniques would be used for people with a diagnoses of schizophrenia. The first controlled studies on cognitive behavior therapy for psychosis (CBTp) emerged in the early 1990s in the United Kingdom, and this treatment

[1]To whom correspondence should be addressed; tel: 44-20-7848-0596, fax: 44-20-7848-5006, e-mail: t.wykes@iop.kcl.ac.uk.

has developed and included some of the theoretical underpinnings of CBT for other disorders. However, unlike CBT for other disorders, which have its roots in Beck Philadelphia Institute, CBTp developed independently perhaps because the main research bases for the 2 types of CBT were separated by the Atlantic Ocean. There is much speculation and argument about why CBTp emerged first in the United Kingdom, but it may be that the different service structures in the United Kingdom within which clinical psychologists worked were more encouraging for nonmedical approaches to drug-resistant psychotic symptoms. The implementation of CBTp was against a tide of skepticism about the development of psychotherapy for people with psychosis in both countries and specific evidence of poor outcomes.[7] There was also optimism about the likelihood of improved medication now somewhat tempered by the Clinical Antipsychotic Trials in Intervension Effectiveness[8] and Cost Utility in the Latest Antipsychotic Drugs in Schizophrenia Study trials[9] and the positive evidence on behavioral approaches to psychosocial rehabilitation which may have contributed to resistance about the development of this form of treatment. However, although the practice of CBTp was enthusiastically grasped by UK clinical psychologists, it has now also developed in the United States with studies approaching the provision of CBTp differently, eg, by more group-based approaches. Again, service structures and the availability of skilled clinical psychology staff probably contribute to these variations. Within the United Kingdom, psychological treatments are typically individualized and based upon an idiosyncratic case formulation (see Tarrier and Calam[10] and Tarrier[11]). In the United States, the recent move to more evidence-based practice has been to counter the past standard practice of providing psychological treatment from a nonsystematic, nonmanualized perspective. Thus, even if the actual intervention techniques used were similar in the United States and the United Kingdom, their strategic application could differ.

Reviews of studies of CBTp have suggested that they are useful for the treatment of schizophrenia.[12–23] The next step is therefore the incorporation of these treatments into services. In the United Kingdom, the National Institute for Clinical Excellence included cognitive behavior therapy (CBTp) in its preferred list of treatments for schizophrenia.[24] UK National Health Services are now implementing this guidance because patients in UK services have the right to expect that this treatment will be available. This therapy has also been considered in the Schizophrenia Patient Outcome Report Team guidance in the United States and has been recommended.[25] There are now more published studies available, and there has been an expansion of the likely symptom targets for CBTp to include, in addition to positive symptoms, negative symptoms, depression, and anxiety, and also some combination therapies to reduce harm for those who have a dual diagnosis and substance misuse.[26] All these symptoms have more recently been recognized as those which hamper recovery and affect quality of life.

Guidance decisions are mainly based on evidence from randomized controlled trials (RCTs), and despite recent criticisms of their appropriateness in mental health,[27,28] RCTs remain the gold standard by which all treatments are judged.[29–31] Such trials reflect the scientific method of demonstrating the value of a new treatment in comparison to an appropriate control group by minimizing all possible sources of bias which could render unsafe the conclusion that the new treatment is beneficial.[29] There are established conventions concerning how clinical trials should be carried out and analyzed[29,32] and these have been formalized in the Consolidated Standards for Reporting of Trials (CONSORT) statement.[33,34]

Despite these guidelines, there is still considerable variability and room for improvement in rigor. Thornley and Adams[35] reported that in 2000 (mainly drug) trials, methodological rigor had not improved to a minimal standard over a period of 35 years. Such rigor is known to affect the estimation of treatment effects. For instance, poorer quality masking of allocation of treatments has been shown to be associated with up to 40% increased estimate of benefit in circulatory and digestive diseases, mental health, obstetrics, and childbirth.[36,37]

This raises the question of whether differences in methodology will also inflate the effects of psychological treatment trials. Marshall et al[38] report that in 150 nondrug trials, one-third of the claims that treatment was superior to control would not have been made if published scales had been used in the assessment. It is of course possible that other aspects of the design and method could also affect the results of psychological treatment trials. But even though meta-analyses have had a direct effect on treatment guidance, methodological rigor of individual trials is hardly referred to,[13,18–23] and the strong claim has been made that the quality of trials included must be investigated in order to ensure that meta-analyses provide valid estimates of the true effects of treatment.[39]

Aspects of psychological therapy, in addition to methodology, may also affect the treatment effect, in particular the content. There have been no investigations of the different elements of treatment in a direct head-to-head comparison, but there may be more general influences that could be investigated in a meta-analysis. CBTp in schizophrenia varies in its emphasis on cognitive and/or behavioral dimensions of therapy and at the extreme end of the continuum merges with some form of psychodynamic treatments. Clinical emphasis in any model is also dependent on the services in which it is provided and the background professional training of the therapists which differs from country to country. There may therefore be particular differences to explore in studies between countries. The superiority of any model has

never been investigated, and as we move from efficacy to effectiveness studies this is an important consideration.

Since the publication of the latest meta-analysis, more data from newer CBTp trials have become available. The aim of this article is to look in more detail at this expanded data set with its extended list of outcomes and investigate the effects of different methodological attributes as well as clinical models in the calculation of benefit. In particular, we will estimate the possible inflation of effect size due to lack of masking of participants to groups because it has been identified as a potent methodological variable in the inflation of the treatment effect.

Methods

Trial Inclusion

Multiple systematic searches of Embase, Medline, Current Contents, Web of Science, PsychInfo, and the Cochrane Collaborative Register of Trials were performed using the following search terms either as key terms or as key words:

(SCHIZO* or SCHIZOPHRENIA or SCHIZOAFFECTIVE DISORDER) AND (COGNITIVE THERAPY or COGNITIVE BEHAVIOUR THERAPY or COGNITIVE BEHAVIOR THERAPY) AND (RANDOM or RANDOMISED CONTROL TRIAL or CLINICAL TRIAL).

One hundred forty-one publications were identified which were hand searched including their reference lists. In addition, the reference lists from reviews and meta-analyses[12–23] were hand searched. This produced a total of 35 articles that were judged potentially eligible for this review and read independently by 2 of the authors. In addition, using our own knowledge of work in this area through conference presentations (2004–2006), the Beck psychosis networks, and familiarity with appropriate research groups around the globe, one further trial was added that had been submitted.[40] This study has subsequently been published. The criteria to retain a publication were

- the studies sample had to contain a majority of people with a diagnosis of schizophrenia;
- all patients received standard psychiatric care (TAU, treatment as usual) including appropriate medication;
- in the experimental group, CBT was an adjunct to TAU;
- there was a control group;
- there was an allocation procedure;
- CBT treatment was targeted at one of the following (positive or negative symptoms of psychosis, functioning, mood, hopelessness/suicidality, or social anxiety).

Trials were excluded if they were uncontrolled, tested forms of psychotherapy other than CBT, or whose outcomes did not include symptoms at the end of treatment. For example, Kemp et al[41] was excluded because its aim and outcome were medication compliance and its method was motivational interviewing rather than more conventional CBT. Two studies were excluded because the authors did not make the data available that was necessary for inclusion in the meta-analysis, leaving a total sample of 34 studies.

Effect Size Calculation

Effect sizes of the treatment trials were calculated from the following equation.[19]

$$\text{Effect size} = (M_t - M_c)/\text{SDc}$$

where M_t is the mean of the CBTp group at posttreatment, M_c is the mean of the control group at posttreatment, and this is divided by the SD of the control group of participants at posttreatment. Outcome was considered in relation to positive symptoms, negative symptoms, functioning, mood, hopelessness, and social anxiety where such data were available, and the measures chosen to reflect these outcomes in the calculation of the effect size was a continuous one. With respect to positive symptoms, this was, in most cases, a summary score from a reliable measure such as the Positive and Negative Syndrome Scale (PANSS) or Brief Psychiatric Rating Scale, but other measures such as Psychotic Symptoms Rating Scales were also used when these were unavailable or when the authors had powered their study on a specific continuous measure. The control group was considered to be the TAU group or a control adjunct treatment that had been hypothesized to be inactive for the main outcome. All effect sizes were recalculated from the study data except in one case in which the data were insufficiently reported and so the effect size reported in a previous meta-analysis was used.[19] The main consideration was the effect after treatment; so we have not included any follow-up data.

