Literature: A Manual for Evidence-Based Clinical Practice (3rd edn). (eds Guyatt G, Rennie D, Meade MO & Cook DJ), pp. 103-106. New York: The McGraw-Hill Companies, Inc.

コクランレビュー

16) Imai H, Tajika A, Chen P, Pompoli A, Guaiana G, Castellazzi M, Bighelli I, Girlanda F, Barbui C, Koesters M, Cipriani A & Furukawa TA (2014) Azapirone versus placebo for panic disorder in adults. **Cochrane Database of Systematic Reviews**, CD010828.
17) Purgato M, Papola D, Gastaldon C, Trespidi C, Magni LR, Rizzo C, Furukawa TA, Watanabe N, Cipriani A & Barbui C (2014) Paroxetine versus other anti-depressive agents for depression. **Cochrane Database of Systematic Reviews**, CD006531.

コクランレビュープロトコル

18) Imai H, Tajika A, Chen L, Pompoli A & Furukawa TA (2014) Psychological therapies versus pharmacological interventions for panic disorder [Protocol]. **Cochrane Database of Systematic Reviews**, CD011170.
19) Kimachi M, Furukawa TA, Kimachi K, Goto Y & Fukuhara S (2014) New oral anticoagulants versus warfarin for preventing stroke and systemic embolic events among atrial fibrillation patients with chronic kidney disease [Protocol]. **Cochrane Database of Systematic Reviews**.
20) Katsura M, Kuriyama A, Takeshima T, Fukuhara S & Furukawa TA (2014) Preoperative inspiratory muscle training for postoperative pulmonary complications in adult patients undergoing cardiac and major abdominal surgery [Protocol]. **Cochrane Database of Systematic Reviews**.
21) Onishi A, Sato A, Iwasaku M & Furukawa TA (2014) Human parathyroid hormone for treating glucocorticoid-induced osteoporosis [Protocol]. **Cochrane Database of Systematic Reviews**.
22) Pompoli A, Furukawa TA, Imai H, Tajika A, Efthimiou O & Salanti G (2014) Psychological therapies for panic disorder with or without agoraphobia in adults [Protocol]. **Cochrane Database of Systematic Reviews**, CD011004.
23) Sakai K, Fujita K, Sozu T, Nakayama T & Furukawa TA (2015) Eradication of helicobacter pylori for iron metabolism [Protocol]. **Cochrane Database of Systematic Reviews**, CD011480.
24) Takeshima N, Furukawa TA, Hayasaka Y, Ogawa Y & Tajika A (2014) Ramelteon for primary insomnia [Protocol]. **Cochrane Database of Systematic Reviews**, CD011049.

G. 知的財産権の出願・登録状況

1. 特許取得　なし
2. 実用新案登録　なし
3. その他　なし

Acta Psychiatrica Scandinavica

Acta Psychiatr Scand 2014: 130: 181–192
All rights reserved
DOI: 10.1111/acps.12275

ACTA PSYCHIATRICA SCANDINAVICA

Meta-analysis

Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis

Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, Chen P, Hunot V, Churchill R. Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis.

Objective: Various control conditions have been employed in psychotherapy trials, but there is growing suspicion that they may lead to different effect size estimates. The present study aims to examine the differences among control conditions including waiting list (WL), no treatment (NT) and psychological *placebo* (PP).
Method: We comprehensively searched for all randomized controlled trials (RCTs) comparing cognitive-behaviour therapies (CBT) against various control conditions in the acute phase treatment of depression, and applied network meta-analysis (NMA) to combine all direct and indirect comparisons among the treatment and control arms.
Results: We identified 49 RCTs (2730 participants) comparing WL, NT, PP and CBT. This network of evidence was consistent, and the effect size estimates for CBT were substantively different depending on the control condition. The odds ratio of response for NT over WL was statistically significant at 2.9 (95% CI: 1.3–5.7). However, the quality of evidence, including publication bias, was less than ideal and none of the preplanned sensitivity analyses limiting to high-quality studies could be conducted, while findings of significant differences did not persist in *post hoc* sensitivity analyses trying to adjust for publication bias.
Conclusion: There may be important differences in control conditions currently used in psychotherapy trials.

T. A. Furukawa[1], H. Noma[2], D. M. Caldwell[3], M. Honyashiki[4], K. Shinohara[4], H. Imai[4], P. Chen[4], V. Hunot[5], R. Churchill[5]

[1]Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, [2]Department of Data Science, Institute of Statistical Mathematics, Tokyo, Japan, [3]School of Social and Community Medicine, University of Bristol, Bristol, UK, [4]Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan and [5]Academic Unit of Psychiatry, School of Social and Community Medicine, University of Bristol, Bristol, UK

Key words: waiting lists; *placebo*; control groups; cognitive therapy; clinical trials

Toshi A. Furukawa, Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501 Japan. E-mail: furukawa@kuhp.kyoto-u.ac.jp

Accepted for publication March 12, 2014

Summations

- We identified a dense, well-connected, homogeneous and consistent network of evidence connecting cognitive-behaviour therapies (CBT), psychological *placebo* (PP), no treatment (NT) and waiting list (WL) in the acute phase treatment of adult depression.
- Applying the network meta-analysis (NMA), the effect size estimates for CBT were substantively different, depending on which control condition it was compared against.
- The indirect comparison between NT and WL revealed that the former was significantly superior to the latter in producing response (OR = 2.9, 95% confidence interval: 1.3–5.7).

Considerations

- The quality of randomized evidence constituting the above network was less than ideal.
- Statistically significant differences in the primary analyses were lost when we applied exploratory sensitivity analyses using network meta-regression to adjust for publication bias.

Introduction

The need for a control condition in evaluating efficacy/effectiveness of any health intervention is accepted beyond doubt but what constitutes an appropriate control in psychotherapy trials has long been hotly debated (1, 2).

The level of control required is dependent on what factors one wishes to control in trials. Maximally, we would like to control for the regression towards the mean, the natural course of the disease, the Hawthorne effect (the effect of being observed and evaluated), and the *placebo* effect (the effect of believing to be treated for benefit). In pharmacotherapy trials, the pill *placebo* control would control for all these factors, and it therefore follows that, if a certain chemical compound turns out to be superior to the pill *placebo* in a comparative trial that compound will be believed to have a specific efficacy for the disorder. Following this model, earlier theorists have advocated for a psychological *placebo* (PP), which would similarly control for the four factors above in psychotherapy trials (3). In reality, such has not been easy to implement and many psychotherapy trials have relied on less than ideal control conditions, including treatment as usual (TAU), no treatment (NT) and waiting list (WL). Many systematic reviews have traditionally lumped all these into one control condition in evaluating psychotherapies (4–6) or various human interventions (7). TAU, sometimes also referred to as usual care (UC), is simply too heterogeneous across studies and countries to be regarded one entity in a review (8). We therefore did not include the so-called TAU or UC conditions in the present review. However, even aside from them, there are now growing theoretical and empirical concerns that different control conditions may lead to different effect size estimates (9–12).

When several alternative interventions or controls are compared in a series of randomized trials, we now have an advanced method of evidence synthesis called network meta-analysis (NMA). Traditional, head-to-head meta-analyses can compare only two interventions at a time. NMA allows assessment of relative effectiveness of all the included interventions by integrating data from both direct and indirect comparisons, thus borrowing strength from the entire network of randomized trials (13, 14). NMA is therefore particularly suitable for the assessment of differences among alternative control conditions, each of which is compared against the active treatment but only a few of which are compared directly among themselves.

Aim of the study

The present study aims to differentially examine no treatment, waiting list and psychological *placebo* control conditions by applying network meta-analysis to the relevant body of randomized controlled trials of cognitive-behaviour therapies for adult depression.

Material and methods

Study eligibility

The eligibility criteria for the studies were as follows.

Study design. All relevant randomized controlled studies. Studies in which two relevant conditions were compared and in which a cointervention (except for protocolized antidepressant treatment intended as combination treatment) was simultaneously provided were accepted when the cointervention was equally administered in both arms. Quasi-randomized controlled trials, in which treatment assignment was decided through methods such as alternate days of the week, were excluded.

Participants. Patients between ages 18 and 75, of both sexes, with a primary diagnosis of acute phase unipolar depression,

i) Diagnosed as such according to any of the following operationalized criteria: Feighner criteria, Research Diagnostic Criteria, DSM-III, DSM-III-R, DSM-IV or ICD-10 or
ii) Identified as such using a validated clinician or self-report depression symptom questionnaire, based on a recognized threshold.

Because differences in cognitive capabilities among younger or elderly people may influence the effectiveness and acceptability of psychotherapies (11, 15), we excluded studies of children and adolescents aged ≤18 or of older people where the mean age of participants was ≥75.

Studies focusing on patients with chronic depression or treatment-resistant depression were excluded. Studies that focused on depression among participants who all had a concurrent primary diagnosis of another mental or physical disorder were excluded. Existence of concurrent secondary diagnosis of another disorder was allowed.

Interventions. The control conditions of interest in this study included the following:

i) PP: a control condition that was regarded as inactive by the researchers but was to be

perceived as active by the participants. Furthermore, the number and duration of the sessions as well as the qualification of the therapists had to be equivalent with those of the active treatment in the same study.

ii) NT: a control condition in which the participants receive no active treatment during the study and in which they do not expect to receive such after the study is over.

iii) WL: a control condition in which the participants receive no active treatment during the study but are forewarned that they can receive one after the study period is over.

In both NT and WL, participants may or may not receive some medical care during the course of the study on a naturalistic basis.

The active intervention chosen to be contrasted in this study was CBT, because it is by far the best studied school of psychotherapy for depression (16) and was therefore likely to form the densest network in differentiating various control conditions. In this study, CBT was broadly conceived as comprising one or more of the following cognitive and behavioural components.

i) Cognitive restructuring, which aims at monitoring and modifying the patient's dysfunctional beliefs that lead to depressed feelings.

ii) Behavioural activation, which aims at helping the patient increase his/her frequency and quality of pleasant activities and/or goal-directed activities.

iii) Problem solving, which teaches the patient a staged and structured approach to pragmatic solution of his/her problems.

iv) Assertion training, which helps the patient initiate and negotiate interpersonal relationships by teaching him/her the skill to express honest feelings and requests.

v) Acceptance and mindfulness, which places less emphasis on rational challenging of thoughts but aims to transform the relationship between the experience of symptoms and difficult thoughts/feelings.

