Protocol

Cognitive Behavioural Therapy: An overview of systematic reviews and meta-analyses

In accordance with the guidelines, our overview of systematic reviews protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) on 10th October 2017 (registration number: CRD42017078690).

Beth Fordham [1], Sally Hopewell [1], Karla Hemming [2], Jeremy Howick [3], Shona Kirtley [1] and Sallie Lamb [1]

Corresponding author: Beth Fordham beth.fordham@ndorms.ox.ac.uk

Author affiliations: [1] Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Windmill Road, Headington, Oxford OX3 7LD

[3] Public Health, Epidemiology and Biostatistics, Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT3[4] Department of Primary Care Health Sciences, Centre for Evidence-Based Medicine, University of Oxford, New Radcliffe House, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK

Email: Sally Hopewell sally.hopewell@csm.ox.ac.uk , Karla Hemming K.Hemming@bham.ac.uk, Jeremy Howick jeremy.howick@phc.ox.ac.uk, Shona Kirtley shona.kirtley@csm.ox.ac.uk and Sallie Lamb sarah.lamb@ndorms.ox.ac.uk.

Contributions: BF is the guarantor. SH provided expertise on design and methodology. KH provided expertise on methodology in particular panoramic meta-analysis. JH provided expertise in methodology especially generalisation framework. SK developed the draft search strategy plan and SL provided expertise in all areas of design, methodology and rigour. All authors read, provided feedback and approved the final manuscript.

Support: This overview of systematic reviews is funded by the National Institute for Health Research, Health Technology Assessment Programme (funding reference number HTA 15/174: Cognitive Behavioural Therapy: An overview of systematic reviews and meta-analyses).

The Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (University of Oxford) is the sponsor of this overview of systematic reviews.

The NIHR HTA programme is funding this overview of systematic reviews. They will provide funding to NDORMS (University of Oxford) who will carry out the research and store the data. The NIHR HTA programme is not involved in any other aspect of the project design, data collection or analyses. They will be made aware of the findings at milestones throughout the project.

Table of Contents

Table of Contents

Introduction

Cognitive Behavioural Therapy (CBT) is an amalgam of interventions which emerge from theoretical cognitive and behavioural psychological models [1]. Interventions can look different across disorders yet a common therapeutic style is recognised [2] and they aim to help recipients' link unhelpful thoughts, emotions and behaviours and to use techniques such as guided discovery to resolve these [3-5]. The successful interventions commonly aim to identify core cognitive factors which maintain the disorder's symptoms and change them [6].

There are many variants of CBT including high versus low intensity. High intensity is formal CBT with a trained health professional predominantly delivered face to face in an individual or group format. Low intensity interventions focus on patient self-help and can be delivered by health professionals with brief CBT training and via several platforms (internet, phone, paper-based). The distinction can become less clear in some forms of CBT where high intensity therapy is combined with low intensity

self-help methods. The majority of CBT is delivered in adherence with CBT process manuals specific to the indication and are technique-driven.

There are further distinctions. High intensity CBT can be delivered via different clinical approaches depending upon the theoretical model being implemented for that patient/group. For example within Schizophrenia a recent meta-analysis has begun to distinguish between different CBT approaches in reducing delusions [7]. The review found that CBT which primarily targeted factors hypothesised to form a patient's delusions were slightly more advantageous than CBT approaches which solely focus on the delusions themselves.

CBT was primarily developed for adult mood disorders but clinical experience and evidential support has expanded its scope. It is now used across a broad range of physical and mental health indications, patient groups and settings. The intervention has a clear theoretical model of action which lends itself for empirical testing. Consequently, CBT was evaluated in trials since early 1980's. There are thousands of randomised controlled trials (RCTs) and consequently hundreds of systematic reviews/meta-analyses examining the effectiveness of CBT across multiple health indications and/or health behaviours.