Measures

Features Used to Assess Quality of Trial Reports. A list of relevant features were extracted from the CONSORT guidelines[34] that are the current convention for describing clinical trials in major medical journals, eg, *Lancet*, *British Medical Journal*, *Journal of the American Medical Association,* and *Archives of General Psychiatry*. Expert opinion from psychologists, psychiatrists, statisticians, and methodologists was then sought on this checklist. These opinions provided face validity. Individual features were differentially weighted based on previous data on methodological characteristics that can influence outcome (eg, Thornley and Adams[35], Moher et al[36], Marshall et al[38], Chalmers et al[42], Jadad et al[43], Juni et al[44], Juni et al[45], Kazdin and Bass[46], and Sterne et al[47]). The resulting list had 15 items grouped into 6 areas of trial design: sample size and recruitment method, allocation to treatment, assessment of outcome, control

groups, description of treatments, and analysis (see below). Items were weighted in score depending on the importance highlighted by previous methodology articles and meta-analyses. For example, the total item score for allocation to treatment group was higher than the section on descriptions of active treatment.

We did not want to confuse the rating of methodological quality with the quality of the report[39,48]; so issues were clarified with the trial researchers, and where necessary data or information were requested that were not available within the published report. Subsequently, we made the ratings of methodological quality of each individual trial available to the trial researchers and asked for their comments.

The methodology areas included within the rating scale (Clinical Trial Assessment Measure: CTAM) are described below. A more detailed account with specific scores is given in Tarrier and Wykes.[49]

Sample Characteristics

Recruitment via volunteers or referrals of "suitable" patients by clinicians will not produce as representative a sample as a geographic or epidemiological cohort. In addition, we have included that sample size should be based on adequate power calculations.

Allocation to Treatment

The process of random allocation needs to be appropriate and clearly described,[50] and the sequence is concealed from the research team because this has been shown to be associated with larger treatment effects.[51]

Assessment of Outcome

Standardized assessment methods should be used[38] and collected independently of treatment by assessors who were unaware of treatment allocation (normally called blinded or masked assessment).

Control Groups

A control treatment that includes standard psychiatric care or TAU is a prerequisite. The use of a control treatment that includes another psychological treatment, such as supportive counselling or a placebo treatment (eg, befriending), is desirable to allow the estimation of the nonspecific effects of treatment to be assessed.

Description of Treatments

Treatments should be described so that they can be independently replicated; this would be aided by a manual or protocol. Assessment of adherence to the treatment protocol or some method of treatment quality assessment should also be carried out. This is an important aspect of psychological treatments that will not be covered by more generic rating scales of trial quality.

Analysis

The most acceptable methods for statistical analysis were judged (by an experienced trial statistician) as being appropriate to the data and the trial design. For instance, the results should be analyzed on an analyzed-as-randomized (sometimes referred to as an intention-to-treat analysis) because it maintains the benefits of randomization.[32] Some commonly used procedures to accommodate missing data were not considered to be appropriate, eg, missing value substitution by "last observation carried forward" because of its underlying assumptions and likely optimism about the precision of the extracted effect size.[32]

Reliability of CTAM. Independent ratings by 2 of the authors of an initial 18 studies showed good blind inter-rater agreement of 0.96. The scale showed adequate internal consistency (Cronbach $\alpha = .697$). The total score and the scores for the 6 domains were used as the outcome measures. The accuracy of the CTAM scores was confirmed by the first author of each specific study in 27 cases. The authors of 6 studies did not comment, and one was not contactable.

Validity of CTAM. Face validity from experts and the CONSORT statement have already been described. In addition, concurrent validity was assessed within an initial 22 studies by correlation with the scores from 3 other scales devised to assess methodological rigor.[42,43,52] The correlations of these scales with the CTAM scores on the corpus of CBTp data were as follows: CTAM and Jadad, $\rho = .960$, $P < .001$; CTAM and Chalmers, $\rho = .93$, $P < .001$; and CTAM and Brown, $\rho = .80$, $P < .001$. This indicates that CTAM had excellent concurrent validity. Predictive validity will be tested in the relationship between CTAM scores and effect sizes. It might be expected that less rigorous studies (ie, more potential for bias) will produce larger effect sizes.

Rating of the Emphasis of the Clinical Model. Descriptions of the clinical characteristics of the underlying model and clinical techniques included in each study were extracted from the written reports, and all identifiers were removed so that the rater could remain masked to the identity of the study. An independent qualified CBTp psychologist ranked the studies on the basis of least to most behavioral. This was defined on the basis of whether there was more emphasis on issues in the past, eg, historical variables, which might have an impact on the future (least behavioral), or whether there was a focus on the here and now and the interpretation of current events (more behavioral).

Results

The Trials

Thirty-four studies of CBTp were identified through published records, conference presentations, and networks of CBTp specialists throughout the world. Data were available either in their published form or, where necessary, were made available to the authors.

Table 1. Number of Studies Providing Specific Outcomes (Numbers in Parenthesis are Those Studies That Specifically Targeted that Outcome)

	Positive Symptom	Negative Symptom	Functioning	Mood	Hopelessness	Social Anxiety	Total
Individual CBTp	27 (24)	19 (1)	12 (2)	12 (0)	3 (0)	0 (0)	27
Group CBTp	5 (4)	4 (1)	3 (0)	3 (0)	1 (0)	2 (2)	7
Total	32 (28)	23 (2)	15 (2)	15 (0)	4 (0)	2 (2)	34

Note: CBTp, cognitive behavior therapy for psychosis.

Twenty-five studies involved the treatment of chronic patients in the community, 7 studies of acutely ill patients, 1 of chronic inpatients, and 1 contained a mixed population. Twenty studies were from the United Kingdom; 5 from the United States; 2 each from Germany, Australia, and the Netherlands; and 1 each from Canada, Italy, and Israel. The average number of participants in each trial was 58.2 (range 11–353), and the median proportion of those lost to follow-up assessment was 14.5% (range 0%–45%). Seven studies (21%) had a dropout rate higher than 25%, which is the level above which many statisticians would question the validity of the study findings.

The studies had varied targets of intervention. Most focused on positive symptoms (one study concentrated on reducing the powerfulness of the voice[53]), although 2 targeted negative symptoms, 2 targeted functioning, and 2 targeted social anxiety. Twenty-seven studies were individual CBTp, while 7 studies were group CBTp. All studies have continuous measures of outcome for the target symptom.

Table 1 shows the total number of studies reporting an outcome with the number of studies whose intervention was targeted at a specific symptom shown in parenthesis. Table 2 shows a summary of the 34 studies included in the analyses including the methodological quality ratings.

Trial Quality

The maximum score for the CTAM is 100, and in this sample of trials, the mean score was 61.2 (SD 18.1) with a median of 56 and a range of 27–100 (see table 2). There was some variability in methodology. All, except 4 studies, had random allocation, and 5 studies did not have independent assessments of outcome measures (but they were different studies). However, few reports adequately described the process of assessor blinding or verified blinding at the end of the study. More than half of the studies did not use a statistical method that was judged to take satisfactory account of dropouts from assessment. Methodological criteria have evolved over time with later trials showing higher CTAM scores (Spearman correlation = .393, $P = .02$) and an increased quality in allocation policy (ie, independence from the research team and true random allocation) (Spearman correlation = 0.463, $P = .006$). Trials with larger sample sizes have better CTAM scores which probably reflects the likelihood of larger investments by funding organizations to be in higher quality trials (Spearman $\rho = .596$, $P < .001$).

Clinical Emphasis

The 34 trials differed in their clinical emphasis with some representing schema-dependent therapy and one even reporting an emphasis on the "links between current symptoms and earlier real life events." At the other end of the scale studies reported an emphasis on behavioral homework, relaxation, and the development of behavioral coping strategies. There was, however, considerable overlap. Behavioral rehearsal, eg, and the reversal of avoidance (or safety behaviors) were common to many of the approaches irrespective of their stated emphasis. The clinical emphasis as ranked by an independent CBTp specialist is given in table 2. Although not significant, group studies were more likely to have a more behavioral emphasis than individual studies ($t = 1.6$, $df = 32$, 95% CI = −14.8 to 1.9).