CBT had to be delivered through face-to-face meetings between the patient and therapist. Both individual and group formats were accepted.

Outcomes. Our primary outcome was the number of patients who responded to treatment, based on changes on Hamilton Rating Scale for Depression (HAMD) (17), Beck Depression Inventory (BDI) (18) or any other validated depression scale at the end of the acute phase treatment. Acute treatment was defined as an 8-week treatment in the analyses. If 8-week data were not available, we used data ranging between 4–16 weeks and the time point given in the original study as the study endpoint was given the preference. Many studies defined response by 50% or greater reduction on the rating scale, but some studies defined it using Jacobson's Reliable Change Index (19). In the latter case, we adopted the original study authors' definitions. If the original authors reported several outcomes, observer-rated scales were preferred to self-reported scales; among the former, HAMD and among the latter, BDI was given preference.

Intention-to-treat analyses were based on the total number of randomly assigned participants, irrespective of how the original study investigators analysed the data, by assuming all drop-outs to be non-responders. For studies in which the exact numbers of participants who had responded were not reported, but means and standard deviations for continuous depression scales were reported, the number of responders was calculated by using a validated imputation method (20, 21).

In a few instances where we were unable to obtain the number of participants randomized at baseline (e.g. the number of total drop-outs after initial randomization is reported but no separate number of drop-outs on each treatment arm is provided nor obtainable from the original authors), we still included such studies using the analysed numbers of participants.

For three- or more-armed trials in which a control condition was compared against two or more variants of CBT (e.g. cognitive therapy and problem solving), all CBT arms were collapsed into one group.

Data collection

To identify the relevant studies, we searched two clinical trials registries created and maintained by the Cochrane Depression, Anxiety and Neurosis Group (CCDAN), the CCDANCTR-Studies and CCDANCTR-References. Reports of trials for inclusion in the Group's registers are collated from weekly, generic searches of MEDLINE (1950–), EMBASE (1974–) and PSYCINFO (1967–) and quarterly searches of the Cochrane Central Register of Controlled Trials. Reports of trials are also sourced from the World Health Organization's trials portal (ICTRP), clinicaltrials.gov, drug companies' websites, the hand searching of key journals, dissertation abstracts, conference proceedings and other non-Cochrane systematic reviews and meta-analyses. Details of CCDAN's generic search strategies can be found in the Cochrane Collaboration Depression, Anxiety and Neurosis

Group's webpage (http://ccdan.cochrane.org/). The CCDANCTR-Studies register was searched using the following terms: Condition = (depress* or dysthymi*) and Intervention = (*therap* or training). This search was supplemented by corresponding searches in CCDAN-References register, CINAHL and PSYNDEX. The additional search strategies for the databases other than the Studies register can be found in Shinohara et al. (22). The reference lists of all selected studies were searched for additional published reports and citations to unpublished studies. Relevant review papers were checked. The most recent updated search for this review was done in February 2012.

Two review authors (RC, VH) examined the abstracts of all publications obtained through the search strategy. Full articles of all the studies identified by any of these review authors, were obtained. Conflicts of opinion regarding eligibility of a study were discussed with a third review author, having retrieved the full paper and consulted the authors if necessary, until consensus was reached. External subject or methodological experts were consulted if necessary.

Data from each study were extracted independently by at least three review authors. Any disagreement was discussed with an additional review author and where necessary, the original study authors were contacted for further information. Information relating to study population, interventions, comparators, potential biases in the conduct of the trial and outcomes were abstracted from the original reports into specially designed paper forms then double-entered into a spreadsheet.

Assessment of risk of bias

Two independent review authors assessed the risk of bias in the selected studies, using the Cochrane Collaboration 'risk of bias' tool (23). The following seven domains were considered:

i) Sequence generation
ii) Allocation concealment
iii) Blinding of therapist
iv) Blinding of participant
v) Blinding of assessor for primary outcome
vi) Incomplete outcome data
vii) Selective outcome reporting

In addition, the following risks of bias specific to psychotherapy trials were systematically appraised.

viii) Researcher allegiance/Conflict of interest: did the researcher(s) have vested interest for or against the therapies under examination?
ix) Treatment fidelity: was the therapy monitored against a manual or a scale through audio- or videotapes?
x) Therapist qualification: were the therapists qualified to deliver psychotherapy and have they had specialist training for the intervention they are providing?
xi) Therapist allegiance/conflict of interest: did the therapists have vested interest for or against the therapies they were providing?
xii) Other sources of bias: was the study apparently free of other problems that could put it at a high risk of bias?

Any disagreement was discussed with a third review author. Where necessary, the authors of the studies were contacted for further information.

Analyses

Individual trials constitute pairwise, head-to-head meta-analyses, which then constitute the network for the NMA. In the following, we therefore first examined the robustness of each pairwise meta-analysis that forms a part of the evidence network in this study. We then proceeded to examine the robustness of the entire NMA. Given the potential clinical heterogeneity of the populations studied as well as the CBT approaches used in the studies, we used the random-effects model in all analyses.

Pairwise meta-analyses. We conducted pairwise meta-analyses by synthesizing studies that compared the same intervention/control condition with a random-effects model to incorporate the assumption that different studies assessed different, yet related, treatment effects (24). We examined between-studies heterogeneity through visual inspection of the forest plots and also by the I^2 statistics (23). We followed the following rough guide to interpretation of I^2 statistics:

0–40%: might not be important;
30–60%: may represent moderate heterogeneity;
50–90%: may represent substantial heterogeneity;
75–100%: considerable heterogeneity.

We also drew funnel plots and used Egger's tests (25) to assess small study effects, including possible publication bias. We conducted these analyses using R version 2.15.2 and metafor version 1.6 (26).

Network meta-analyses. We implemented random-effects network meta-analyses, taking into account the heterogeneity of treatment effects across studies in the Bayesian framework using OPENBUGS 3.2.1. We modelled the binary outcomes in every

treatment group of every study using the logistic regression model, incorporating the heterogeneity across studies by random effects. We evaluated the comparative response rates of CBT and the three control conditions. We assessed the Bayesian estimates and 95% credible intervals (CIs) of the odds ratios, and evaluated significance using the CIs (according to whether the CI included the null values).

A key assumption of the NMA model is the consistency of the network, that is, that direct and indirect evidence on the same comparisons does not disagree beyond random errors. For all triangles contained in the network, the difference between the direct and indirect estimates was examined by Bucher's test of inconsistency (27). We also evaluated model fit of the consistent and inconsistent models using the residual deviance statistics and deviance information criterion (DIC) (28). In a well-fitting model, the residual deviance should be close to the number of data points, and the model fit should be better for the consistency model than for the inconsistency one in terms of the residual deviance as well as DIC.

Sensitivity analyses. We *a priori* had selected allocation concealment, assessor blinding, treatment fidelity and imputation of numbers of responders as potentially important effect modifiers to be examined in sensitivity analyses to limit the included studies to those at low risk of bias. We conducted additional meta-regression analyses using random-effects network meta-regression models (29) to examine potential effect moderators such as the mean age of participants, the type of rating scales (clinician-rated vs. self-rated), publication status (published vs. dissertation) and therapy format (individual vs. group). The effect of assuming all drop-outs to be non-responders was examined by the completers analysis. In addition, leave-one-out analyses (i.e. omitting one study at a time) were performed to evaluate the influence of individual studies to the overall NMA results.

When substantive small study effects were noted, we conducted meta-regression analyses to evaluate their influences involving the study-specific variances as a covariate (30).

Results

Study selection

Of 6710 studies identified through electronic search and reference search, 186 full text articles were retrieved, of which 57 studies satisfied the eligibility criteria for the present study (Fig. 1). Of these, we were unable to include eight studies in the present study, as there was critical information lacking in the original reports and unavailable from the original authors upon request (31–38).

Thus, we were able to include 49 randomized studies, representing 117 treatment arms and involving 2730 participants. Figure 2 shows the final evidence network. There were nine studies comparing CBT against PP, 14 studies comparing CBT against NT, and 28 studies comparing CBT against WL. One study compared CBT, PP and NT (39), and another compared CBT, PP and WL (40), thus contributing one study each to the comparison PP vs. NT and to the comparison PP vs. WL, respectively.

The characteristics of the included studies are listed in Table 1. The mean drop-out rates were 19.5% on CBT, 25.5% on PP, 35.0% on NT and 10.8% on WL. The risks of bias of all the included studies are illustrated in Fig. 3.

Pairwise meta-analyses

Table 2 summarizes the results of pairwise meta-analyses. In comparison with the comparator arm CBT, PP tended to be inferior (OR = 1.60, 95% CI: 0.95–2.67), NT was definitely inferior (2.07, 1.35–3.18) and WL gravely inferior (OR = 3.99, 2.76–5.77). Two single studies comparing NT and WL against the common comparator PP were in line with the above pairwise meta-analyses, showing a larger effect size for PP against WL than against NT.

Though not definitive, because the 95% CI were wide, the I^2 statistics hinted at low to moderate heterogeneity for all these pairwise meta-analyses. The funnel plot analysis suggested strong small study effects in the comparison CBT vs. WL but not in the comparisons CBT vs. NT or CBT vs. PP.

Network meta-analysis

The baseline characteristics of included studies were similarly distributed across the comparisons and consistency was therefore likely to hold across the network. There were two closed loops in our network (Fig. 2), neither of which was suggestive of inconsistency beyond chance according to Bucher's test ($P = 0.60$ for CBT-PP-NT, and $P = 0.49$ for CBT-PP-WL).

