Cognitive–behavioural therapy for a variety of conditions: an overview of systematic reviews and panoramic meta-analysis

Beth Fordham,[1*] Thavapriya Sugavanam,[1] Katherine Edwards,[1] Karla Hemming,[2] Jeremy Howick,[3] Bethan Copsey,[1] Hopin Lee,[1] Milla Kaidesoja,[4] Shona Kirtley,[1] Sally Hopewell,[1] Roshan das Nair,[5,6] Robert Howard,[7] Paul Stallard,[8] Julia Hamer-Hunt,[9] Zafra Cooper[10] and Sarah E Lamb[1,11] on behalf of the Cognitive Behavioural Therapy – Overview Expert Consultation Group

[1]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[2]Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[3]Faculty of Philosophy, University of Oxford, Oxford, UK
[4]Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland
[5]Department of Psychiatry and Applied Psychology, University of Nottingham, Nottingham, UK
[6]Institute of Mental Health, Nottinghamshire Healthcare NHS Foundation Trust, Nottingham, UK
[7]Division of Psychiatry, University College London, London, UK
[8]Department for Health, University of Bath, Bath, UK
[9]Public and patient representative, Oxford, UK
[10]Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA
[11]College of Medicine and Health, University of Exeter, Exeter, UK

*Corresponding author beth.fordham@ndorms.ox.ac.uk

Declared competing interests of authors: Zafra Cooper reports occasional fees for lectures and workshops on cognitive–behaviour therapy (CBT) for eating disorders, and payment to provide various clinical and research groups with supervision in CBT for eating disorders. Roshan das Nair reports being chairperson of the National Institute for Health Research (NIHR) Research for Patient Benefit East Midlands Regional Advisory Committee (2019 to present); he was also a NIHR Health Services and Delivery Research funding panel member (2018–20). Sally Hopewell reports membership of the Health Technology Assessment (HTA) Clinical Evaluation and Trials Committee from 2018 to the present. Robert Howard reports grants from NIHR HTA and Efficacy and Mechanism Evaluation programmes during the conduct of the study. He was a member of the HTA Commissioning Committee (2013–18) and the HTA Commissioning Sub-board (2016–17). Milla Kaidesoja reports grants from the Helsinki Institute of Life Science during the conduct of the study. Sarah E Lamb was on the HTA Additional Capacity Funding Board (2012–15), the HTA End of Life Care and Add-on Studies Board (September 2015), the HTA Prioritisation Group Board (2010–15) and the HTA Trauma Board (2007–8).

Published February 2021
DOI: 10.3310/hta25090

Scientific summary

CBT for a variety of conditions

Health Technology Assessment 2021; Vol. 25: No. 9
DOI: 10.3310/hta25090

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Cognitive–behavioural therapy is an amalgam of interventions that emerged from cognitive and behavioural psychological models. It aims to improve quality of life by changing maladaptive cognitions that maintain problematic symptoms. An overview of cognitive–behavioural therapy systematic reviews was conducted in 2012 and included 269 reviews, concluding that cognitive–behavioural therapy was effective across many conditions. However, only 11 of the included reviews synthesised randomised controlled trials. Since then, there have been many more randomised controlled trials and subsequent reviews. In parallel, there has been more guidance on improving trial and review quality. Hence, the time was right to undertake an updated overview, focused on high-quality randomised controlled trial evidence, to introduce new methods to understand how consistent the effects are across different conditions and to understand where future research resources would be best invested.

Objectives

This overview aimed to comprehensively map the existing evidence base to identify where we have high-quality evidence of the effectiveness of cognitive–behavioural therapy and where we have evidence gaps. Then we examined the consistency of the effectiveness of cognitive–behavioural therapy across different conditions and, when appropriate, generated an across-condition general effect estimate. Finally, we considered the extent to which the existing evidence base could be used to guide treatment, commissioning and research investment decisions.

To answer these research aims, we undertook two steps: (1) a mapping exercise – we identified all available systematic reviews of cognitive–behavioural therapy, assessed their quality and stratified them by quality, condition, context and population; and (2) a panoramic meta-analysis – we selected higher-quality reviews with sufficient quantitative data and conducted panoramic meta-analyses for the primary outcome of health-related quality of life and for the secondary outcomes of depression, anxiety and pain.

Finally, we considered the implications of the mapping and panoramic meta-analytic data. We used a model of generalisation to examine how the data of the overview answer the questions needed to inform treatment, commissioning and research investment decisions.