Clinical Model and Trial Quality

There was no significant association between the emphasis of the clinical model and methodological rigor of the trials as measured by the CTAM total score ($\rho = -.19$, $P = .28$). Neither the total score nor the domain scores for CTAM were associated with clinical emphasis.

Study Origin and Trial Quality

Those studies originating outside the United Kingdom had smaller sample sizes. In fact, only 2 non-UK studies had a sample size that was above 27 participants per group, which is the smallest sample size likely to identify a modest clinically significant effect.[46] Overall trial quality was significantly higher in UK trials (mean CTAM score: UK 67.6, non-UK 52.1; $t = -2.69$, 95% CI = −27.4 to −3.8). The differences in method, apart from sample size, were (1) in the way the sample was drawn—non-UK sites mainly relying on convenience samples, (2) the quality of random allocation, and (3) the quality of control group. There were no differences in the target symptom effect sizes or the clinical emphasis of the model

Table 2. Included Studies and Their Characteristics

Trial	Year	United Kingdom	Individual/ Group	Primary Aim	Positive Symptom Effect Size	Total CTAM (Max 100)	Sample (Max 10)	Allocation (Max 16)	Assessment (Max 32)	Control (Max 16)	Analysis (Max 15)	Treatment Description (Max 11)	Clinical Model
Milton et al[54]	1978	Y	I	P	0.78	52	0	13	26	10	0	3	25
Tarrier et al[55]	1993	Y	I	P	0.35	49	2	10	16	16	5	0	32
Garety et al[56]	1994	Y	I	P	0.55	39	2	0	16	6	9	6	14.5
Bentall et al[57]	1994	Y	I	P	0.29	53	2	10	23	10	5	3	11
Drury et al[58]	1996	Y	I	P	0.93	53	2	13	16	10	9	3	23
Kuipers et al[59]	1997	Y	I	P	0.37	63	7	16	13	6	15	6	4.5
Tarrier et al[60]	1998	Y	I	P	0.73	96	10	16	32	16	11	11	32
Daniels[61]	1998	N	G	N	0.64	42	2	0	26	6	0	8	34
Levine et al[62]	1998	N	G	P	2.36	56	0	10	26	6	11	3	32
Pinto et al[63]	1999	N	I	P	0.99	44	2	10	16	10	0	6	28
Haddock et al[64]	1999	Y	I	P	−0.49	56	2	10	26	10	5	3	24
Halperin et al[65]	2000	N	G	SA	n/a	27	2	10	6	6	0	3	8.5
Sensky et al[66]	2001	Y	I	P	0.14	81	7	16	26	10	11	11	2.5
Bailer et al[67]	2001	N	I	N	0.34	38	2	0	16	6	11	3	20
Barrowclough et al[26]	2001	Y	I	P	0.26	80	10	16	29	6	11	8	14.5
Lewis et al[68]	2002	Y	I	P	0.12	100	10	16	32	16	15	11	20
Turkington et al[69]	2002	Y	I	P	0.23	77	10	16	26	6	11	8	14.5
Durham et al[70]	2002	Y	I	P	−0.32	84	7	16	29	16	5	11	20
Hall and Tarrier[71]	2002	Y	I	P	0.88	41	2	16	6	6	5	6	29.5
Valmaggia et al[72]	2002	N	I	P	0.32	62	2	13	26	10	11	0	8.5
Granholm et al[73]	2002	N	I	F	0.62	40	2	10	16	6	0	6	14.5
Gumley et al[74]	2003	Y	I	P	0.19	53	10	10	16	6	5	6	6
Rector et al[75]	2003	N	I	P	0.28	55	2	13	26	6	5	3	7
Jolley et al[76]	2003	Y	I	P	−0.10	75	5	16	26	6	11	11	2.5
Kingsep et al[77]	2003	N	G	SA	n/a	56	2	0	26	6	11	11	29.5
Trower et al[53]	2004	Y	I	CH	1.75	71	7	16	26	6	5	6	1
Wiersma et al[78]	2004	N	I	P	0.65	54	7	16	16	6	6	3	26
Bechdolf et al[79]	2004	N	G	P	0.02	67	7	16	26	10	5	3	20
Startup et al[80]	2005	Y	I	P	0.44	64	10	16	16	6	5	11	45
Cather et al[81]	2005	N	I	P	0.04	56	2	16	26	6	0	11	27
Granholm et al[82]	2005	N	I	F	−0.07	87	10	13	32	6	15	11	14.5
Wykes et al[83]	2005	Y	G	P	0.02	79	10	16	26	6	15	6	20
Gaudiano and Herbert[84]	2006	N	I	P	0.47	45	2	14	3	6	15	6	10
Barrowclough et al[40]	2006	Y	G	P	0.04	87	10	16	29	6	15	11	14.5

Note: CTAM, Clinical Trial Assessment Measure; primary aim: P, positive; N, negative; SA, social anxiety; CH, command hallucinations; F, functioning; n/a, not applicable; y, yes; N, no.

Table 3. Results of Meta-analyses

	Mean Weighted Effect Size	95% Confidence Interval	Heterogeneity Test (*df*), Significance Level	No. of Studies	Sample Size
Target symptom	0.400	0.252, 0.548	74.1 (32), significant at the 5% level	33	1964
Positive symptoms	0.372	0.228, 0.516	61.7 (31), significant at the 5% level	32	1918
Negative symptoms	0.437	0.171, 0.704	118.1 (22), significant at the 5% level	23	1268
Functioning	0.378	0.154, 0.602	36.7 (14), significant at the 5% level	15	867
Mood	0.363	0.079, 0.647	52.7 (12) significant at the 5% level	15	953
Hopelessness	−0.190	−0.547, 0.166	10.0 (3), not significant	4	431
Social anxiety	0.353	n/a	n/a	2	61

between studies originating inside and outside the United Kingdom although it was clear that the UK studies spanned the whole continuum of emphasis whereas non-UK sites tended to have a more behavioral emphasis.

Several meta-analyses were conducted on this corpus of data in relation to the various outcomes as well as the targeted outcomes. The outcome under investigation for each meta-analysis was the total symptom score on an appropriate scale, eg, PANSS positive symptom score, PANSS negative symptom score, Global Assessment of Functioning for functioning, Calgary Depression Scale for depression, Brief Social Phobia Scale for social anxiety, and Beck Hopeless Scale for hopelessness/suicidality (see Method for decisions on choice of outcomes).

Meta-analyses of All CBTp Trials in Relation to Various Symptom and Functioning Outcomes

Six separate meta-analyses were carried out on the data using a random-effects model (see Everitt and Pickles[32] and Fleiss[85]) which was applied to the effect sizes obtained from the CBTp trials which reported the various outcomes

- target symptoms (as specified by the research team in their publication),
- positive symptoms,
- negative symptoms,
- functioning,
- mood,
- hopelessness.

Heterogeneity of effect sizes was assumed (and later confirmed for many analyses, see table 3); so a random-effects (RE) model rather than a fixed-effects model was adopted[32,85] (see table 3). Even in the few instances where the test for homogeneity of effect sizes was not significant, the RE model was still used because of concerns about the test's lack of power. The results of the individual meta-analyses are given in table 3. The table also provides the estimated overall effect size, the 95% confidence intervals (CIs), and the usual chi-square test of the homogeneity of the effects. A significant chi-square value indicates heterogeneity in the effect sizes. Not all trials reported all outcomes, and not all trials targeted particular outcomes; so the initial analysis reports the results of a meta-analysis of the indicated target for the study. In the 34 trials, only 33 CBTp trials reported this outcome for the symptom that the trial was targeting. Because there were only 2 studies for the assessment of social anxiety, neither heterogeneity nor CIs are provided.

In terms of the relationships between different outcomes, there are significant correlations between improvements in positive symptoms and a worsening in hopelessness ($R = -0.978$, $P = .022$, $N = 4$), improvements in negative symptoms ($R = 0.830$, $P < .001$, $N = 23$), and some evidence of improved functioning, but this did not quite reach significance ($R = 0.510$, $P = .052$, 2 tailed, $N = 15$). Improvements in negative symptoms were related to improvements in functioning ($R = 0.656$, $P = .021$, $N = 12$). There is also a positive relationship between improvements in functioning and improvements in mood ($R = 0.954$, $P = .003$, $N = 6$). These results suggest that there is a relationship between different outcomes in that targeting one outcome may have positive (and sometimes detrimental) effects on others.