There was little to choose between the inconsistency and consistency models both in terms of residual deviance and DIC (Residual deviance and DIC were 108.5 and 181.3 for consistency model and 108.4 and 181.3 for inconsistency model,

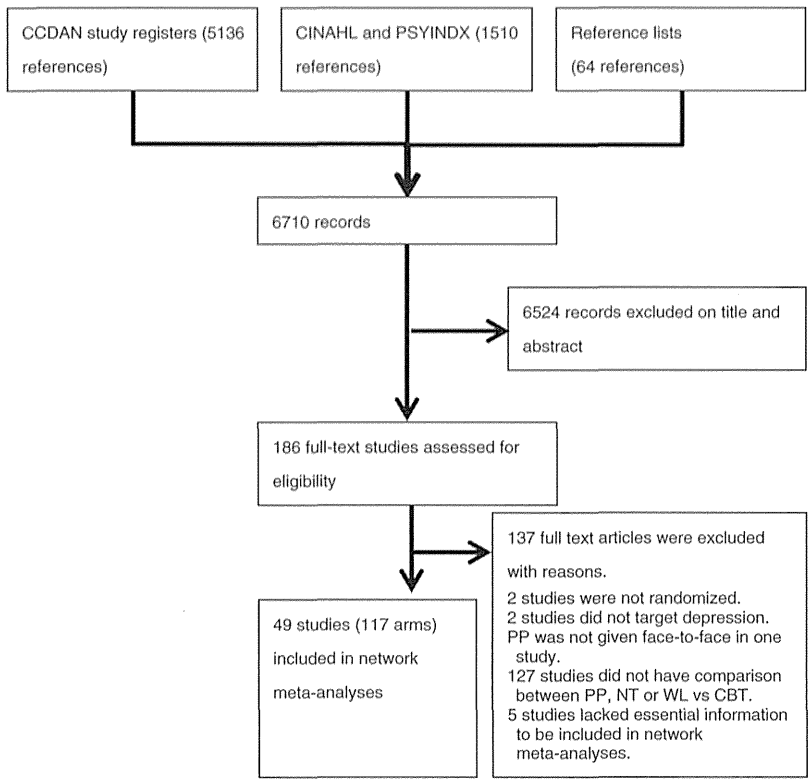

Fig. 1. PRISMA flowchart for the identification of relevant trials.

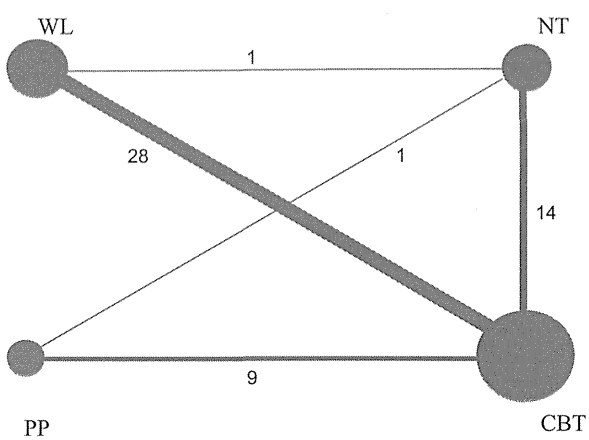

Fig. 2. Evidence network constituted by the 52 included comparisons.

respectively, for 98 data points). Owing to the way in which the residual deviance is calculated, zero cells (in which the number of responders is zero), can cause computational difficulties. We therefore further explored model fit by (i) excluding trials with zero cells and (ii) by applying a continuity correction of 0.5 to 0 cells. Given the reasonable concordance between the residual deviance and the number of data points when we excluded zero cells (83.0 for 82 data points) and applied continuity correction (102.1 for 98 data points), in conjunction with the results of Bucher's test above, the consistency model was judged to have satisfactory model fit. These results are shown in Table 3.

Sensitivity analyses

We were unable to conduct any of the preplanned sensitivity analyses, as there were only up to a quarter of studies at low risk of bias for each of them (14 for allocation concealment, 7 for assessor blinding, 12 for treatment fidelity and 4 for imputation). None of the regression coefficients of the meta-regression examining possible effect moderators turned out to be statistically significant [−0.024 (95% CI: −0.056 to 0.006) for age, −0.899 (−1.843 to 0.024) for rating scale, −0.442 (−1.399 to 0.520) for publication status, and 0.004 (−0.798 to 0.762) for therapy format].

The results of the NMA based on the completers only were essentially similar to our primary results. For example, the odds ratio of response for NT over WL was statistically significant at 2.6 (95% CI: 1.2–5.9). Also, all of the leave-one-out estimates were distributed around the overall NMA estimates with narrow ranges, suggesting that the

Table 1. Characteristics of the included studies

Study	No of participants	Included disorders	Baseline BDI	Types of CBT	Control conditions	Format	No of sessions	Concomitant pharmaco-therapy	Outcome scale
Arean (1993) (44)	48	MDD	22.5	PS	WL	G	12	Not allowed	HAMD
Ayen (2004) (45)	22	MDD+	22.0	Other CBT	WL	G	12	Allowed	BDI
Barrera (1979) (46)	20	Depression	25.1	BA	WL	G	5	Unclear	MMPI-D
Besyner (1979) (47)	20	Depression	24.9	CT	PP	G	4	Unclear	BDI
Brown (1984) (48)	63*	MDD+	31.8	CWD	WL	G/I	12	Allowed	BDI
Carrington (1979) (49)	20*	Depression	26.9	CT	WL	I	12	Unclear	BDI
Collins (1996) (50)	89	MDD	23.4	CT	WL	G	12	Allowed	BDI
Dowrick (Rural Finland) (1996)† (51)	50	MDD+	21.1	PS	NT	I	6	Allowed	BDI
Dowrick (Urban Finland) (1996)† (51)	47	MDD+	21.3	PS	NT	I	6	Allowed	BDI
Dowrick (Ireland, Urban&Rural) (1996)† (51)	38	MDD+	23	CWD	NT	G	8	Allowed	BDI
Dowrick (Rural Norway) (1996)† (51)	61	MDD+	19.2	CWD	NT	G	8	Allowed	BDI
Dowrick (Urban Norway) (1996)† (51)	67	MDD+	21	CWD	NT	G	8	Allowed	BDI
Dowrick (Urban Spain) (1996)† (51)	30	MDD+	22	PS	NT	I	6	Allowed	BDI
Dowrick (Rural UK) (1996)† (51)	49	MDD+	26	PS	NT	I	6	Allowed	BDI
Dowrick (Urban UK) (1996)† (51)	84	MDD+	24.8	PS, CWD	NT	I/G	6/8	Allowed	BDI
Embling (2002) (52)	38	MDD	31.0	CT	WL	G	12	Allowed	BDI
Epstein (1987) (53)	22	MDD+	25.8	CT	WL	G	8	Not allowed	BDI
Faramarzi (2008) (54)	82	Depression	19.9	CT	NT	G	10	Not allowed	BDI
Fuchs (1977) (40)	28*	Depression	22.8	Other CBT	PP, WL	G	6	Unclear	BDI
Hamamci (2006) (55)	24	Depression	28.4	CT	NT	G	11	Not allowed	BDI
Hamdan-Mansour (2009) (56)	84	Depression	24.1	Other CBT	NT	G	10	Unclear	BDI
Hautzinger (2004) (57)	100	MDD+	NA	CT	WL	G	12	Allowed	HSCL-D
Hayman (1980) (58)	28	Depression	18.5	AT	WL	G	8	Not allowed	BDI
Hegerl (2010) (59)	120	MDD+	NA	CT	PP	G	10	Allowed	HAMD
Hess-Homeier (1981) (60)	14	Depression	24.5	CT	WL	I	24	Not allowed	BDI
Kelly (1982) (61)	24	MDD	25.1	BA, REBT	PP	G	6	Allowed	BDI
Malouf (1984) (62)	53	Depression	20.6	PS, REBT	WL	G	4	Unclear	BDI
Miranda (2003) (63)	179	MDD	NA	CT	NT	I/G	8	Not allowed	HAMD
Nezu (1986) (64)	21	MDD	22.5	PS	WL	G	8	Not allowed	BDI
Nezu (1989) (65)	43	MDD	27.0	PS	WL	G	10	Not allowed	HAMD
Pace (1977) (66)	16	Depression	22.3	BA	PP	I	7	Unclear	BDI
Pace (1993) (67)	99	Depression	17.0	CT	WL	I	7	Not allowed	BDI
Pecheur (1980) (68)	21	MDD	22.7	CT	WL	G	8	Not allowed	HAMD
Pellowe (2006) (69)	54	Depression	13.8	ACT	PP	G	4	Allowed	BDI
Propst (1980) (39)	47	Depression	15.4	Other CBT	PP, NT	G	8	Not allowed	BDI
Propst (1992) (70)	49*	MDD+	17.3	CT, other CBT	WL	I	18	Not allowed	HAMD
Ross (1985) (71)	67	MDD	27.8	CT	WL	I/G	12	Allowed	BDI
Schmidt (1983) (72)	44	Depression	24.9	Other CBT	WL	I/G	8	Unclear	BDI
Schmitt (1988) (73)	40	MDD	27.1	PS, AT	WL	G	12	Unclear	HAMD
Serfaty (2009) (74)	137	MDD+	26.8	CT	PP	I	12	Allowed	BDI
Taylor (1977) (75)	28	Depression	21.2	CT, BA, other CBT	WL	I	6	Not allowed	BDI
Usaf (1990) (76)	60	MDD	27.5	CWD	WL	G	10	Not allowed	BDI
Wilson (1982) (77)	21*	Depression	26.0	BA	PP	I	7	Unclear	BDI
Wilson (1983) (78)	25	Depression	24.0	BA, CT	WL	I	8	Not allowed	HAMD
Wollersheim (1991) (79)	16	MDD	25.8	CBT other	WL	G	10	Unclear	BDI
Wong (2008a) (80)	337	MDD	20	CT	WL	G	10	Allowed	BDI
Wong (2008b) (81)	96	MDD	23.9	CT	WL	G	10	Allowed	BDI
Wright (2000) (82)	45	MDD	29.7	CT, CBT other	WL	I	9	Not allowed	HAMD

PP, psychological placebo; NT, no treatment; WL, waiting list; CBT, cognitive-behaviour therapy; MDD, major depressive disorder diagnosed by operationalized diagnostic criteria; MDD+, major depressive disorder and depressive disorder NOS as diagnosed; ACT, acceptance and commitment therapy, AT, assertion training; BA, behavioural activation; CT, cognitive therapy; CWD, coping with depression course; PS, problem solving; REBT, rational emotive-behaviour therapy; G, group; I, individual; BDI, beck depression inventory; HAMD, Hamilton Rating Scale for depression; HSCL-D, Hopkins Symptom Checklist Depression Scale; MMPI-D, Minnesota Multiphasic Personality Inventory Depression Scale by operationalized diagnostic criteria.