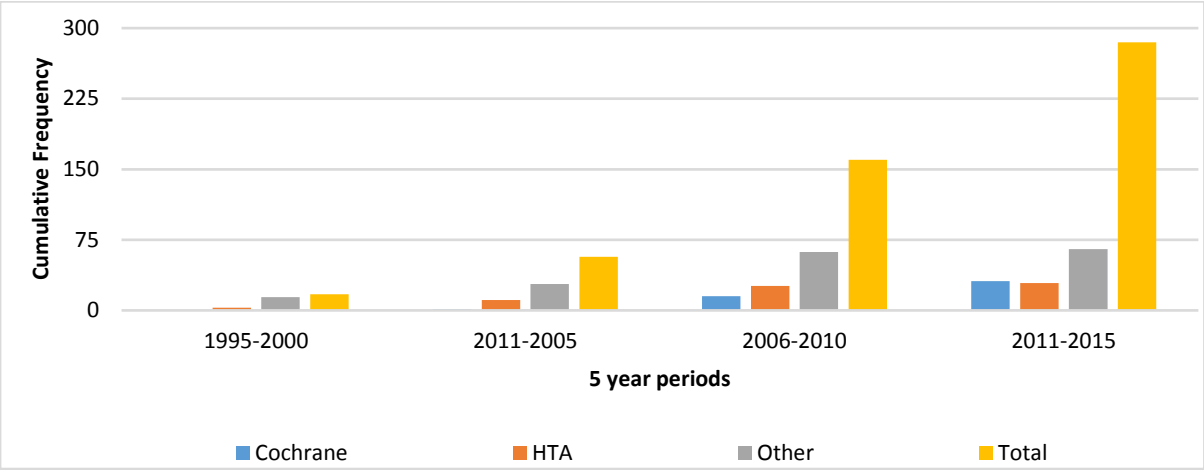
In 2014, the Economist published an article detailing that 43% of all therapy courses in Britain are CBT and how over 6000, new therapists were trained between 2007 and 2014 [8]. In 2013 to 2014 the NHS Improving Access to Psychological Therapy service reported 1,145,957 CBT face to face and 43,323 computerised CBT appointments were attended [9]. The majority (28%) of expenditure on psychological treatment research between 2008 and 2013 was spent on CBT (£21.54 million) [10]. Some researchers have argued that CBT is in a 'virtuous circle: money pours into research, evidence accumulates, more financial support is given to… [CBT]… and other forms of psychotherapy are excluded' (Peter Fonagy, UCL [8]).

In order to achieve a perspective of the scale of CBT research we ran scoping searches for RCTs examining CBT effectiveness in EMBASE and this yielded 6092 (un-sifted) hits. We recognise that, several NICE clinical guidelines have reviewed contemporary evidence [3,4] but only focus upon one specific indication rather than the generic effects of the health technology.

In 2012, a smaller scale review of CBT meta-analyses across indications was conducted [11]. This review identified 269 meta-analyses published between 2000 and 2011. This study grouped the findings into 16 categories: addiction and substance use disorder, schizophrenia and other psychotic disorders, depression and dysthymia, bipolar disorder, anxiety disorders, somatoform disorders, eating disorders, insomnia, personality disorders, anger and aggression, criminal behaviours, general stress, distress due to general medical indications, chronic pain and fatigue, pregnancy complications and female hormonal indications and special populations (children and elderly). The review summarised treatment effects within these indications separately, but made no attempt to pool or compare effects over different indications, address the issue of generalisability and source of variation of effects, and areas with a lack of evidence. This review provides a good estimate of the number of reviews and the number of different indications our review could include.

We searched the Centre for Research and Dissemination Database of Abstracts of Reviews of Effects (DARE) (from inception to March 2015 when the database was no longer updated) and have mapped the cumulative frequency of systematic reviews on CBT produced by Cochrane, HTA and other sources, please see Figure 1. There has been an increase in the number of systematic reviews of CBT within each 5 year time period.

Figure 1: Cumulative frequency of CBT reviews (20 years)



Amongst the reviews published in 2014-2015 and found, as expected, that the majority (67%) examined mental health indications (n=34) and the remaining 33% were evenly split between physical health indications (n=10) and maladaptive behaviours (n=7), please see Figure 2.

Figure 2: 2014-2015 Topic distribution of CBT systematic reviews

The majority of the mental health reviews focussed on mood disorders (depression and anxiety) across various populations, however they did span other mental health indications such as schizophrenia and specific symptoms such as fear of falling in older adults. The physical health indication systematic reviews commonly focussed on treating co-morbid mood disorders e.g. depression in people with chronic obstructive pulmonary disease but did also primarily focus on physical indications, such as insomnia and physical symptoms such as pain and sleep. The behavioural reviews either focussed solely on behavioural outcomes such as adherence to medication but predominantly focussed on maladaptive behaviours which have been classified as mental health indications such as addictive gaming, substance misuse and abusive violent behaviours.

This snapshot of the available evidence base suggests that we will find the most comprehensive evidence within mental health indications and it might be possible to generalise generic effects from mental health domains across to co-morbid symptoms within physical health indications. For example if we find comprehensive evidence that CBT improves symptoms of feeling depressed whether this is depression by itself or as a co-morbidity to living with a physical health indication, we can examine whether this effect can be generalised to all co-morbid depression in all physical health indications.

The vast literature leaves a confusing picture to communicate to patients, the general public, clinicians, researchers and research funders. There are two key needs:

(1) The need for a systematic and a complete synopsis of the current evidence. Developing evidence map and a 'gap map.'
(2) To understand the extent to which the current evidence could suggest generic effects that might be generalised to all or a selection of other indications or between different formats of CBT.

A huge amount of funding and clinical hours are invested in CBT interventions. In providing a comprehensive cross section of the available evidence we can provide a clear picture of where CBT should or should not be used clinically and where we do or do not need to fund further research, potentially allowing a fairer and more evidence based distribution of funds and resources.

Aim

The overarching aim is to identify for which populations there is / is not sufficient evidence (i.e. many good quality RCTs synthesised into good quality systematic reviews) to recommend CBT as clinically effective or not. In addition, we will explore the potential for generalising existing evidence of effects to other indications where there is currently insufficient systematic review evidence. Finally we will recommend where further research would be beneficial (e.g. in primary trials or indication specific reviews, indirect comparisons).