Methods

We worked with a cognitive–behavioural therapy expert consultation group consisting of clinical academics ($n = 6$), research academics ($n = 8$) and patient representatives ($n = 4$) throughout the overview process to guide the protocol development, synthesis strategy, data analysis, and interpretation.

Data sources and search strategy

The Database of Abstracts of Reviews of Effects (up to March 2015), Cochrane Database of Systematic Reviews, MEDLINE (via Ovid), EMBASE (via Ovid), PsycINFO (via Ovid), Cumulative Index to Nursing and Allied Health Literature (via EBSCO*host*), Child Development and Adolescent Studies (via EBSCO*host*) and OpenGrey databases were searched up until January 2019. Publication year was restricted to after 1992 to eliminate superseded reviews.

Inclusion criteria

A systematic review of cognitive–behavioural therapy in any condition [recognised in the *International Classification of Diseases*, Eleventh Revision (ICD-11)] across any age group or setting was considered for inclusion if:

- the review fulfilled at least four of the five Centre for Reviews and Dissemination criteria to qualify as a systematic review
- the intervention was reported as cognitive–behavioural therapy or included at least one cognitive and one behavioural element
- the cognitive–behavioural therapy trials were qualitatively or quantitatively summarised
- one of the following outcomes was considered in the review – health-related quality of life, depression, anxiety or pain
- it was available in English.

Stage one: mapping

Data extraction

The Covidence (Melbourne, VIC, Australia) platform was used for sifting and article management, and Microsoft Excel® (Microsoft Corporation, Redmond, WA, USA) was used for data extraction and management. Article screening, data extraction and quality assessment were conducted independently in duplicate by two researchers. A third researcher resolved conflicts. The online A MeaSurement Tool to Assess systematic Reviews (AMSTAR)-2 was used to assess the quality of the included reviews (i.e. high, moderate, low or critically low).

Reviews were categorised by the condition that they aimed to improve with cognitive–behavioural therapy. Physical conditions were classified by primary codes of the ICD-11, whereas mental conditions were represented by secondary codes under the primary code of 'Mental, behavioural and neurodevelopmental disorders.' We extracted descriptive information, such as participant characteristics (age, sex, ethnicity, country of residence), intervention (intensity, timing, context), control groups (active or non-active), outcomes, follow-up duration and patient perspectives (satisfaction, dropout rates, acceptability).

Evidence synthesis

The mapping exercise included producing (1) a bubble chart through TIBCO Spotfire® (TIBCO, Software Inc., Palo Alto, CA, USA) software to present the volume of evidence, in terms of number of reviews, randomised controlled trials and participants across all conditions; (2) summary tables to present the descriptions of the included reviews as per the ICD-11 classification; and (3) gaps maps, sectioned by condition, population, context and quality to highlight gaps in the evidence base.

Stage 2: panoramic meta-analysis

Data extraction

From the reviews identified in stage one, we selected reviews that contained quantitative data suitable for extraction. From these, we selected those reviews that were rated as 'moderate' or 'high' on the AMSTAR-2 checklist (henceforth referred to as 'higher-quality' reviews). Then we compared these reviews to identify if any review shared the same randomised controlled trial as another review. When we identified reviews that included the same randomised controlled trial, we chose (1) the review with the longest follow-up, (2) the review with the highest AMSTAR-2 rating, (3) the most recent review or (4) the review with the largest number of trials.

The primary analysis was conducted using the higher-quality reviews with suitable quantitative data. The analyses were conducted on the primary outcome of health-related quality of life and the secondary outcomes of depression, anxiety and pain.

Data synthesis

A panoramic meta-analysis using a two-step frequentist approach (random-effects model) was conducted with continuous end-point data for each outcome in Stata® versions 13 and 16 (StataCorp LP, College Station, TX, USA). The analysis produced a within- and across-condition I^2 heterogeneity statistic. If the heterogeneity (I^2) was < 75%, we proceeded to pool the estimates across (1) the within-condition reviews and (2) across the condition estimates. We produced standardised mean differences with 95% confidence intervals for all analyses. We also calculated prediction intervals for the primary analyses. For meta-analyses with > 10 reviews, we produced funnel plots and conducted Egger's test to detect publication bias and small-study effects. Next, we conducted subgroup analyses based on the ages of participants (children and adolescents, adults, older adults), cognitive–behaviour therapy intervention intensity (high, low), comparator groups (active, non-active) and duration of follow-up [short (< 12 months), long (≥ 12 months)]. We performed a sensitivity analysis that combined all of the lower-quality reviews (rated 'low' or 'critically low' on the AMSTAR-2 checklist) with the higher-quality review data.