Individual and Group CBTp: Comparison of Outcome in Relation to the Target Symptom

To address this question, a random-effects model was applied separately to the individual CBTp studies (number of studies = 26, total number of participants = 1565) and to the group CBTp studies (number of studies = 7, total number of participants = 399) in relation to the outcome for the target symptom. For the individual CBTp studies, the estimated effect size was 0.415 (SE = 0.08); the corresponding figure for the group CBTp studies was 0.386 (SE = 0.20). The estimated 95% CI for the difference is (−0.384, 0.442) suggesting that there is no evidence of any difference in effect size between individual and group CBTp.

Exploration of Heterogeneity of Outcomes Using Individual CBTp Aimed at Positive Symptoms

There was heterogeneity of effect size noted in the meta-analyses reported above. The purpose of the following analyses is to investigate this heterogeneity in more detail with specific reference to only those studies that indicated

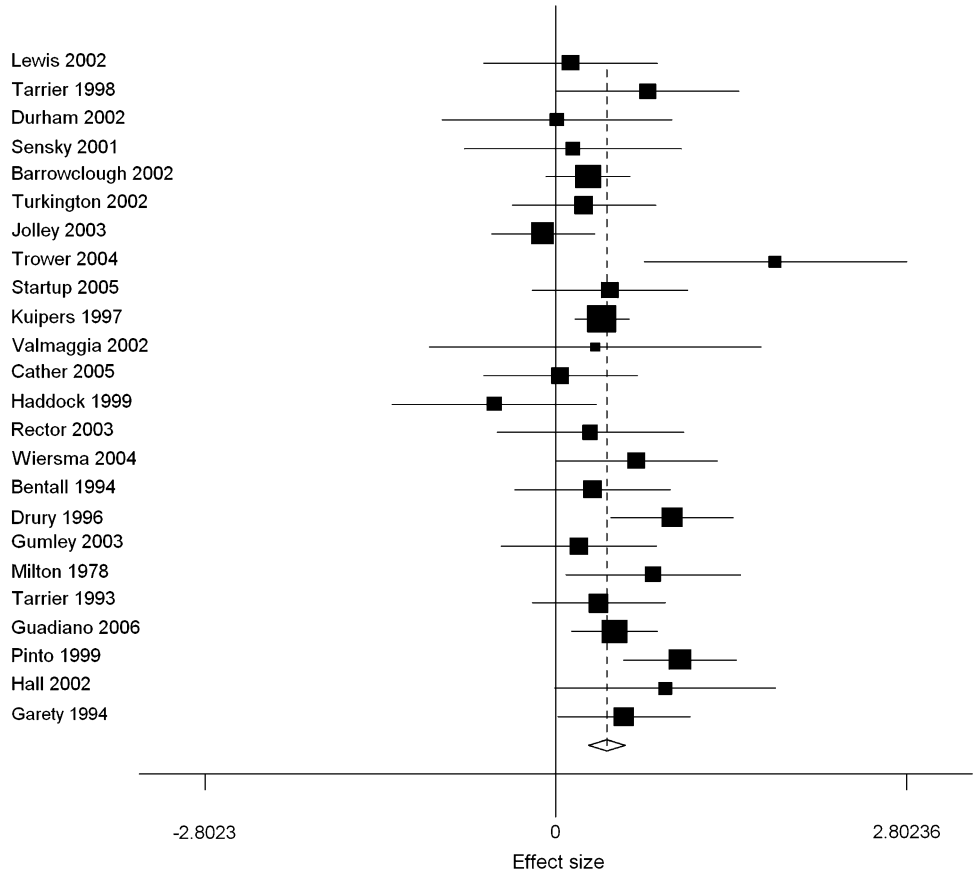

Fig. 1. Forest Plot of the Effect Sizes for the Trials Shown in Table 2.

that the target symptom was positive symptoms. We also chose those that provided CBTp in a more traditional approach by an individual therapist. This has 2 benefits—the studies are homogenous in terms of their mode of treatment provision and treatment target but also report on CBT because it was originally developed as a therapy aimed at residual symptoms. This gave a large sample because the majority of studies were individual CBTp aimed at reducing positive symptoms (number of studies = 24, total number of participants = 1450). These studies were therefore investigated in more detail to understand whether methodological rigor (or lack of it) and clinical emphasis affected the estimate of effectiveness.

Meta-analysis of Individual CBTp Trials Aimed at Positive Symptoms. A random-effects model was applied to the effect sizes obtained from the 24 individual CBTp trials aimed at positive symptoms, leading to an estimated overall effect size of 0.399 with a 95% CI of 0.243, 0.556. The usual chi-square test of the homogeneity of the trials took the value 40.0 with 23 *df*, which is significant at the 5% level. There is evidence of heterogeneity of effect size in the 24 studies. A *forest plot* (see Everitt[86]) of the effect sizes and associated 95% CIs ordered by CTAM score is shown in figure 1.

Relationship Between Methodological Quality, Clinical Emphasis, and Effect Size. To investigate the various relationships, a weighted analysis is necessary because the estimated effect sizes clearly have different precisions and any unweighted analysis ignores this feature of the data. The weight applied to a study was the reciprocal of the sum of the estimated between study variance and the estimated variance of the effect size for the study (see Everitt[86]). The former is found from the random-effects model used in the meta-analysis (see above), and the latter is approximated by the sum of the sample sizes for the experimental and control groups divided by the product of these sample sizes (see Fleiss[85]). Because the Trower et al[53] trial had a distinct focus of intervention (command hallucinations), the results of some analyses were repeated to check the effects of this study on the outcome of the analysis.

Relationship of CTAM and Effect size. The simple correlation was significant whether or not Trower et al[53] study was excluded (Spearman $\rho = -.485$, $P < .02$). However, in a weighted effect size analysis including the Trower et al[53] study, the estimated regression coefficient of effect size on CTAM (found from a weighted regression—see above) was $-.0056$ with 95% CI

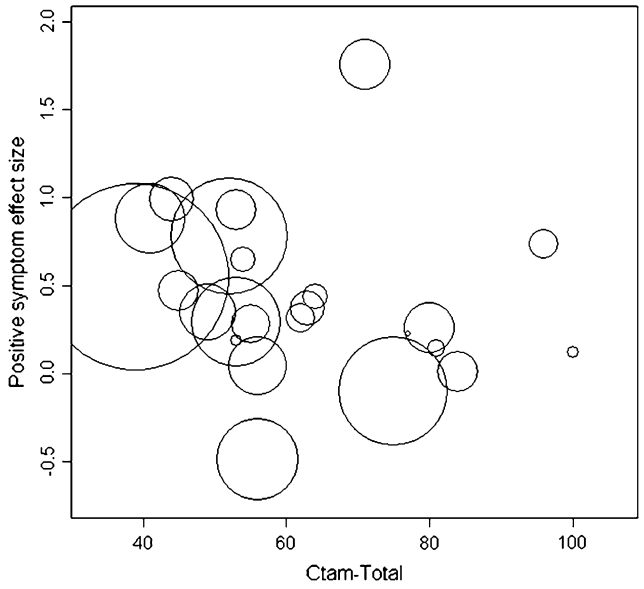

Fig. 2. A Bubbleplot of Effect Size Against Clinical Trial Assessment Measure (CTAM) Score (Radii of the Circles Represent the SE of the Effect Size).

[−0.0152, 0.004]. A bubbleplot (see Everitt[86]) of effect size against CTAM in which the radii of the circles represent the SE of the effect size of a study is shown in figure 2. The tendency is, as expected, for smaller CTAM scores to be associated with larger effect sizes, but the relationship is weak suggesting that other factors not included in the CTAM scale also affect the trial results.

Relationship Between Domains of CTAM and Effect Size. The estimated regression coefficients for the various domains of CTAM and effect size and their estimated CIs are given in table 4. In these analyses, none of the regression coefficients are significant. However, when the Trower et al[53] study was excluded, the relationship of the assessment domain and effect size became significant (the estimated regression coefficient was −.017 (95% CI = −0.032, −0.002). All but one domain is in the expected direction of higher effect sizes with poorer methodology within a particular domain.