*For these studies, randomized N was not available. Instead, we used number of participants analysed.

†Dowrick (1996) ENREF_41 reports nine independently conducted, albeit according to concerted protocols, RCTs. Two of these RCTs conducted in Ireland were reported in an amalgamated form in the definitive report (83) and is therefore treated as one trial in this meta-analysis.

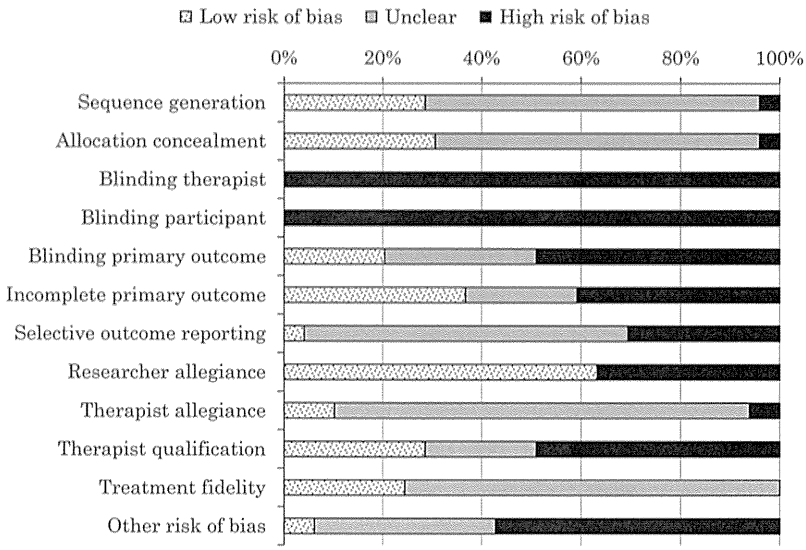

Fig. 3. Risks of bias of the 52 included trials.

Table 2. Results of pairwise meta-analyses

	Number of studies	Number of participants	OR	*P*	I^2, % (%)	Egger's test *P*
CBT vs. PP	9	446	1.60 (0.95–2.67)	0.077	18.8 (0–81.3)	0.69
CBT vs. NT	14	829	2.07 (1.35–3.18)	<0.001	38.8 (0–79.0)	0.34
CBT vs. WL	28	1486	3.99 (2.76–5.77)	<0.001	31.5 (0–76.9)	<0.001
PP vs. NT	1	24	2.04 (0.40–10.56)	0.394	–	–
PP vs. WL	1	20	6.00 (0.53–67.7)	0.147	–	–

CBT, cognitive-behaviour therapies; PP, psychological *placebo*; WL, waiting list.

Table 3. Results of network meta-analysis

	PP	NT	WL
CBT	1.65 (0.76–3.13)	2.36 (1.31–4.26)	6.26 (3.90–10.1)
PP	–	1.62 (0.62–3.59)	4.31 (1.74–9.46)
NT	–	–	2.87 (1.33–5.65)

OR shows the effectiveness of the intervention on the left over that on the top. 95% credible intervals in parentheses. Resdev = 109.2 (Data points: 98), DIC = 181.2.
PP, psychological *placebo*; NT, no treatment; WL, waiting list; CBT, cognitive-behaviour therapies.

overall NMA estimates obtained in this study are stable and no influential outliers were involved.

Given the strong small study effects in the comparison between CBT and WL, we ran two additional meta-regressions incorporating the assumption that there is such a bias favoring CBT over WL (30). Two models were hypothesized: The first model assumed that there is such a bias not only between CBT and WL but that there are other similar (exchangeable) levels of biases in all the other comparisons in the network; the second model, on the other hand, assumed that there is no small study effects between CBT and PP but that there are exchangeable small study effects biases in all the other comparisons. The DIC suggested that the second model was a better fitting model. In this model, the regression coefficient representing small study effects for all comparisons other than CBT vs. PP was statistically significant at −1.255 (−2.713 to −0.470). (Table 4).

Discussion

We identified a fairly dense, well-connected, homogeneous and consistent evidence network around CBT and its control conditions in the acute phase

Table 4. Results of the network meta-regressions adjusting for small study effects

	PP	NT	WL
CBT	1.34 (0.70–2.24)	1.63 (0.78–3.15)	1.79 (0.97–2.87)
PP	–	1.31 (0.54–2.79)	1.44 (0.64–2.78)
NT	–	–	1.22 (0.48–2.42)

OR shows the effectiveness of the intervention on the left over that on the top. 95% credible intervals in parentheses. Resdev = 96.97 (Data points: 98), DIC = 162.2.
PP, psychological *placebo*; NT, no treatment; WL, waiting list; CBT, cognitive-behaviour therapies.

treatment of depression. The effect size estimates for CBT were substantively different, depending on the control condition; the odds ratios for response (50% or greater reduction in depression severity) was not statistically significant at 1.7 (0.8–3.1) when CBT was compared against PP, but were 2.4 (1.3–4.3) in comparison with NT and 6.3 (3.9–10.1) in comparison with WL. Surprisingly, the network meta-analytical estimate of the odds ratio of NT over WL was significantly greater than unity at 2.9 (1.3–5.7).

There are, however, several factors that may undermine the robustness of these estimates. First of all, the methodological standards of the included studies were often less than ideal, as has already been often pointed out by several authors with regard to psychotherapy literature (41, 42). Apart from therapist and participant blinding, which essentially is impossible in psychotherapy trials, only a quarter of the included studies, on average, were rated to be at low risk of bias on various aspects of methodological rigour. We were unable to run the preplanned sensitivity analyses, limiting the included studies to those with high quality. We must remember, however, that these are still all randomized trials satisfying the minimum level of evidence quality and that this is the best evidence body we currently possess.

The apparent existence of small study effects is another major threat. There were notable small study effects in the comparison between CBT and WL. Why small study effects were so preponderant in the comparison against WL is not clear. To gauge the influence of the small study effects bias on our comparison of interest between NT and WL, we ran sensitivity analyses, trying to statistically adjust for such biases through meta-regression and found the odds ratio of NT over WL was no longer statistically significant in the better fitting of the two hypothesized models. However, there is no established method to adjust for funnel plot asymmetry in pairwise let alone network meta-analyses (30). It is possible that our methods may have over-corrected for the small study effects as none of the ORs for CBT over control conditions, including NT and WL, was significant in this adjusted model. These results therefore remain exploratory.

The differential drop-out rates cannot be the reason for the observed difference in effect sizes because we assumed all drop-outs to be non-responders, making the comparison most favourable to WL with the smallest drop-out rate and least favourable to NT with the highest drop-out rate (43). The differences in the drop-out rates would have worked, if any, in decreasing or even subverting the observed differences between NT and WL.

One may wonder how WL can be 'less effective' than NT. In both conditions, participants had earlier shown interest in receiving CBT for their depression, had provided informed consent to be randomized but then allocated to inactive control conditions. Ethically, participants on both NT and WL are allowed to receive some medical care of their own choice during the course of the study on a naturalistic basis. Those allocated to WL may, however, be more motivated to remain depressive so that they can receive their originally desired therapy after the study period is over, while those allocated to NT may more actively seek other treatments, either by oneself or by others, for their ailment.

In summary, the currently available best evidence, analysed by use of NMA, suggested that different control conditions lead to substantively different treatment effect estimates and that WL control may generate bigger effect sizes estimates for CBT than NT or PP. In other words, WL could be regarded a nocebo condition if it indeed is inferior to NT, that is, doing nothing. Unfortunately, the less than ideal quality of the evidence body, including probable publication bias, undermines the strength of this finding. However, there are several research implications of this study that remain valid. We will need to pay more attention to the differences in the control conditions in future psychotherapy research. In individual trials of psychotherapy, the use of WL as control should be more carefully deliberated, as it probably cannot be equated with NT condition controlling for regression towards the mean and the natural course of the disease but instead it may introduce negative psychological expectation of 'waiting for the desired active treatment.' In systematic reviews and meta-analyses of psychotherapies, we probably should not lump different control conditions into one comparative arm.

Acknowledgements

This review is one publication of the High Impact Reviews of Effectiveness in Depression (HIRED) project in which a group of researchers within the Cochrane Collaboration Depression, Anxiety and Neurosis Group conducted systematic reviews of all available evidence for all psychological therapies for treating depression.

Declaration of interest

T. A. Furukawa has received honoraria for speaking at CME meetings sponsored by Asahi Kasei, Eli Lilly, GlaxoSmith-Kline, Mochida, MSD, Otsuka, Pfizer, Shionogi and

Tanabe-Mitsubishi. He is diplomate of the Academy of Cognitive Therapy. He has received royalties from Igaku-Shoin, Seiwa-Shoten and Nihon Bunka Kagaku-sha. He is on advisory board for Sekisui Chemicals and Takeda Science Foundation. The Japanese Ministry of Education, Science, and Technology, the Japanese Ministry of Health, Labor and Welfare, and the Japan Foundation for Neuroscience and Mental Health have funded his research projects. DC has received honoraria for delivering training at Pfizer. DC is funded by the Medical Research Council, UK. All the other authors have no conflicts of interest to declare.

