Original article

Clinical relevance of findings in trials of CBT for depression

P. Lepping [a,b,c,*], R. Whittington [d], R.S. Sambhi [c], S. Lane [d], R. Poole [a], S. Leucht [e], P. Cuijpers [f], R. McCabe [g], W. Waheed [h,i]

[a] Betsi Cadwaladr University Health Board, Centre for Mental Health and Society, Bangor University, North Wales, UK
[b] Mysore Medical College and Research Institute, Mysuru, India
[c] Department of Psychiatry, Betsi Cadwaladr University Health Board, North Wales, UK
[d] Department of Health Services Research, University of Liverpool, Liverpool, UK
[e] Department of Psychiatry and Psychotherapy, Technische Universität München, München, Germany
[f] Department of Clinical Psychology and EMGO Institute for Health and Care Research, VU University and VU University Medical Center, Amsterdam, The Netherlands
[g] Mersey Care NHS Trust, Liverpool, UK
[h] Lancashirecare NHS Trust, Preston and the University of Manchester, Manchester, UK
[i] Centre for Primary care, The University of Manchester, Manchester, UK

A R T I C L E   I N F O

A B S T R A C T

Cognitive behavioural therapy (CBT) is beneficial in depression. Symptom scores can be translated into Clinical Global Impression (CGI) scale scores to indicate clinical relevance. We aimed to assess the clinical relevance of findings of randomised controlled trials (RCTs) of CBT in depression. We identified RCTs of CBT that used the Hamilton Rating Scale for Depression (HAMD). HAMD scores were translated into Clinical Global Impression – Change scale (CGI-I) scores to measure clinical relevance. One hundred and seventy datasets from 82 studies were included. The mean percentage HAMD change for treatment arms was 53.66%, and 29.81% for control arms, a statistically significant difference. Combined active therapies showed the biggest improvement on CGI-I score, followed by CBT alone. All active treatments had better than expected HAMD percentage reduction and CGI-I scores. CBT has a clinically relevant effect in depression, with a notional CGI-I score of 2.2, indicating a significant clinical response. The non-specific or placebo effect of being in a psychotherapy trial was a 29% reduction of HAMD.

© 2017 Published by Elsevier Masson SAS.

1. Introduction

Depressive disorders are common throughout the world. They are by far the most common cause of years lost due to disability in high-, as well as low- and middle-, income countries, in both men and women [1]. Antidepressants and a range of psychological therapies have been shown to be beneficial in the treatment of depression and are recommended in treatment guidelines [2–4]. People suffering from depression often prefer psychological treatments to medication [5]. There is variation in individual responses to treatment, including long-term outcome. A range of psychotherapeutic interventions have been developed and refined to improve treatment outcomes [4]. Clinical trials of psychotherapies suggest that differences between these treatments are small and may have little clinical significance [6].

Previous research has explored the relationship between changes in commonly used symptom rating scales and clinical improvement. Based on data from clinical trials, statistical methods to convert symptom scores into notional Clinical Global Impression scale severity and change (CGI-S and CGI-C/CGI-I) scores have been developed [7–9]. These two scales quantify clinician's' overall impression of clinical severity and clinical change in participants' psychiatric condition with an aim to bring clinical relevance into trial results. The conversion helps in understanding the relationship between statistically significant results and meaningful clinical improvement.

Using a published systematic database of randomised trials of psychotherapies for adult depression [10], we aimed to assess the clinical relevance of outcomes reported for cognitive behavioural therapy trials in depression. We used a broad definition of CBT therapies, which includes cognitive bibliotherapy, but not behavioural activation therapy. We used a notional CGI translation of scores from the Hamilton Rating Scale for Depression (HAMD), ([11]: ECDEU Assessment Manual for Psychopharmacology), one of the most commonly used scales in CBT clinical trials [9,12].

* Corresponding author. Wrexham Academic Unit, Centre for Mental Health and Society, Technology Park, Wrexham, Wales LL13 7TP, United Kingdom. Fax: +44 1978 726 600.
  E-mail address: peter.lepping@wales.nhs.uk (P. Lepping).

2. Material and methods

We searched a comprehensive database of randomised controlled clinical trials of psychological therapies for depression [10,13]. A detailed description of the development of this database has been published and can be accessed online (www. evidencebasedpsychotherapies.org) [14]. Systematic review principles were used to include all possible RCT trials using CBT for the treatment of depression.