Table 4. Relationship Between Methodological Rigor and Effect Size

Domain	Regression Coefficient	95% Confidence Interval
Sample	−.018	(−0.065, 0.039)
Allocate	.002	(−0.058, 0.062)
Assess	−.014	(−0.035, 0.007)
Control	−.013	(−0.056, 0.030)
Analysis	−.020	(−0.057, 0.017)
Treat	−.002	(−0.049, 0.045)

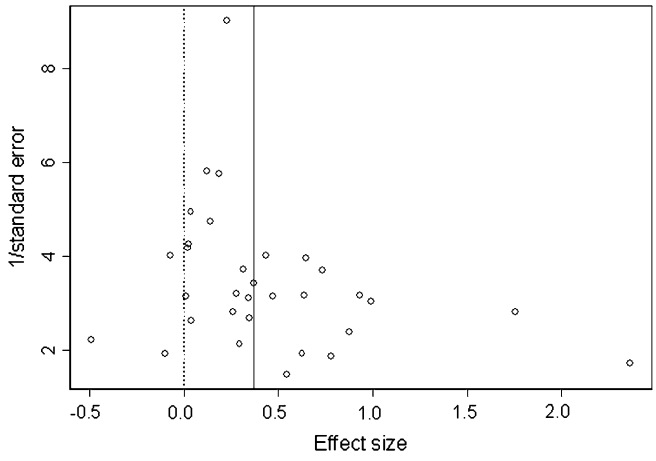

Fig. 3. Funnel Plot.

Masked and Unmasked Assessment. One of the main questions of interest in this study was whether masking of the assessor to the treatment group allocation can lead to different effect sizes because this has been shown for a number of medication studies. There is also some indication of a relationship because the assessment domain includes group masking which was found in the results above. To address the specific question of masking, a random-effects model was applied separately to the 14 masked studies and to the 10 unmasked studies. For the masked studies, the estimated effect size was 0.307 (95% CI = 0.087, 0.527); the corresponding figure for the unmasked studies was 0.492 (95% CI = 0.312, 0.672). There is a tendency for the unmasked studies to be overoptimistic about the effects of CBTp, with effect sizes of 50%–100% higher than those found in masked studies. Even more telling perhaps is the lower end of the 95% CI (0.087, 0.527) for the masked studies.

Relationship Between Clinical Emphasis and Effect Size. The analyses were carried out excluding the Trower et al[53] study because the clinical emphasis in this study of so different. The estimated regression coefficient of .014 (0.0004, 0.028) was significant and suggests that more behavioral treatments produce greater effect sizes.

Publication Bias. Possible publication bias was investigated by a *funnel plot* (effect size against precision) (see figure 3). The absence of studies in the left-hand corner of this plot is usually taken as an indication of possible publication bias. The current plot does not appear to indicate any evidence of a worrying publication bias and so suggests that the estimated effect size found from the random-effects model applied to the 24 studies is realistic.

Table 5. Effect Sizes by Methodological Quality

	Mean Weighted Effect Size	95% Confidence Interval	Heterogeneity Test (*df*), Significance Level	No. of Studies	Sample Size	95% Confidence Interval of Difference
Target symptom						
High CTAM	0.223	0.017, 0.428	27.73 (11), significant at the 5% level	12	1124	0.038, 0.584
Low CTAM	0.534	0.343, 0.725	35.35 (20), significant at the 5% level	21	840	
Positive symptom						
High CTAM	0.222	0.016, 0.427	27.83 (11), significant at the 5% level	12	1124	−0.002, 0.532
Low CTAM	0.487	0.311, 0.664	27.35 (19), not significant	20	794	
Negative symptom						
High CTAM	0.206	−0.104, 0.516	28.33 (8), significant at the 5% level	9	631	−0.100, 0.908
Low CTAM	0.610	0.200, 1.020	83.33 (13), significant at the 5% level	14	637	
Functioning						
High CTAM	0.147	−0.172, 0.466	8.49 (4), not significant	5	347	−0.058, 0.782
Low CTAM	0.509	0.221, 0.797	23.04 (9), significant at the 5% level	10	520	
Mood						
High CTAM	0.084	−0.154, 0.322	9.21 (5), not significant	6	685	0.048, 1.144
Low CTAM	0.680	0.174, 1.186	32.96 (8), significant at the 5% level	9	268	

Note: CTAM, Clinical Trial Assessment Measure.

Relationship Between Methodological Quality and Effect Size in Each of the Outcome Domains

Because there was some relationship between methodological quality and effect size, the outcomes shown in table 3 were investigated in terms of the relationship between studies where the methodology by current standards might be considered adequate. Because there was no specific domain that was poor in all the studies, a cutoff score for the CTAM total of 65 was taken to indicate adequate methodology. This produced 12 studies with adequate methodology and 22 with poorer methodology. The results of the meta-analyses in each of these groups are shown in table 5. For each symptom area, the effect size is larger for the low CTAM studies. This difference is significant for the target symptom and for assessments of mood, and the CIs for the difference is highly skewed for all the other measures. The CIs for the weighted effect sizes in higher CTAM scoring studies are also not significant for negative symptoms, functioning, and mood. However, even when the more stringent criterion is used to define the groups, there are still modest effect sizes for positive symptoms and the targeted symptom.

Discussion

What Variability Is There Between Studies?

This is the largest review of CBTp trials containing 20 more trials than the most comprehensive meta-analysis[21] and 475 more participants. The field has certainly now matured so that there are now nearly 6 times as many patients as in the first reported meta-analysis.[19] The average number of participants in each trial is relatively small (average 66), but this is the same as for drug trials in schizophrenia over the past 35 years.[35]

The measure of trial quality, CTAM, detected methodological variability even in studies that had already passed some methodological inclusion criteria. In fact, in other assessments of quality, these studies would probably have been considered "good research evidence."[25] The methodology scores were not bimodal, and there was also variability in items scoring highly among the trials. Although there was variability, there is some evidence that trials improved their quality and their size over time with larger more rigorously controlled trials appearing more recently. Similar variation has been noted in CBT and suicidal behavior and therapies involving virtual reality although the trials investigated here were of higher quality and were more likely to be large enough to identify clinically significant effects.[87,88]

As well as variability in the methodologies adopted there is also variability in the types of therapy offered under the single umbrella of CBTp. We tried to capture this in our ranking of emphasis, but even if the emphasis was similar, the ingredients used both within a particular study and certainly within a particular participant are likely to be variable (see Tarrier and Wykes for a more detailed discussion[49]). Our independent rating of clinical emphasis was not linked to methodological rigor. As expected, studies from the United Kingdom had higher methodology scores because this site had developed the CBTp study models which then led to the larger and more rigorous studies. Other sites are now building from initial pilot and feasibility studies to more significant RCTs.

Is CBTp Effective?

The obvious further question is, for what outcome? Because CBTp was designed to specifically target positive symptoms, we have used this data set to provide the

answer on efficacy and the simple answer to this question is yes it has modest effects. This is the conclusion of other meta-analyses in this area.[12–16,18–23] The current meta-analysis was methodologically rigorous, for instance, we carried out assessments based on weighted effect sizes that earlier meta-analyses failed to do (eg, Rector and Beck[18] and Gould et al[19]), included all eligible trials that others failed to do (eg, Pilling et al[13]), and considered methodological variability that others failed to do (eg, Pfammatter et al[20]). Although the overall effect size on positive symptoms was lower than assessed by others (eg, Rector and Beck[18] and Gould et al[19]), it was in line with recent evaluations using a smaller data set but similar rigorous methods[21] (effect size = 0.35, no studies = 15). Even when a stringent methodological criterion is adopted a modest significant positive effect size remains.

But perhaps more interesting is that this meta-analysis shows that CBTp may have an effect on other outcomes even if these were not the specific targets of the therapy. The effects for all outcomes were modest, and with the exception of hopelessness (4 studies) the effects were significant overall although these results are more speculative because they seem to be affected by the methodological rigor of the study. However, for hopelessness, not only did this outcome showed homogeneity of effect size but also 3 out of the 4 studies showed negative effects suggesting that current CBTp approaches are not beneficial for this particular outcome and may even be detrimental.