References

1. Mohr DC, Spring B, Freedland KE et al. The selection and design of control conditions for randomized controlled trials of psychological interventions. Psychother Psychosom 2009;**78**:275–284.
2. Borkovec TD, Sibrava NJ. Problems with the use of placebo conditions in psychotherapy research, suggested alternatives, and some strategies for the pursuit of the placebo phenomenon. J Clin Psychol 2005;**61**:805–818.
3. Rosenthal R, Frank JD. Psychotherapy and the placebo effect. Psychol Bull 1956;**53**:294–302.
4. Cuijpers P, Smit F, Bohlmeijer E, Hollon SD, Andersson G. Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. Br J Psychiatry 2010;**196**:173–178.
5. Jakobsen JC, Hansen JL, Storebo OJ, Simonsen E, Gluud C. The effects of cognitive therapy versus 'no intervention' for major depressive disorder. PLoS ONE 2011;**6**:e28299.
6. Stevens SE, Hynan MT, Allen M. A meta-analysis of common factor and specific treatment effects across the outcome domains of the phase model of psychotherapy. Clin Psychol Sci Pract 2000;**7**:273–290.
7. Hrobjartsson A, Gotzsche PC. Placebo interventions for all clinical conditions. Cochrane Database Syst Rev 2010: CD003974.
8. Baker AL, Hiles SA, Thornton LK, Hides L, Lubman DI. A systematic review of psychological interventions for excessive alcohol consumption among people with psychotic disorders. Acta Psychiatr Scand 2012;**126**:243–255.
9. Baskin TW, Tierney SC, Minami T, Wampold BE. Establishing specificity in psychotherapy: a meta-analysis of structural equivalence of placebo controls. J Consult Clin Psychol 2003;**71**:973–979.
10. Cuijpers P, van Straten A, Warmerdam L, Smits N. Characteristics of effective psychological treatments of depression: a metaregression analysis. Psychother Res 2008;**18**:225–236.
11. Watanabe N, Hunot V, Omori IM, Churchill R, Furukawa TA. Psychotherapy for depression among children and adolescents: a systematic review. Acta Psychiatr Scand 2007;**116**:84–95.
12. Vancampfort D, Vansteelandt K, Scheewe T et al. Yoga in schizophrenia: a systematic review of randomised controlled trials. Acta Psychiatr Scand 2012;**126**:12–20.
13. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. Stat Med 1996;**15**:2733–2749.
14. Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. Ann Intern Med 2013;**159**:130–137.
15. Pinquart M, Duberstein PR, Lyness JM. Treatments for later-life depressive conditions: a meta-analytic comparison of pharmacotherapy and psychotherapy. Am J Psychiatry 2006;**163**:1493–1501.
16. Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. J Consult Clin Psychol 2008;**76**:909–922.
17. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;**23**:56–62.
18. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961;**4**:561–571.
19. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;**59**:12–19.
20. Furukawa TA, Cipriani A, Barbui C, Brambilla P, Watanabe N. Imputing response rates from means and standard deviations in meta-analyses. Int Clin Psychopharmacol 2005;**20**:49–52.
21. da Costa BR, Rutjes AW, Johnston BC et al. Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: meta-epidemiological study. Int J Epidemiol 2012;**41**:1445–1459.
22. Shinohara K, Honyashiki M, Imai H et al. Behavioural therapies versus other psychological therapies for depression. Cochrane Database Syst Rev 2013;**10**:CD008696.
23. Higgins JP, Green S, eds. Cochrane handbook for systematic reviews of interventions version 5.1.1 [updated March 2011] Available from: www.cochrane-handbook.org, 2011.
24. Dersimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;**7**:177–188.
25. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;**315**:629–634.
26. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw 2010;**36**:1–48.
27. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997;**50**:683–691.
28. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). J R Stat Soc Series B 2002;**64**:583–639.
29. Dias S, Sutton AJ, Welton NJ, Ades AE. Nice, DSU technical support document 3: heterogeneity: subgroup, meta-regression, bias and bias-adjustment. Available at: http://www.nicedsu.org.uk/TSD3%20Heterogeneity.final%20report.08.05.12.pdf (accessed 26 March 2014).
30. Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. Res Synth Methods 2012;**3**:161–176.
31. Thompson LW, Gallagher D, Breckenridge JS. Comparative effectiveness of psychotherapies for depressed elders. J Consult Clin Psychol 1987;**55**:385–390.
32. Comas-Diaz L. Effects of cognitive and behavioral group treatment on the depressive symptomatology of Puerto Rican women. J Consult Clin Psychol 1981;**49**:627–632.
33. Cullen JM. Testing the effectiveness of behavioral activatin therapy in the treatment of acute unipolar depression. [PhD dissertation]. Western Michigan University, 2002.
34. Fry PS. Cognitive training and cognitive-behavioral variables in the treatment of depression in the elderly. Clin Gerontol 1984;**3**:25–45.
35. Morris NE. A group self-instruction method for the treatment of depressed outpatients. [PhD dissertation]. University of Toronto, 1975.
36. Neimeyer RA, Weiss ME. Cognitive and symptomatic predictors of outcome of group therapies for depression. J Cogn Psychother 1990;**4**:23–32.

37. Shaw BF. Comparison of cognitive theerapy and behavior therapy in the treatment of depression. J Consult Clin Psychol 1977;**45**:543–551.
38. Zeiss AM, Lewinsohn PM, Munoz RF. Nonspecific improvement effects in depression using interpersonal skills training, pleasant activity schedules, or cognitive training. J Consult Clin Psychol 1979;**47**:427–439.
39. Propst LR. The comparative efficacy of religious and nonreligious imagery for the treatment of mild depression in religious individuals. Cogn Ther Res 1980;**4**: 167–178.
40. Fuchs CZ, Rehm LP. A self-control behavior therapy program for depression. J Consult Clin Psychol 1977;**45**:206–215.
41. Cuijpers P, van Straten A, Bohlmeijer E, Hollon SD, Andersson G. The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. Psychol Med 2010;**40**:211–223.
42. Lynch D, Laws KR, McKenna PJ. Cognitive behavioural therapy for major psychiatric disorder: does it really work? A meta-analytical review of well-controlled trials. Psychol Med 2010;**40**:9–24.
43. Mazzotti E, Barbaranelli C. Dropping out of psychiatric treatment: a methodological contribution. Acta Psychiatr Scand 2012;**126**:426–433.
44. Arean PA, Perri MG, Nezu AM, Schein RL, Christopher F, Joseph TX. Comparative effectiveness of social problem-solving therapy and reminiscence therapy as treatments for depression in older adults. J Consult Clin Psychol 1993;**61**:1003–1010.
45. Ayen I, Hautzinger M. Kognitive verhaltenstherapie bei depressionen im klimakterium: eine kontrollierte, randomisierte interventionsstudie. Z Klin Psychol Psychother 2004;**33**:290–299.
46. Barrera M Jr. An evaluation of a brief group therapy for depression. J Consult Clin Psychol 1979;**47**:413–415.
47. Besyner JK. The comparative efficacy of cognitive and behavioral treatments of depression: A multi-assessment approach [PhD dissertation]. Graduate Faculty of Texas Tech University, 1978.
48. Brown RA, Lewinsohn PM. A psychoeducational approach to the treatment of depression: comparison of group, individual, and minimal contact procedures. J Consult Clin Psychol 1984;**52**:774–783.
49. Carrington CH. A comparison of cognitive and analytically oriented brief treatment approaches to depression in black women [PhD dissertation]. University of Maryland, 1979.
50. Collins RW. The treatment of depression: An integrative psychotherapy model [PhD dissertation]. San Francisco, CA: Saybrook Institute, 1996.
51. Dowrick C, Dunn G, Ayuso-Mateos JL et al. Problem solving treatment and group psychoeducation for depression: multicentre randomised controlled trial. Outcomes of Depression International Network (ODIN) Group. BMJ 2000;**321**:1450–1454.
52. Embling S. The effectiveness of cognitive behavioural therapy in depression. Nurs Stand 2002;**17**:33–41.
53. Epstein D. Aerobic activity versus group cognitive therapy: An evaluative study of contrasting interventions for the alleviation of clinical depression [PhD dissertation]. University of Nevada Reno, 1986.
54. Faramarzi M, Alipor A, Esmaelzadeh S, Kheirkhah F, Poladi K, Pash H. Treatment of depression and anxiety in infertile women: cognitive behavioral therapy versus fluoxetine. J Affect Disord 2008;**108**:159–164.
55. Hamamci Z. Integrating psychodrama and cognitive behavioral therapy to treat moderate depression. Arts Psychother 2006;**33**:199–207.
56. Hamdan-Mansour AM, Puskar K, Bandak AG. Effectiveness of cognitive-behavioral therapy on depressive symptomatology, stress and coping strategies among Jordanian university students. Issues Ment Health Nurs 2009;**30**: 188–196.
57. Hautzinger M, Welz S. Kognitive verhaltenstherapie bei depressionen im alter. Ergebnisse einer kontrollierten vergleichsstudie unter ambulanten bedingungen an depressionen mittleren schweregrads [Cognitive behavioral therapy for depressed older outpatients–a controlled, randomized trial]. Z Gerontol Geriatr 2004;**37**:427–435.
58. Hayman PM, Cope CS. Effects of assertion training on depression. J Clin Psychol 1980;**36**:534–543.
59. Hegerl U, Hautzinger M, Mergl R et al. Effects of pharmacotherapy and psychotherapy in depressed primary-care patients: a randomized, controlled trial including a patients' choice arm. Int J Neuropsychopharmacol 2010;**13**:31–44.
60. Hess-Homeier MJ. A comparison of Beck's cognitive therapy and jogging as treatments for depression [PhD dissertation]. University of Montana, 1981.
61. Kelly LM. Rational emotive therapy versus Lewinsohnian based approaches to the treatment of depression [PhD dissertation]. Athens, GA: University of Georgia, 1982.
62. Malouff JM. A study of a brief, cognitive treatment for depression personsl who have recently experienced a martial separation [PhD dissertation]. Arizona State University, 1984.
63. Miranda J, Chung JY, Green BL et al. Treating depression in predominantly low-income young minority women: a randomized controlled trial. JAMA 2003;**290**:57–65.
64. Nezu AM. Efficacy of a social problem-solving therapy approach for unipolar depression. J Consult Clin Psychol 1986;**54**:196–202.
65. Nezu AM, Perri MG. Social problem-solving therapy for unipolar depression: an initial dismantling investigation. J Consult Clin Psychol 1989;**57**:408–413.
66. Pace FR. Behavioral techniques in the treatment of depression [PhD dissertation]. University of New South Wales, 1977.
67. Pace TM. Changes in depressive self-schemata and depressive symptoms following cognitive therapy. J Couns Psychol 1993;**40**:288–294.
68. Pecheur DR. A comparison of the efficacy of secular and religious cognitive behavior modification in the treatment of depressed Christian college students [PhD dissertation]. Rosemead Graduate School of Professional Psychology, 1980.
69. Pellowe ME. Acceptance and commitment therapy as a treatment for dysphoria [PhD dissertation]. University of Wyoming, 2006.
70. Propst LR, Ostrom R, Watkins P, Dean T, Mashburn D. Comparative efficacy of religious and nonreligious cognitive-behavioral therapy for the treatment of clinical depression in religious individuals. J Consult Clin Psychol 1992;**60**:94–103.
71. Ross M, Scott M. An evaluation of the effectiveness of individual and group cognitive therapy in the treatment of depressed patients in an inner city health centre. JR Coll Gen Pract 1985;**35**:239–242.
72. Schmidt MM, Miller WR. Amount of therapist contact and outcome in a multidimensional depression treatment program. Acta Psychiatr Scand 1983;**67**:319–332.