To aid interpretation of these results, we transformed the standardised mean difference into an approximate mean difference on the most common outcome measure (e.g. Beck Depression Inventory for depression). To do this, we multiplied the overall pooled estimate for each outcome by the standard deviation of the outcome measure to produce an estimate of the mean difference for each measure. We identified the standard deviation from a trial, judged as having a low risk of bias, in a higher quality review.

Generalisation of the evidence

The expert consultation group helped form a list of pertinent questions regarding the generalisability of this evidence. The questions were as follows:

- Is there evidence of a general effect of cognitive–behaviour therapy across conditions?
- Is this effect robust across the conditions represented in each ICD-11 code?
- Is the effect robust across conditions that are represented by lower-quality reviews only?
- Is the effect robust across the populations and contexts we have tested?
- Can we infer that the effect might be observed across conditions that are not included in the current overview?

We drew on an established model of generalisation to guide the interpretation of the overview data.

Results

Mapping

We mapped 494 reviews (2052 trials, 221,128 participants). The most common reason for a review to be excluded was because it did not include a synthesis of the included cognitive–behaviour therapy trials or it did not fulfil the Centre for Reviews and Dissemination criteria to qualify as a systematic review. Ten per cent (237/2454) of the full-text reviews were excluded because full texts were not available in English.

Most of the included reviews (284/494, 57%) were published in the preceding 5 years (2015 onwards). Of the 494 reviews, only 142 reviews (29%) were rated as 'high' or 'moderate' on the AMSTAR-2.

The 494 reviews included 13 out of 20 ICD-11 mental condition categories and 14 out of 20 ICD-11 physical condition categories. 'Mood disorders' were the most researched condition (92 reviews, 272 trials, 42,676 participants). Most reviews considered the effects of high-intensity cognitive–behaviour therapy (397/494, 80%) delivered as a standard treatment (463/494, 93%) in the short term (402/494, 81%) in the adult population (378/494, 77%). Research with older adults was limited (30/494, 6%).

The effects of cognitive–behaviour therapy as a preventative intervention (29/494, 6%) or as part of relapse prevention (7/494, 1%) were under-researched. Reporting on condition severity (247/494, 50%) and the setting whence participants were recruited was also poor (283/494, 57%). Nearly half of the included reviews did not report details on sex (218/494, 44%) or the country where the trials were conducted (218/494, 44%), and the majority did not report the ethnicity of the participants (458/494, 93%). Only a very small proportion of reviews had included trials from the Asian, South American and African continents (45/494, 9%).

Panoramic meta-analysis

Of the 494 reviews, 71 (207 trials, 20,862 participants) were high-quality reviews with data suitable for inclusion in the panoramic meta-analyses.

Health-related quality of life

Estimates from 24 reviews (49 trials, 4304 participants) representing 10 different conditions demonstrated low heterogeneity ($I^2 = 32\%$). The analysis produced a modest effect in favour of cognitive–behavioural therapy (standardised mean difference 0.23, 95% confidence interval 0.14 to 0.33, prediction interval −0.03 to 0.50). This translates to an estimated mean change of 3 points on the Short Form questionnaire-36 items tool. No publication bias or small-study effects were identified ($p = 0.18$).

The sensitivity analysis found that the inclusion of an additional 10 lower-quality reviews increased the heterogeneity ($I^2 = 71\%$), but did not alter the effect estimates (standardised mean difference 0.28, 95% confidence interval 0.17 to 0.38). The effect was larger for cognitive–behavioural therapy compared with non-active comparator groups than for cognitive–behavioural therapy compared with active comparator groups. The interaction effect between these two types of reviews (active and non-active comparator groups) was statistically significant. None of the other subgroup analyses reported significant interaction effects between the groups.

All of the analyses from the primary, condition-specific, subgroups and sensitivity analyses produced effect estimates consistent with the general effect.

Depression

There was too much heterogeneity within and between conditions in the depression analyses; therefore, we did not pool any reviews together. No publication or small-study bias was detected ($p = 0.87$).

Anxiety

The heterogeneity across the 12 conditions represented by the anxiety analysis was acceptable ($I^2 = 62\%$). We pooled across 34 high-quality reviews (59 trials, 4673 participants) and identified a small effect in favour of cognitive–behavioural therapy (standardised mean difference 0.30, 95% confidence interval 0.18 to 0.43, prediction interval −0.28 to 0.88). This translates to an estimated mean change of 4 points on the Beck Anxiety Inventory. No publication or small-sample bias was detected ($p = 0.70$). All of the analyses from the primary, conditions, subgroups and sensitivity analyses produced effect estimates consistent with the general effect.