There are also relationships between outcomes such that some outcome improvements seem to be correlated suggesting that irrespective of the actual target, CBTp has wider beneficial impacts. The mode of transmission of these benefits and the timing of such overlapping benefits is not clear and would require further prospective investigation using individual level data.

The mode of provision of CBTp also seems to have little effect with group studies showing the same modest effects. This may make them more cost-effective for particular outcomes and is a mode of presentation that is more highly developed in the United States. However, there are fewer studies testing group therapy and further work in this area is merited to investigate how much improvement is necessary and to reduce the heterogeneity of the effect sizes. Work by both Wykes et al[83] and Barrowclough et al[40] on the clustering of effects within groups provides an analytic strategy for identifying factors that predict good outcomes and therefore the likelihood of increased efficacy of group treatment.

Can We Account for Heterogeneity in Effect Size?

We had thought that by choosing a more homogenous sample of studies, ie, those that provided CBTp individually and which specifically targeted positive symptoms, we would reduce the heterogeneity in the effect sizes, but again although the effects were modest (0.399) there was still significant heterogeneity.

In our subsample, the trials with poorer methodology, as measured by CTAM, tended to have larger effect sizes, although the relationship is relatively weak. Only about 60% of meta-analyses carry out quality assessments of methodology and even fewer (50%) actually link these trial quality assessments to the interpretation of the findings in meta-analyses[89] or specifically to psychological treatment studies. The measure we derived for the quality assessment is reliable and has evidence of both internal and external validity, and although there may be further developments it is clear that the CTAM list provides a start for improving the methodological rigor specifically of psychological treatment trials.

Some assessments of trial quality have been found seriously to affect the assessment of treatment effectiveness in other areas of treatment.[90] However, in our assessments, no specific domain dominated. This may be because of the heterogeneity of the trial methodologies themselves[91] suggesting that some studies are particularly adept at one aspect of methodological rigor and not others. This is hopeful because it suggests that with enough effort all aspects could be improved and high levels of methodological rigor would be possible. Methodological quality also affected the effect sizes for the other outcomes investigated making most not significant. So potential methodological quality variation can (and did) lead to bias and reduction in precision of the estimates of the therapy's effectiveness.[90–92] These results are no different to those that have been found for drug and other medical treatment, and the results reported here are salutary for all psychiatric studies.

The most influential individual methodological variable was masked assessment, which is known to be difficult to carry out in psychological treatment trials. When masked assessment was attempted (we do not know how successful this was), then there was nearly a 60% reduction in the effect size. This compares with the 34% inflation of effect size reported in other studies.[39] In order to ensure true randomization, the random allocation sequence should be irreversibly administered and should be concealed to the individuals in charge of enrolment. Knowledge of possible upcoming allocation may permit selective allocation of patients through some form of manipulation. This might result in those more amenable or more likely to benefit from the treatment being allocated differentially between groups. The effects on subsequent measures are also essential to consider. It is obvious that assessors may well be biased in their assessment, particularly their expectation of positive changes if they know the group allocation. This is a problem not only in the assessment of psychological treatments but in many other fields. In a recent assessment based in Denmark, the majority of the trials

reviewed had inadequate concealment irrespective of whether they used data in the trial protocol or the trial publication.[93]

Other meta-analyses have suggested that CBTp for positive symptoms is more effective for acute populations compared with chronic ones, but these conclusions may also be tempered by overall trial quality. For instance, in the studies identified by Zimmerman et al,[21] the acute studies had only 1 in 3 with adequate methodology (as defined in this article) but for the chronic group 5 out of 10 had adequate methodology. These methodology quality differences are likely to have an impact on the effect size calculations, as was seen in table 5. However, even with the reduction in effect size, the results of the meta-analysis on positive symptoms indicates that for this increased corpus of trials there is a benefit for CBTp on the positive symptoms of schizophrenia although the lower end of the CI suggests only marginal benefits. The general results therefore support other published meta-analyses and systematic reviews that CBTp should be included in guidance on treatments for schizophrenia.

There was a trend for clinical models of CBTp that emphasize more behavioral aspects of treatment to produce larger effect sizes, and this relationship could not be accounted for by differences in methodological rigor. So, studies from all schools of therapy were carried out as well (or as badly) as each other. Our assessment of the level of behavioral emphasis was based only on the published information on the type of therapy approach. It is of course quite possible that every client in any study received a different amount of the therapy and therefore a mixture of the behavioral and more cognitive aspects of the treatment. Assessments of fidelity measure if the therapy complies with what is expected of CBTp but rarely do they measure whether all aspects of the therapy manual were received. If some patients received less of a "dose" of therapy, it might be expected that this would lead to lower clinical effectiveness (see Tarrier and Wykes[49] for a full discussion of the validity issues in CBT studies). In the future, trials of CBTp or any psychological treatment need to include some measure of the "effective dose" of a specific therapy. This is certainly not the same as measures for medication therapies which are simple and generally only refer to the number of days of treatment at a specified level. A sophisticated but simple-to-apply measure for psychological treatments needs to include aspects of the therapy process—eg, how many specific CBTp sessions has the person received. This might be gleaned from a simple process measure collected by the therapist and later tested for reliability.

In addition, we were not able to assess adequately some of the external validation factors, such as treatment generalization, because most studies recruited patients from a convenience sample (ie, through referrals from other agencies) and were relatively small. How participants enter into a trial and the pool from which they were recruited can greatly affect the ease of treatment and potentially the treatment response. Developments in background mental health services over time and the differences between available services between different locations, such as the United Kingdom and United States, can also make comparisons difficult. This is true both between different studies and within a study between experimental and control groups when both receive standard care or TAU. The nature of TAU can be extremely varied. An investigation of this was beyond the scope of this review and must wait until there are larger numbers of studies. We hoped to capture some of the differences by looking at country of origin, but this is not a very clean variable. Participant willingness to enter a trial may also be affected by the nature and accessibility of alternative care from standard services. These are issues which might be expected to potentially influence trial results but are rarely considered or assessed other than in the folklore of the academic community.

While still supporting the use of CBTp in schizophrenia we wish to caution against exaggerated claims of the magnitude of treatment benefit. Many of the studies involved in our meta-analysis were carried out by or under the supervision of the experts in CBTp. There have been some studies showing that after specialist training CBTp can be of value in local service either by training psychiatrists or nurses in key skills. But in order to optimize the treatment effects in local services, it may be necessary to provide specific training in the clinical models that are most successful. There is some indication of a bias toward the use of behavioral features although the specific successful ingredients of the underlying models have not yet been investigated. How effective novel treatments are disseminated into the wider mental health service is a challenge that sometimes results in unanticipated difficulties.[94] The hope that CBTp would become widely available through brief training to less well-qualified and skilled staff is not borne out by experience. But the cost of CBTp being delivered by well-trained and experienced psychologists should not preclude the adoption of CBTp into mental health services. Even if group treatments were an alternative and perhaps cost-effective, these too require experienced and trained personnel to carry them out. The fact that organ transplants can only be carried out by experienced and skilled surgeons is not advanced as a reason to restrict transplant surgery. In mental health services, cheap and ineffective alternatives should not be a viable option to useful, valued, and effective but more expensive ones.

Despite the current acceptance of CBTp in treatment guidance and in some service settings—particularly the United Kingdom—we still need more information if we are to provide CBTp effectively and efficiently. The

concern of CBTp specialists is that research funders will assume that, once the therapy has been accepted, it requires no further investment; so we have set out some key features of a future CBTp research program. It should include

- methodologically rigorous efficacy trials of well-defined treatment programs,
- measures of treatment process which allow an estimate of the dose of treatment that is more sophisticated than "number of sessions,"
- outcome measures that are acceptable not only to the clinical and academic community but also to the patients themselves,
- studies of effectiveness with different models, staff training methods, and background service provision,
- long-term follow-up studies of the durability of treatment effects.

This program will produce a refinement of treatment that has been tested using rigorous methods and is therefore likely to offer the most benefit to people with continuing positive symptoms of schizophrenia.

Acknowledgments

We would wish to thank the authors of all participating studies for their help and assistance and also G. Haddock, N. Rector, A. Gumley, B. Gaudiano, and L. Valmaggia for providing further then unpublished data; R. Gould and K. Mueser for details of their meta-analysis; G. Dunn and T. Johnson for comments on design and statistics; and E. Peters for rating the clinical emphasis of the models of therapy. Both Wykes and Tarrier have carried out CBTp trials.