73. Schmitt SG. Clinical depression: A comparative outcome study of two treatment approaches [PhD dissertation]. Fairleigh Dickinson University, 1988.
74. Serfaty MA, Haworth D, Blanchard M, Buszewicz M, Murad S, King M. Clinical effectiveness of individual cognitive behavioral therapy for depressed older people in primary care: a randomized controlled trial. Arch Gen Psychiatry 2009;**66**:1332–1340.
75. Taylor FG. Experimental analysis of a cognitive-behavioral therapy for depression. Cogn Ther Res 1977;**1**:59–72.
76. Usaf SO, Kavanagh DJ. Mechanisms of improvement in treatment for depression: test of a self-efficacy and performance model. J Cogn Psychother 1990;**4**:51–70.
77. Wilson PH. Combined pharmacological and behavioural treatment of depression. Behav Res Ther 1982;**20**:173–184.
78. Wilson PH, Goldin JC, Charbonneau-Powis M. Comparative efficacy of behavioral and cognitive treatments of depression. Cogn Ther Res 1983;**7**:111–124.
79. Wollersheim JP, Wilson GL. Group treatment of unipolar depression: a comparison of coping, supportive, bibliotherapy, and delayed treatment groups. Prof Psychol Res Pr 1991;**22**:496–502.
80. Wong DF. Cognitive and health-related outcomes of group cognitive behavioural treatment for people with depressive symptoms in Hong Kong: randomized wait-list control study. Aust N Z J Psychiatry 2008;**42**:702–711.
81. Wong DF. Cognitive behavioral treatment groups for people with chronic depression in Hong Kong: a randomized wait-list control design. Depress Anxiety 2008;**25**:142–148.
82. Wright JH, Wright AS, Albano AM et al. Computer-assisted cognitive therapy for depression: maintaining efficacy while reducing therapist time. Am J Psychiatry 2005;**162**:1158–1164.
83. Dunn G, Maracy M, Dowrick C et al. Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up. Br J Psychiatry 2003;**183**:323–331.

Cogn Ther Res (2014) 38:249–260
DOI 10.1007/s10608-014-9599-7

ORIGINAL ARTICLE

Specificity of CBT for Depression: A Contribution from Multiple Treatments Meta-analyses

Mina Honyashiki · Toshi A. Furukawa · Hisashi Noma · Shiro Tanaka · Peiyao Chen · Kayoko Ichikawa · Miki Ono · Rachel Churchill · Vivien Hunot · Deborah M. Caldwell

Published online: 29 January 2014

Abstract The "Dodo bird verdict," which claims that all psychotherapies are equally effective, has been a source of bewilderment and intense controversy among psychiatrists and psychologists. To examine this issue, we focused on cognitive-behavior therapy (CBT) and applied the newly developed review method known as multiple treatments meta-analysis (MTM). We identified randomized controlled trials comparing CBT against a psychological placebo (PP) and/or no treatment (NT) controls during the acute phase treatment of adults with depression. A random-effects MTM was conducted within a Bayesian framework. All the analyses were performed on an intention-to-treat basis. The MTM of the evidence network from 18 studies (39 treatment arms, 1,153 participants) revealed that CBT was significantly more likely to yield a response than NT (OR 2.24, 1.32–3.88) and that CBT was nominally, but not significantly, superior to PP (OR 1.30, 0.53–2.94), which in turn was superior to NT (OR 1.73, 0.67–4.84). The intervention effects in MTM were associated with the number of sessions, and the specificity of CBT increased as the number of sessions increased. The specific component of CBT was estimated to constitute 50.4 % (19.7–85.0) when CBT was given for ten or more sessions. Despite the quantitatively and qualitatively limited body of randomized evidence examining this issue, the present study strongly suggested a non-null specific component of CBT when given for an adequate length.

Keywords Multiple treatments meta-analysis · Cognitive behavior therapy · Dodo bird verdict · Common factor · Specific factor

Introduction

It was Rosenzweig (1936) who first conceptualized psychotherapy as consisting of (1) common (non-specific)

M. Honyashiki · T. A. Furukawa (✉) · P. Chen
Department of Health Promotion and Human Behavior, School of Public Health, Kyoto University Graduate School of Medicine, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan
e-mail: furukawa@kuhp.kyoto-u.ac.jp

T. A. Furukawa
Department of Clinical Epidemiology, School of Public Health, Kyoto University Graduate School of Medicine, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan

H. Noma
Department of Data Science, The Institute of Statistical Mathematics, Tokyo, Japan

S. Tanaka
Department of Pharmacoepidemiology, School of Public Health, Kyoto University Graduate School of Medicine, Kyoto, Japan

K. Ichikawa
Department of Health Information, School of Public Health, Kyoto University Graduate School of Medicine, Kyoto, Japan

M. Ono
Department of Psychiatry, Kyoto University Graduate School of Medicine, Kyoto, Japan

R. Churchill · V. Hunot
Academic Unit of Psychiatry, School of Social and Community Medicine, University of Bristol, Bristol, UK

D. M. Caldwell
School of Social and Community Medicine, University of Bristol, Bristol, UK

factors found in many different treatment approaches, and (2) specific factors proper to a particular treatment method and theory. This conceptualization later paved the way for Rosenthal and Frank's proposal of placebo psychotherapy, modeling pill placebo control in drug therapy trials, to establish the specific effectiveness of psychotherapies (Rosenthal and Frank 1956). They wrote in 1956: "...improvement under a special form of psychotherapy cannot be taken as evidence for (a) correctness of the theory on which it is based or (b) efficacy of the specific technique used, unless improvement can be shown to be greater than or qualitatively different from that produced by [...] a nonspecific form of psychotherapy."

The ensuing research efforts, however, have largely resulted in disappointing findings that are known as the Dodo bird verdict, which essentially states that all psychotherapies are equally effective (Baardseth et al. 2013; Luborsky et al. 2002; Luborsky and Singer 1975; Smith and Glass 1977; Wampold et al. 1997). The term originated from Rosenzweig's citation from Lewis Carroll's novel "Alice's Adventures in Wonderland," in which the characters get wet and have to dry themselves and the Dodo bird calls for a competition to run around the lake. When asked who won, the Dodo bird declares, "Everybody has won, and all must have prizes" (Rosenzweig 1936). The effectiveness of psychotherapies are thus postulated to be due to common factors, which include expectancy, relationship (empathy, warmth, alliance), and an explanatory framework (Greenberg and Newman 1996; Omer and London 1989).

However, the seminal papers cited above are subject to one or more of the following conceptual and methodological weaknesses.

1. As rightly criticized by Chambless et al. (Chambless 2002; Siev et al. 2010), the authors of these papers (Baardseth et al. 2013; Luborsky et al. 2002; Luborsky and Singer 1975; Smith and Glass 1977; Wampold et al. 1997) amalgamated very different comparisons for extremely diverse conditions among a wide spectrum of participants ranging from worried normal to psychotic inpatients. Their pooled effect size is therefore clinically uninterpretable. No one would choose his/her cancer therapy based on a meta-analysis of all therapies including all drugs, surgeries and radiation therapies for all stages of cancers of any histopathology and in any organ in the body.
2. Their dismissal of the obtained pooled effect size of 0.20 as small and clinically insignificant is factually and theoretically mistaken. First, one-third of established and acknowledged interventions in both medicine and psychiatry have effect sizes smaller than 0.3 in comparison with a placebo (Leucht et al. 2012). How can one expect a larger effect size when comparing active treatments? Second, an effect size of 0.20 corresponds with a number needed to treat (NNT) of around 15 for control event rates between 20 and 50 % (Furukawa 1999). A common mental disorder often has a 12-month prevalence of 1–5 %, which would translate into two to ten million sufferers per year in the USA alone; a therapy with an NNT of 15 could thus bring about 200,000–1,000,000 additional responses or remissions per year that an alternative therapy cannot achieve. This is not meaningless by any humane measure.
3. They base their arguments on the point estimate and ignore the uncertainties around it. In fact, the 95 % confidence interval of their obtained effect size is very wide, surpassing 0.50, which signifies a moderate effect according to Cohen's rule of thumb (Cohen 1988) and may, in fact, be more powerful than more than half of the established and currently practiced medical interventions (Leucht et al. 2012). The correct statistical interpretation of the obtained pooled effect size in these studies should be: no firm evidence to exclude neither clinically powerful difference in effect or no difference in effect, and not evidence of no clinically meaningful difference in effect.
4. It is most surprising that these meta-analyses are not based on a systematic search of all available evidence on a particular clinical topic, in view of the disconcerting magnitude of publication bias that has become widely known (Dickersin 1990; Song et al. 2000). For example, Wampold and colleagues' reviews limited their search to four English journals only (Ahn and Wampold 2001; Wampold et al. 1997). Luborsky based their analyses on, alas, "our collection of meta-analyses" (Luborsky et al. 2002).