Pain

The heterogeneity of effect estimates generated for the outcome of pain, across abdominal, leukaemia-related, non-specific chest, osteoarthritis, spinal, back and neck pain was high, but acceptable ($I^2 = 64\%$). The overall pooled effect, from 10 high-quality reviews (22 trials, 2581 participants), was modest and in favour of cognitive–behavioural therapy (standardised mean difference 0.23, 95% confidence interval 0.05 to 0.41, prediction interval −0.28 to 0.74). The effect translated to a change of 6 mm on the 100-mm visual analogue scale. No publication or small-sample bias was detected ($p = 0.19$). All of the analyses from the primary, conditions, subgroups and sensitivity analyses produced effect estimates consistent with the general effect.

Generalisation

From our mapping and panoramic meta-analyses, we found that cognitive–behavioural therapy produced a general effect of improving health-related quality of life across different conditions.

The effects we found remained consistent across all conditions tested and when considering the broader number of health conditions represented by comorbidities of these patients. We suggested that this effect was robust across the conditions represented in the ICD-11 primary (physical conditions) and secondary (mental conditions) codes.

The consistency of the general effect leads us to suggest that it is robust across the populations (age, sex) and contexts (health-care setting, intervention delivery/timing, condition severity) that have been represented by this overview. We are less sure of the consistency of the effect across ethnic groups, as this was poorly reported. Nor are we sure of the effect in countries in Africa, Asia and South America, as these were under-researched.

The debate of whether or not the general effect can be generalised across conditions that are not represented in this overview remains contentious. There is no evidence to suggest that cognitive–behavioural therapy would not be effective or would be harmful. The expert consultation group did not reach agreement that cognitive–behavioural therapy effects change through shared mechanisms for every condition: some members of the group felt that additional detailed information on mechanistic data would be needed to make broader generalisations, but several members of the investigator team felt that it was sufficient that the statistical and remaining principles of generalisation had been met. Therefore, it remains uncertain if the effect would be replicated in conditions not represented in this overview.

Conclusion

The best-quality evidence available has estimated that cognitive–behavioural therapy produces a general improvement in health-related quality of life and reduces specific contributing symptoms (pain and anxiety). The effect is observed when cognitive–behavioural therapy is delivered via high- or low-intensity formats and is evident when data are collected > 12 months after a patient has received cognitive–behavioural therapy. The effect becomes much smaller when cognitive–behavioural therapy is compared with active comparators such as pharmacotherapy, relaxation or exercise therapy. However, we did not identify any condition for which there was evidence in favour of the comparator group (i.e. a statistically significant effect in favour of the comparator). Cognitive–behavioural therapy has been tested in participants with 22 different conditions. Given that there is no condition for which it has been demonstrated that there is no benefit of cognitive–behavioural therapy, cognitive–behavioural therapy is likely to work across most, if not all, conditions. However, some of our expert consultation group were not in agreement that we can make this generalisation across conditions not represented in the overview without more evidence on the mechanisms of how cognitive–behavioural therapy effects change.

We suggest that this effect is applicable to children, adolescents and adults, who are male or female, living in Europe, North America and Australasia. We recommend that future research examines (1) if ethnicity can moderate the effectiveness of cognitive–behavioural therapy, (2) if older adults experience the same effect as adults and children/adolescents, (3) the preventative use of cognitive–behavioural therapy and (4) targeting reviews published in languages other than English to try and identify evidence from countries in Africa, Asia and South America.

The main limitation of the panoramic meta-analyses is that we extracted and analysed data at the review level. Many reviews synthesised cognitive–behavioural therapy randomised controlled trial evidence in combination with other therapies or types of study design. We were often unable to extract

the purely cognitive–behavioural therapy randomised controlled trial data in isolation; therefore, we could not use the data from that review. If we had been able to return to the randomised controlled trial data sources, then we could have included the individual randomised controlled trials in the panoramic meta-analyses, but this was beyond the scope of this study.

Study registration

This study is registered as PROSPERO CRD42017078690.

Funding

This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 25, No. 9. See the NIHR Journals Library website for further project information.

Health Technology Assessment

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 3.370

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, the Cochrane Library and Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 15/174/24. The contractual start date was in January 2018. The draft report began editorial review in April 2020 and was accepted for publication in July 2020. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).