The inclusion criteria for our review were:

- participants: a diagnosis of depression with no psychiatric or physical co-morbidity;
- interventions: at least one type of psychological therapy classified in the database as CBT 14;
- comparator: any comparator or control. Any arm with a placebo is classed as a control;
- outcome measures: percentage change in mean Hamilton Depression Rating Scale (HAMD/HDRS) score, given directly or calculated from baseline to endpoint data;
- design randomised controlled trial;
- reporting:
  - published in a peer reviewed journal and included in the database up to 31.12.2013, the point to which the data base was complete at the time of the study. Books and conference posters were excluded,
  - published in any language,
  - sample size for each study arm reported,
  - available as electronic full-text or as paper full-text.

We obtained the full-text version of identified papers. Data were extracted by two reviewers independently (RSS and RW) and differences were resolved by consensus. Leucht et al. have demonstrated that changes in HAMD scales can be translated into notional Clinical Global Impressions scale, severity and improvement (CGI-S and CGI-I) scores (Table 1) [9]. The method used for the translation was equi-percentile linking of HAMD17 and CGI ratings from 43 drug trials in patients with major depressive disorder (MDD) treated with mirtazapine ($n = 7131$). This method has some limitations, but generates an acceptably robust translation, allowing an objective and reliable estimate of the clinical relevance of published findings. The same or similar methods have been used to evaluate clinical relevance of antipsychotic trial data [8], transcranial magnetic stimulation trial data [15], antidepressant trial data [16], and cognitive therapies for schizophrenia trial data [17].

To determine the CGI ratings from the reported HAMD data in our analysis, the mean percentage change from baseline to the last follow-up point was calculated for each study using an Excel spreadsheet. An overall mean percentage change was then calculated for aggregated treatment and control arms, as well as subgroups of particular treatments and control conditions. Hypothesis tests were carried out using the independent sample t-test, at the 5% significance level to access the statistical significance of mean percentage change between CBT subgroup

Table 1
CGI translation.

HAMD % change	CGI-I	Interpretation
−84%	1	Very much improved
−59%	2	Much improved
−33%	3	Minimally improved
−9%	4	No change
+8%	5	Minimally worse
+27.5%	6	Much worse
+60%	7	Very much worse

and the remaining four subgroups. The percentage HAMD change was plotted with CGI-I scores on a graph. The CGI-I score was extracted manually for each point (PL), as the conversion graphs are not linear [9].

The database classifies a therapy as CBT when cognitive restructuring (the evaluating, challenging, and modifying a patient's dysfunctional beliefs) was one of the core elements of the therapy. The subgroups are classified in accordance with the database. The study arms were categorised in discussion with the research team. The included studies are of diverse quality. They have different degrees of blinding or no blinding at all. Whilst all were RCTs, the recruitment processes were variable and may have relied on volunteers recruited by advertisements. This is in keeping with shortcomings of the psychotherapy literature in general, which are well described [3]. Study arms were classified into five subgroups:

- CBT alone ($n = 61$ study arms). See above for definition criteria;
- other psychological monotherapy ($n = 57$). This group includes a variety of active psychotherapies, including behavioural activation, provided only one therapy is applied to that group of participants;
- pharmacological monotherapy ($n = 21$). This group includes any antidepressant therapy where only one active drug was used;
- combinations of active therapies ($n = 31$). This group includes any combination of therapies, whether psychological or medication;
- controls ($n = 44$), this group includes waiting list controls, treatment as usual and placebo.

For a comprehensive list of the included study arms and their categorization see the list of included studies or contact the authors.

3. Results

The database was accessed in December 2013 generating 421 titles (Fig. 1). In the next stage, we excluded duplicates and papers with insufficient data. We obtained full-text versions of the remaining 393 papers, of which a further 311 were excluded. The remaining 82 papers were included in the analysis. Overall raw disagreement between the two independent extractors occurred with 13.7% data points. All disagreements were identified and resolved by re-extracting the data. The 82 studies included had between 2 and 10 study arms, yielding 170 datasets relating to a psychological therapy. By adding control arms, a total of 214 datasets were included in the analysis. These datasets included 6330 individual participants. HAMD scores at baseline confirmed that the participants met research criteria for depression, above the accepted threshold score of 8 or higher.

Table 2 shows the mean percentage change in HAMD scores and the translation into notional CGI-I scores for all five categories. The mean percentage HAMD change for the aggregated treatment arms was 53.66%, and 29.81% for aggregated control arms. This difference was statistically significant ($P < 0.05$). The notional CGI-I score for treatment was 2.25 (1 = "very much improved", 2 = "much improved", 3 = "minimally improved", 4 = "no change"). For controls, it was 3.1.

Amongst the different types of psychological and pharmacological treatments, the notional CGI-I score was lowest for the category "combination of active therapies", indicating the biggest clinical improvement. This was followed by "CBT alone", followed by "other psychological monotherapies" and "pharmacological monotherapies". Compared to "CBT alone", the HAMD percentage change of the category "combination of active therapies" showed statistical significance in favour of the "combination of active