On the other hand, there have also been attempts to refute the Dodo bird verdict by quantifying the specific versus non-specific components in the effectiveness of psychotherapies, the most well-known of which is the one by Lambert and Barley (2001). Based on "a subset of more than 100 studies that provided statistical analyses of the predictors of outcome" they concluded that specific techniques explained 15 % of the total improvement in psychotherapy, the remaining being explained by common factors (30 %), expectancy (15 %) and extra therapeutic change (40 %). Stevens et al. (2000) were more specific: they calculated effect sizes for 80 outcome studies that each contained no treatment (NT), a common factor, and treatment groups. The effect size in terms of symptom improvement was 0.58 for treatment versus NT, which then was roughly additive of that between treatment and the common factor (0.26) and that between the common factor

and NT (0.35). Bowers and Clum (1988) did a similar analysis for behavior therapy by performing a meta-analysis of studies that had both a placebo condition and a NT condition: the overall effect size of the treatment was 0.76, of which 0.55 was specific and 0.21 was non-specific. Barker et al. (1988) limited themselves to credible placebo controls and found that the overall effect size of the treatment was 1.06, of which 0.55 was specific and 0.47 was non-specific. In other words, of the effectiveness of psychotherapies over NT, the percentage contributed by specific factors ranged widely, with values of 25, 45, 72, and 52 %, respectively. None of these figures may be clinically meaningless, but unfortunately all these reviews are subject to some or all of the criticisms described above.

Therefore, it is timely to ask how much specific versus non-specific components there are in the effectiveness of a specific psychotherapy for a well-delineated clinical condition using a modern systematic review methodology. The current study represents a secondary analysis of the Cochrane systematic reviews of six major psychotherapy schools for depression in adults (Hunot et al. 2013; Shinohara et al. 2013). The six schools included behavior therapies, cognitive-behavior therapies (CBT), third-wave cognitive therapies, psychodynamic therapies, humanistic therapies and integrative therapies. In order to quantitatively assess the specific versus non-specific components, the present study focuses on a triangular comparison between CBT, which were the most thoroughly researched of the six schools, and a psychological placebo (PP) and NT. We also applied a new meta-analysis technique, known as multiple treatments meta-analysis (MTM) or network meta-analysis (Higgins and Whitehead 1996), to this triangular comparison to combine the direct and indirect comparisons contained therein, so that we can make the maximal use of the available randomized evidence.

Methods

Criteria for Considering Studies for this Review

We included only randomized controlled studies comparing CBT with PP and/or NT in the acute phase treatment of adults with depression. Quasi-randomized studies, such as those using allocation by day of the week, date of birth, or alternate allocations, were not eligible because a lack of allocation concealment leads to overestimation (Schulz et al. 1995). Both open and single-blinded (assessor-blinded) studies were eligible, as it is impossible to blind the therapists or participants in psychotherapy trials.

Depression could either be defined as unipolar major depression according to any of the operationalized diagnostic criteria (Feighner criteria, Research Diagnostic Criteria, DSM-III, DSM-III-R, DSM-IV, ICD-10) or as scoring above the accepted threshold of a validated depression screening instrument. Studies focusing on chronic or treatment-resistant depression were excluded.

Cognitive-behavior therapy includes cognitive therapy (Beck et al. 1979), rational emotive behavior therapy (Ellis 1979), problem-solving therapy (D'Zurilla and Goldfried 1971), self-control therapy (Fuchs and Rehm 1977), coping with depression course (Lewinsohn et al. 1984) and others that use both cognitive and behavioral skills for the treatment of depression.

Psychological placebo is defined as an experimental condition used in an attempt to control for non-specific factors. The criteria for a control condition to be regarded PP were as follows: (1) intervention is regarded as lacking active components by researchers in a trial but is explained as active to the participants; (2) the number and duration of the face-to-face session is equivalent with active treatment in the same study and; (3) the qualification of the therapists is equivalent to that for the active treatment. We did not include pill placebo controls because they control for the regression towards the mean, the natural course and treatment expectancy but not the common therapeutic factors of psychotherapy (Hollon and DeRubeis 1981).

No treatment consists of patients who did not receive either active or non-specific interventions. This control condition controls for the regression towards the mean and the natural course of the condition. We did not include waiting list controls, which are often used in psychotherapy research, among the NT controls.

Study Selection and Data Extraction

To identify relevant studies, we searched two clinical trial registries created and maintained by the Cochrane Depression, Anxiety and Neurosis Group (CCDAN), the CCDANCTR-Studies and CCDANCTR-References, supplemented by corresponding searches in CINAHL, PSYINDEX, and reference searches. The details of the search strategies for these registries can be found on the Cochrane Collaboration Depression, Anxiety and Neurosis Group's webpage (http://ccdan.cochrane.org/). The most recent updated search for this review was done in February 2012. The quality ratings were operationalized, and studies were categorized into either a low risk of bias, a high risk of bias, or an unclear risk of bias for each domain. All the assessments were performed by two independent review authors, and disagreements were resolved by discussion between two authors and, where necessary, in consultation with a third author. Missing information was sought by contacting the original authors, whenever possible.

Outcome Measures

Acute treatment was defined as an 8-week treatment in the analyses. If 8-week data were not available, we used data ranging between 4 and 16 weeks, and the time point given in the original study as the study endpoint was given preference.

Response was our pre-defined primary outcome, as this allows the inclusion of all dropouts and thus enables a conservative estimate of the treatment effect according to the intention-to-treat principle. We defined response as the proportion of patients who showed a reduction of at least 50 % from the baseline score on the Hamilton Rating Scale for Depression (HAM-D), the Montgomery-Asberg Depression Rating Scale (MADRS), or any other validated depression scale at the above-defined time point. If the original authors reported several outcomes, we gave preference to the BDI for a self-rating scale and the HAM-D for an observer-rating scale. Observer-rated scales were preferred to self-reported scales.

Intention-to-treat analyses were based on the total number of randomly assigned participants, irrespective of how the original study investigators analyzed the data, by assuming that all dropouts were non-responders. For studies in which the exact numbers of participants who had responded were not reported, but the means and standard deviations for continuous depression scales were reported, the number of responders was calculated using a validated imputation method (da Costa et al. 2012; Furukawa et al. 2005).

Analysis

Multiple Treatments Meta-analyses, and Examination of Inconsistency/Heterogeneity

We conducted multiple treatments meta-analyses. To ensure that the network was connected, a network diagram was constructed. Random-effects MTM, allowing for the heterogeneity of treatment effects across studies, was conducted in a Bayesian framework using OpenBUGS 3.2.1. These methods combine direct and indirect evidence for all three pairs of treatments. A key assumption of MTM is that of consistency, i.e., that direct and indirect evidence do not disagree beyond chance. In the first instance, one should ensure that the subsets of trials forming the network are similar in factors which could modify the treatment effect. Where feasible, consistency should also be statistically evaluated. Here, we used the posterior mean of the residual deviance as a global goodness of fit statistic to assess consistency. In a well-fitting model the residual deviance should be close to the number of data points. In case with considerable inconsistency, we investigated the possible sources.

Quantifying Specific Versus Non-specific Components

The relative contributions of specific effects and non-specific effects were estimated by dividing log ($OR_{CBT,PP}$) or log ($OR_{PP,NT}$) by log ($OR_{CBT,NT}$), where $OR_{X,Y}$ represents the odds ratio of treatment X over treatment Y.

Publication Bias and Sensitivity Analyses

To assess publication bias, we drew funnel plots for pairwise comparisons if the number of studies contributing to that comparison was ten or greater. To examine if the obtained results were preserved when we limited the included studies to only high-quality ones, we had planned a priori to examine the following variables: risk of biases (limiting to trials with a low risk of bias at allocation concealment, blinding of assessor, and treatment fidelity), included disorders, and response imputation.

Meta-regression

The following sources of possible clinical heterogeneity, which had been listed a priori, were examined as effect modifiers in network meta-analyses: number of sessions, group versus individual format, baseline depression severity, and concomitant pharmacotherapy.

Results

Selection and Inclusion of Studies

Out of 6,710 studies identified through an electronic search and reference search, 195 full-text articles were retrieved, of which 18 studies (comprising 39 treatment arms, and 1,153 participants) satisfied the eligibility criteria for the present study (Fig. 1).

Characteristics of the Included Studies

Figure 2 shows the network of evidence comparing CBT, PP, and NT. The characteristics of the included studies are listed in Table 1. The contents of the PP conditions are listed in Table 2. Two of the 18 studies had two CBT arms. Five of the 18 studies used an individual format for CBT or PP, 11 studies used a group format, and the remaining two used both formats. The number of sessions ranged from 4 to 12 sessions. Ten of the 15 studies allowed concomitant pharmacotherapy, while five studies did not. Only two