References

1. Beck AT. Successful out-patient psychotherapy of a chronic schizophrenic with a delusion based on borrowed guilt. *Psychiatry*. 1952;15:205–212.
2. Shapiro MB, Ravenette AT. A preliminary experiment on paranoid delusions. *J Ment Sci*. 1959;105:295–312.
3. Liberman RP, Teigen J, Patterson R, Baker V. Reducing delusional speech in chronic paranoid schizophrenics. *J Appl Behav Anal*. 1973;6:57–64.
4. Meichenbaum D, Cameron R. Training schizophrenics to talk to themselves: a means of developing attentional control. *Behav Ther*. 1973;4:515–534.
5. Brown GW, Birley JLT, Wing JK. Influence of family life on course of schizophrenic disorders—replication. *Br J Psychiatry*. 1972;121:241–258.
6. Butzlaff RL, Hooley JM. Expressed emotion and psychiatric relapse: a meta-analysis. *Arch Gen Psychiatry*. 1998;5: 547–552.
7. Mueser KT, Berenbaum H. Psychodynamic treatment of schizophrenia: is there a future? *Psychol Med*. 1990;20: 253–262.
8. Meltzer HY, Bobo WV. Interpreting the efficacy findings in the CATIE study: what clinicians should know. *CNS Spectr*. 2006;11:14–24.
9. Jones PB, Barnes TR, Davies L, et al. Randomized controlled trial of the effect on quality of life of second- vs first-generation antipsychotic drugs in schizophrenia: Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1). *Arch Gen Psychiatry*. 2006;63:1079–1087.
10. Tarrier N, Calam R. New developments in cognitive-behavioural case formulation. Epidemiological, systemic and social context: an integrative approach. *Cogn Behav Psychother*. 2002;30:311–328.
11. Tarrier N. *Case Formulation in Cognitive Behaviour Therapy: The Treatment of Challenging and Complex Clinical Cases*. London, UK: Routledge; 2006.
12. Cormac I, Jones C, Campbell C, Silveira da Mota Neto J. Cognitive behaviour therapy for schizophrenia (Cochrane Review). *Cochrane Library*. Issue 1Oxford, UK: Update Software; 2003.
13. Pilling S, Bebbington P, Kuipers E, Garety P, Orbach G, Morgan C. Psychological treatments in schizophrenia: I. Meta-analysis of family intervention and cognitive behaviour therapy. *Psychol Med*. 2002;32:763–782.
14. Thornicroft G, Susser E. Evidence-based psychotherapeutic interventions in the community care of schizophrenia. *Br J Psychiatry*. 2001;178:2–4.
15. NHS Centre for Reviews and Dissemination Psychosocial interventions for schizophrenia. *Eff Health Care*. 6. 2000. pp. 1–8.
16. Dickerson FB. Cognitive behavioral psychotherapy for schizophrenia: a review of recent empirical studies. *Schizophr Res*. 2000;43:71–90.
17. Dickerson FB. Update on cognitive behavioural psychotherapy for schizophrenia: review of recent studies. *J Cogn Psychother*. 2004;18:189–205.
18. Rector NA, Beck AT. Cognitive behavioral therapy for schizophrenia: an empirical review. *J Nerv Ment Dis*. 2001;189:278–287.
19. Gould RA, Mueser KT, Bolton E, Mays V, Goff D. Cognitive therapy for psychosis in schizophrenia: an effect size analysis. *Schizophr Res*. 2001;48:335–342.
20. Pfammatter M, Junghan U, Brenner H. Efficacy of psychological therapy in schizophrenia: conclusions from meta-analyses. *Schizophr Bull*. 2006;32:S64–S80.
21. Zimmermann G, Favrod J, Trieu V, Pomini V. The effect of cognitive behavioural treatment on schizophrenia spectrum disorders: a meta-analysis. *Schizophr Res*. 2005; 77:1–9.
22. Sensky T. The effectiveness of cognitive therapy for schizophrenia: what can we learn from the meta-analyses. *Psychother Psychosom*. 2005;74:131–135.
23. Gaudiano BA. Is symptomatic improvement in clinical trials of cognitive-behavioral therapy for psychosis clinically significant? *J Psychiatr Pract*. 2006;12:11–23.
24. *NICE Schizophrenia: Core interventions in the Treatment and Management of Schizophrenia in Primary and Secondary Care*. Available at: http:\\www.nice.org.uk. Accessed October 19, 2007.
25. Lehman AF, Kreyenbuhl J, Buchanan R, et al. The Schizophrenia Patient Outcomes Research Team (PORT): updated treatment recommendations 2003. *Schizophr Bull*. 2004;30: 193–217.
26. Barrowclough C, Haddock G, Tarrier N, et al. Randomized controlled trial of motivational interviewing, cognitive

behaviour therapy, and family intervention of patients with co-morbid schizophrenia and substance use disorder. *Am J Psychiatry*. 2001;158:1706–1713.

27. Richardson A, Baker M, Burns T, Lilford RJ, Muijen M. Reflections on methodological issues in mental health research. *J Ment Health*. 2000;9:463–470.
28. Slade M, Priebe S. Are randomised controlled trials the only gold that glitters? *Br J Psychiatry*. 2001;179:286–287.
29. Pocock SJ. Clinical trials with multiple outcomes: a statistical perspective on their design, analysis, and interpretation. *Control Clin Trials*. 1997;18:530–545.
30. Doll R. Controlled trials: the 1948 watershed. *BMJ*. 1998;317:1217–1220.
31. Everitt BS, Wessely S. *Clinical Trials in Psychiatry*. Oxford, UK: Oxford University Press; 2003.
32. Everitt BS, Pickles A. *Statistical Aspects of the Design and Analysis of Clinical Trials*. London, UK: Imperial College Press; 2000.
33. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA*. 1996;276:637–639.
34. Moher D, Schultz K, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomisation. *JAMA*. 2001;285:1987–1991.
35. Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ*. 1998;317:1181–1184.
36. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*. 1998;352:609–613.
37. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effect in controlled trials. *JAMA*. 1995;280:178–180.
38. Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales; a major source of bias in randomised controlled trials of schizophrenia. *Br J Psychiatry*. 2000;176:249–252.
39. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomised controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials*. 1995;16:62–73.
40. Barrowclough C, Haddock G, Lobban F, et al. Group cognitive-behavioural therapy for schizophrenia: randomised controlled trial. *Br J Psychiatry*. 2006;189:1–7.
41. Kemp R, Kirov G, Everitt B, Hayward P, David A. Randomised controlled trial of compliance therapy. 18-month follow-up. *Br J Psychiatry*. 1998;17:413–419.
42. Chalmers TC, Smith H, Blackburn B, et al. A method for assessing the quality of a randomized control trial. *Control Clin Trials*. 1981;2:31–49.
43. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*. 1996;17:1–12.
44. Juni P, Witschi A, Bloch R, Eggar M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA*. 1999;282:1054–1060.
45. Juni P, Altman G, Eggar M. Assessing the quality of controlled clinical trials. *BMJ*. 2001;323:42–46.
46. Kazdin AE, Bass P. Power to detect differences between alternative treatments in comparative psychotherapy outcome research. *J Consult Clin Psychol*. 1989;57:138–147.
47. Sterne JAC, Juni P, Schulz KF, Altman DG, Barlet C, Eggar M. Statistical methods for assessing the influence of study characteristics on treatment effects in meta-epidemiological research. *Stat Med*. 2002;21:1513–1524.
48. Huwiler-Muntener K, Juni P, Junker C, Eggar M. Quality of reporting of randomised trials as a measure of methodologic quality. *JAMA*. 2002;287:2801–2804.
49. Tarrier N, Wykes T. Is there evidence that cognitive behaviour therapy is an effective treatment for schizophrenia? A cautious or cautionary tale? *Behav Res Ther*. 2004;42:1377–1401.
50. Schulz KF, Altman DG, Moher D. Allocation concealment in clinical trials. *JAMA*. 2002;288:2406–2407.
51. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomised trials: explanation and elaboration. *Ann Intern Med*. 2001;134:663–694.
52. Brown SA. Measurement of quality of primary studies for meta-analysis. *Nurs Res*. 1991;40:352–355.
53. Trower P, Birchwood M, Meaden A, Byrne S, Nelson A, Ross K. Cognitive therapy for command hallucinations: randomised controlled trial. *Br J Psychiatry*. 2004;184:312–320.
54. Milton F, Patwa VK, Hafner RJ. Confrontation vs belief modification in persistently deluded patients. *Br J Med Psychol*. 1978;51:127–130.
55. Tarrier N, Beckett R, Harwood S, Baker A, Yusupoff L, Ugarteburu I. A controlled trial of two cognitive behavioural methods of treating drug-resistant residual psychotic symptoms in schizophrenic patients: I. Outcome. *Br J Psychiatry*. 1993;162:524–532.
56. Garety PA, Kuipers L, Fowler D, Chamberlain F, Dunn G. Cognitive behavioural therapy for drug resistant psychosis. *Br J Psychiatry*. 1994;67:259–271.
57. Bentall RP, Haddock G, Slade PD. Cognitive behaviour therapy for persistent auditory hallucinations: from theory to therapy. *Behav Ther*. 1994;25:51–66.
58. Drury V, Birchwood M, Cochrane R, Manmillan F. Cognitive therapy and recovery from acute psychosis: a controlled trial. I. Impact on psychotic symptoms. *Br J Psychiatry*. 1996;169:593–601.
59. Kuipers E, Garety P, Fowler D, et al. London-East Anglia randomised controlled trial of cognitive-behavioural therapy for psychosis. I: effects of the treatment phase. *Br J Psychiatry*. 1997;171:319–327.
60. Tarrier N, Yusupoff L, Kinney C, et al. A randomised controlled trial of intensive cognitive behaviour therapy for chronic schizophrenia. *BMJ*. 1998;317:303–307.
61. Daniels L. A group cognitive-behavioral and process-oriented approach to treating the social impairment and negative symptoms associated with chronic mental illness. *J Psychother Pract Res*. 1998;7:167–176.
62. Levine J, Barak Y, Granek I. Cognitive group therapy for paranoid schizophrenics: applying cognitive dissonance. *J Cogn Psychother*. 1998;12:3–12.
63. Pinto A, La Pia S, Mennella R, Giorgio D, De Simone L. Cognitive-behavioural therapy and clozapine for clients with treatment-refractory schizophrenia. *Psychiatr Serv*. 1999;50:901–904.
64. Haddock G, Tarrier M, Morrison AP, Hopkins R, Drake R, Lewis S. A pilot study evaluating the effectiveness of individual inpatient cognitive behavioural therapy in early psychosis. *Soc Psychiatry Psychiatr Epidemiol*. 1999;34:254–258.
65. Halperin S, Nathan P, Drummond P, Castle D. A cognitive-behavioural, group-based intervention for social anxiety in schizophrenia. *Aust N Z J Psychiatry*. 2000;34:809–813.