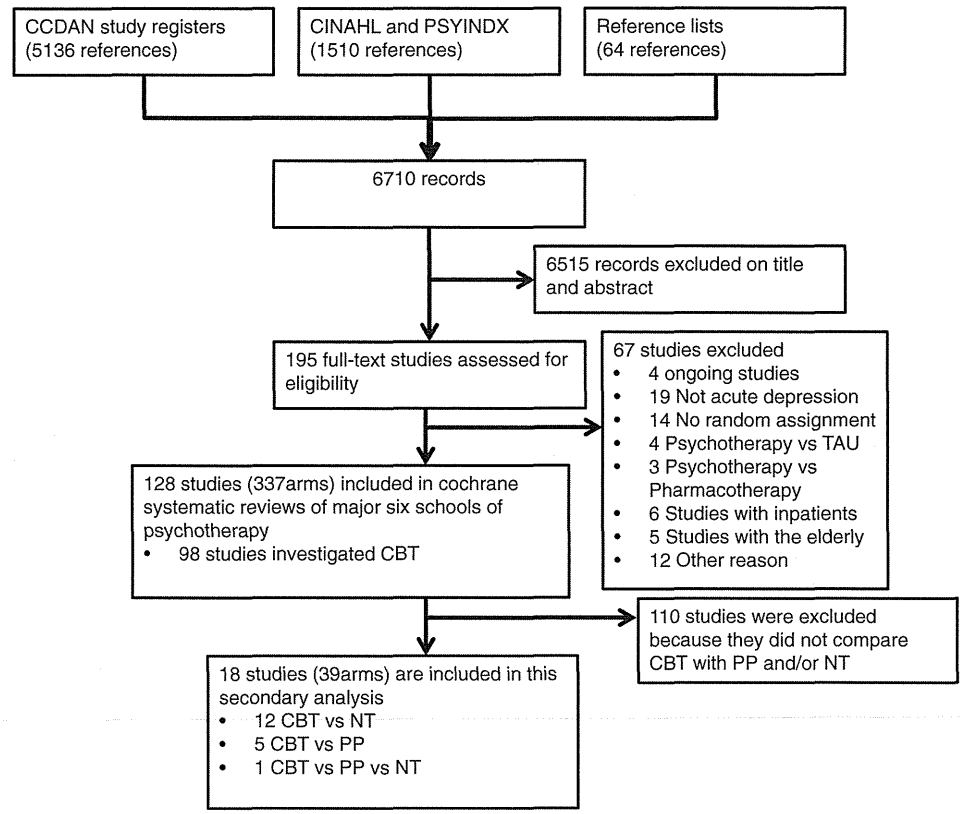

Fig. 1 Flowchart for selection of studies

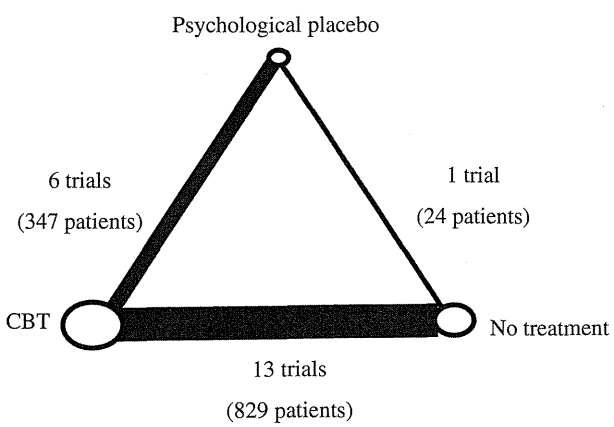

Fig. 2 Evidence network the size of each dot is proportional to the number of patients allocated and the width of line to the number of trials. Numbers do not add up to numbers in Table 1 because of a multi-arm trial by Propst 1980

studies used an observer scale (HAMD) as an outcome measure, while the other 16 studies used a self-rating scale (BDI). The mean baseline severity on the BDI was minimal (14–19) in one study, mild (20–28) in 14 studies, and moderate (>28) in one study. The quality of the included studies varied but was generally moderate. Ten studies reported adequate allocation concealment. One out of two studies using an objective scale reported the blinding of the assessors. Three studies reported fidelity monitoring for CBT or PP. Twelve studies included patients with major depressive disorder diagnosed according to operationalized diagnostic criteria, while the remaining six included patients scoring above the accepted threshold of a validated depression screening instrument. We had to use the imputed response rates based on the continuous severity score at the end of treatment in 16 studies. All but one study provided data on the numbers of randomized patients. We used the number of participants assessed at the end of treatment as the denominator for the remaining study.

Pair-wise Meta-analyses

We conducted CBT versus PP and CBT versus NT pair-wise meta-analyses (Table 3). These analyses showed that CBT was significantly more effective than NT in bringing about a response. The CBT versus PP comparison was not significant. Overall, the heterogeneity was moderate, although for all comparisons the 95 % CI included values that showed very high or no heterogeneity, reflecting the small number of included studies for each pair-wise comparison.

Table 1 Selected characteristics of the included studies

Study	N of arms in:			N	Included disorders	Baseline BDI	Format	N of sessions	Con- comitant pharmaco-therapy	Outcome scale	Risk of Bias			Response imputed
	CBT	PP	NT								Allocation concealment	Blinding of assessors	Treatment fidelity	
Besyner1979 (Besyner 1979)	1	1		20	Other	24.9	Grp	4	Unclear	BDI	Unclear	High	Unclear	Imputed
Dowrick_Finland[a] Rural1996 (Dowrick et al. 2000; Dunn et al. 2003)	1		1	50	MDD+	21.1	Ind	6	Allowed	BDI	Low	High	Unclear	Imputed
Dowrick_Finland[a] Urban1996 (Dowrick et al. 2000; Dunn et al. 2003)	1		1	47	MDD+	21.3	Ind	6	Allowed	BDI	Low	High	Unclear	Imputed
Dowrick_Ireland[a] UrbanRural1996 (Dowrick et al. 2000; Dunn et al. 2003)	1		1	38	MDD+	23	Grp	8	Allowed	BDI	Low	High	Unclear	Imputed
Dowrick_Norway[a] Rural1996 (Dowrick et al. 2000; Dunn et al. 2003)	1		1	61	MDD+	19.2	Grp	8	Allowed	BDI	Low	High	Unclear	Imputed
Dowrick_Norway[a] Urban1996 (Dowrick et al. 2000; Dunn et al. 2003)	1		1	67	MDD+	21	Grp	8	Allowed	BDI	Low	High	Unclear	Imputed
Dowrick_Spain[a] Urban1996 (Dowrick et al. 2000; Dunn et al. 2003)	1		1	30	MDD+	22	Ind	6	Allowed	BDI	Low	High	Unclear	Imputed
Dowrick_UK[a] Rural1996 (Dowrick et al. 2000; Dunn et al. 2003)	1		1	49	MDD+	26	Ind	6	Allowed	BDI	Low	High	Unclear	Imputed
Dowrick_UK[a] Urban1996 (Dowrick et al. 2000; Dunn et al. 2003)	2		1	84	MDD+	24.8	Ind/ Grp	6/8	Allowed	BDI	Low	High	Unclear	Imputed
Faramarzi 2008 (Faramarzi et al. 2008)	1		1	82	Other	19.9	Grp	10	No	BDI	Unclear	High	Unclear	Imputed
Fuchs1977 (Fuchs and Rehm 1977)	1	1		18[b]	Other	NA	Grp	6	Unclear	BDI	Unclear	High	Unclear	Imputed
Hamamci2006 (Hamamci 2006)	1		1	24	Other	28.4	Grp	11	No	BDI	Unclear	High	Unclear	Imputed
Hamdan-Mansour2009 (Hamdan-Mansour et al. 2009)	1		1	84	Other	24.1	Grp	10	Unclear	BDI	Low	High	Low	Imputed
Hegerl2010 (Hegerl et al. 2010)	1	1		120	MDD+	NA	Grp	10	No	HAMD	Unclear	Unclear	Low	No
Kelly1982 (Kelly 1982)	1	1		16	MDD+	25.4	Grp	6	Allowed	BDI	Unclear	High	Unclear	Imputed
Miranda2003 (Miranda et al. 2003)	1		1	179	MDD+	NA	Ind/ Grp	8	No	HAMD	Low	Low	Unclear	No

Table 1 continued

Study	N of arms in:			N	Included disorders	Baseline BDI	Format	N of sessions	Con- comitant pharmaco-therapy	Outcome scale	Risk of Bias			Response imputed
	CBT	PP	NT								Allocation concealment	Blinding of assessors	Treatment fidelity	
Propst1980 (Propst 1980)	2	1	1	47	Other	15.4	Grp	8	No	BDI	Unclear	High	Unclear	Imputed
Serfaty2009 (Serfaty et al. 2009)	1	1		137	MDD+	26.8	Ind	12	Allowed	BDI	Low	High	Low	Imputed

CBT cognitive behavior therapies, *PP* psychological placebo, *NT* no treatment, *BDI* beck depression inventory, *MDD+* major depressive disorder diagnosed by operationalised diagnostic criteria

[a] Dowrick et al. (2000) reports nine independently conducted, albeit according to concerted protocols, RCTs. Two of these RCTs conducted in Ireland were reported in an amalgamated form in the definitive report Dunn et al. (2003) and is therefore treated as one trial in this meta-analysis

[b] For Fuchs and Rehm (1977), randomized N was not available. Instead we used number of participants assessed at the end of intervention

Multiple Treatment Meta-analyses and Examination of Inconsistency/Heterogeneity

The consistency model provided an adequate fit to the data, with a posterior mean residual deviance of 37.8 for 37 data points, although an index of heterogeneity (the median between-trials standard deviation) was relatively high ($\sigma = 0.70$). Table 4 summarizes the results of the MTM. CBT was significantly superior to NT. CBT was not significantly different from PP, nor was PP from NT.

Publication Bias and Sensitivity Analyses

We drew a funnel plot for the primary outcome of the studies comparing CBT and NT. Egger's test was not significant ($P = 0.34$). For other comparisons, the number of comparisons was too small for a funnel plot.

There were not enough studies to conduct MTM for sensitivity analyses, so we only conducted pair-wise meta-analyses. Among them, limiting the studies to high-quality trials did not change the overall results (see Table 3).

Meta-regression

We conducted meta-regressions for MTM to examine the effects of selected covariates on efficacy. The association between the treatment effect and the number of sessions was significant (slope −0.21; 95 % CrI −0.42 to −0.002). We found no indication that the treatment efficacy was significantly associated with the baseline depression severity according to the BDI (slope −0.05; 95 % CrI −0.21 to 0.10), nor did we find an association between the effect size and the CBT format (slope: −0.04; 95 %CrI: −1.28 to 1.18) or concomitant pharmacotherapy (slope −0.52; 95 % CrI −1.56 to 0.45).

Figure 3 shows the estimated relationship between the number of sessions and the specificity of CBT. Table 4 presents a post hoc meta-regression dichotomizing the number of session into "≥10" and "<10". The specific component now contributed 50.4 % (95 % CrI 19.7–85.0 %) of the total efficacy of CBT over NT when the number of sessions was 10 or over. The interaction was qualitative (Table 4), suggesting that CBT is specifically beneficial only if it is given in 10 or more sessions.

Discussion

A systematic comprehensive search of the literature yielded a network of evidence of 18 studies (comprising 39 arms, and 1153 patients) comparing CBT, PP, and NT. The MTM of the evidence network was consistent, revealing that CBT was significantly more likely to yield a response