66. Sensky T, Turkington D, Kingdon D, et al. A randomised controlled trial of cognitive-behavioural therapy for persistent symptoms in schizophrenia resistant to medication. *Arch Gen Psychiatry*. 2000;57:165–172.
67. Bailer J, Takats I, Westermeier C. Efficacy of individualized cognitive-behavioral therapy for schizophrenic patients with negative symptoms and social disabilities: a controlled trial. *Z Klin Psychol Psychother*. 2001;30:268–278.
68. Lewis SW, Tarrier N, Haddock G, et al. Randomised controlled trial of cognitive-behaviour therapy in early schizophrenia: acute phase outcomes. *Br J Psychiatry*. 2002;181: suppl 4391–97.
69. Turkington D, Kingdom D, Turner T. The Insight Group.Effectiveness of a brief cognitive behavioural therapy intervention in the treatment of schizophrenia. *Br J Psychiatry*. 2002;180:523–528.
70. Durham RC, Guthrie M, Morton RV, et al. Tayside-Fife clinical trial of cognitive-behavioural therapy for medication-resistant psychotic symptoms: results to 3-month follow-up. *Br J Psychiatry*. 2003;182:303–311.
71. Hall PL, Tarrier M. The cognitive-behavioural treatment of low self-esteem in psychotic patients: a pilot study. *Behav Res Ther*. 2003;41:317–332.
72. Valmaggia L, van der Gaag M, Tarrier N, Pijnenborg GHM, Sloof CJ. A randomized controlled trial of cognitive behavior therapy with treatment refractory positive symptoms of schizophrenia. *Br J Psychiatry*. 2005;186:324–330.
73. Granholm E, McQuaid JR, McClure FS, Pedrelli P, Jeste DV. A randomized controlled pilot study of cognitive behavioural social skills training for older patients with schizophrenia. *Schizophr Res*. 2002;53:167–169.
74. Gumley A, O'Grady M, McNay L, Reilly J, Power KG, Norrie J. Early intervention for relapse in schizophrenia: results of a 12-month randomised controlled trial of cognitive behavioural therapy. *Psychol Med*. 2003;33:419–431.
75. Rector N, Seeman MV, Segal ZV. Cognitive therapy for schizophrenia: a preliminary randomised controlled trial. *Schizophr Res*. 2003;63:1–11.
76. Jolley S, Garety P, Craig T, Dunn G, White J, Aitken M. Cognitive therapy in early psychosis: a pilot randomized controlled trial. *Behav Cogn Psychother*. 2003;31:473–478.
77. Kingsep P, Nathan P, Castle D. Cognitive behavioural group treatment for social anxiety in schizophrenia. *Schizophr Res*. 2003;63:121–129.
78. Wiersma D, Jenner J, Nienhuis F, van de Willige G. Hallucination focused integrative treatment improves quality of life in schizophrenia patients. *Acta Psychiatr Scand*. 2004;109:194–201.
79. Bechdolf A, Knost B, Kuntermann C, et al. A randomized comparison of group cognitive-behavioural therapy and group psychoeducation in patients with schizophrenia. *Psychiatr Scand*. 2004;110:21–28.
80. Startup M, Jackson M, 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;34:413–422.
81. Cather C, Penn D, Otto M, Yovel I, Mueser K, Goff D. A pilot study of functional Cognitive Behavioral Therapy (fCBT) for schizophrenia. *Schizophr Res*. 2005;74:201–209.
82. Granholm E, McQuaid JR, McClure FS, et al. A randomized, controlled trial of cognitive behavioral social skills training for middle-aged and older outpatients with chronic schizophrenia. *Am J Psychiatry*. 2005;162:520–529.
83. Wykes T, Hayward P, Thomas N, et al. What are the effects of group cognitive behaviour therapy for voices? A randomised control trial. *Schizophr Res*. 2005;77:201–210.
84. Gaudiano B, Herbert J. Acute treatment of inpatients with psycholtic symptoms using acceptance and commitment therapy: pilot results. *Behav Res Ther*. 2006;44:415–437.
85. Fleiss JL. Measures of effect size for categorical data. In: Cooper H, Hedges LV, eds. *The Handbook of Research Synthesis.* New York, NY: Russell Sage Foundation; 1993:245–281.
86. Everitt BS. *Modern Medical Statistics: A Practical Guide.* London, UK: Arnold; 2002.
87. Tarrier N, Taylor K, Gooding P. Cognitive behavioural interventions to reduce suicidal behaviour: a systematic review and meta-analysis. *Behav Modif*. In press.
88. Gregg L, Tarrier N. Virtual reality in mental health: a review of the literature. *Soc Psychiatry Psychiatr Epidemiol*. 2007;42:343–354.
89. Moja L, Telaro E, D'Amico R, Moschetti I, Coe L, Liberati A. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. *BMJ*. 2005;330:1053.
90. Balk E, Bonis P, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA*. 2002; 287:2973–2982.
91. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA. Incorporating variations in the trial quality of individual randomised trial into meta-analysis. *J Clin Epidemiol*. 1992;45:255–265.
92. Gotzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials*. 1989;10:31–56.
93. Pildal J, Chan A-W, Hróbjartsson A, Forfang E, Altman D, Gøtzsche P. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. *BMJ*. 2005;330:1049.
94. Tarrier N, Barrowclough C, Haddock G, McGovern J. The dissemination of innovative cognitive-behavioural treatments for schizophrenia. *J Ment Health*. 1999;8:569–582.