Objectives

The objectives will be to complete:

(1) A systematic evidence mapping exercise to identify systematic reviews of CBT treatment effectiveness/harm on generic and indication-specific outcomes, intermediary (mechanistic) variables, and indication-specific outcomes). The indication categorisation will remain at as lower level of granularity as possible within the ICD-10 classification. Within this system an example of the primary level is 'Mental, Behavioral and Neurodevelopmental disorders" and an example of the secondary level would be 'Mood [affective] disorders" and tertiary would be "Major depressive disorder, recurrent." We shall aim to group reviews by secondary level classification however we will proceed to tertiary level categorisation if the evidence dictates. We shall also indicate if a review is up to date (published within past 5 years) or not.

(2) A systematic review of systematic reviews and where appropriate the synthesis of pooled estimates in a "panoramic meta-analysis". Where sensible, we will report the consistency of effect sizes and pooled effect sizes within ICD-10 categories and/or broad indications and where outcomes are sufficiently similar to allow pooling or comparison. For example, we will seek to determine if the effect of CBT in painful physical indications is consistent over a range of ICD-10 categories or sub-categories in for example, pain measures. Also, for example, whether intervention on depression renders a similar effect regardless of underlying indication.

(3) An assessment of the generalisability of the available evidence. We will seek expert and patient opinion (Expert Consultation Group) on indications that are sufficiently similar to make sensible and meaningful comparisons. This will include considering where existing evidence might be generalised legitimately into areas without evidence (for example, could evidence of CBT effectiveness in Osteoarthritis of the Knee be indicative of a class effect on all Osteoarthritis pain). This will involve some additional literature searching for evidence relating to the mechanisms of action of CBT. We will use an established framework model to investigate the extent to which generalisation of evidence can be justified across indications where evidence is currently lacking

(4) To explore major variants of the application of the CBT that will make a meaningful difference to the costs of health service provision. Specifically low versus high intensity CBT.

(5) Throughout we will use expert and PPI consultation, using a recognised nominal group technique to ensure an appropriate balance between external advice and unwarranted influence and bias being introduced into the selection, grading, presentation and interpretation of evidence.

Methods

Eligibility criteria

To be included in the evidence map and overview of systematic reviews, studies must meet the following criteria:

Type of studies

We will include systematic reviews of RCTs which evaluate the effects of CBT. We will include systematic reviews which include both randomised and non-randomised trials so long as the randomised evidence can be extracted independently from the review. To be included, systematic reviews must fulfil a minimum of 4 methodological criteria as defined by the Centre for Reviews and Dissemination, University of York, as part of the Database of Reviews of Effects (DARE) database (http://www.crd.york.ac.uk/crdweb) [12]: (1) inclusion/exclusion criteria reported; (2) adequate search strategy; (3) included studies synthesised; (4) quality of the included studies assessed; (5) sufficient details about the included studies reported.

Type of participants

We will include systematic reviews of randomised trials, which include data from all age groups including children, adolescents and adults. Within these age categories participants can be male or female. We will include all indications recognised within the International Classification of Diseases manual version 10 (ICD-10) and their alternative nominal categorisation. For example ICD-10 uses the term 'mental retardation' whereas research papers may refer to 'intellectual disability.' The participants included in these RCTs will have participated in either a control or intervention arm examining the effectiveness of CBT in any format (high or low intensity delivery). Details of the target population will be provided in the data summary presentations and a priori subgroup comparisons have been planned based on indication type and age category.

Setting

We will include systematic reviews of randomised trials that have been conducted in any context including community settings, NHS primary care, private practice, institutions, residential care etc and across any country. The setting/context information will be included in our data summary presentation however no a priori subcategory analyses have been planned to compare between settings.

Intervention

We will include systematic reviews of RCTs which examined an intervention which meets our working definition of CBT. This may be amended in light of the review from the consultation group but at present it stands as:

> 1) The intervention is explicitly or implicitly based on the CB model (where the use of CBT/CB in relation to the intervention is explicitly stated OR where the connection between thoughts, feelings and behaviours in relation to the intervention is implicitly described); and
>
> 2) Uses specific techniques to both change cognitions and change behaviours.'

We will include an examination of manualised CBT predominantly, but systematic reviews may also include modular CBT. We will include high and low intensity CBT review. High intensity will be defined as using a highly specialist trained CBT therapist and low intensity is all other types of CBT. No a priori subcategory analyses have been planned to compare between high and low intensity CBT.

Comparator

We will include systematic reviews if they explore comparisons of CBT to either: 1) a non CBT comparator intervention (e.g. other psychological, behavioural, pharmacological interventions, placebo CBT (attention control), 2) no intervention or treatment as usual or (3) Another format of CBT (i.e. computerised CBT versus face to face). No a priori subcategory analyses have been planned to compare between different types of comparators.

Outcomes

We will include systematic reviews which report information on at least one of the following primary outcomes:

- patient reported outcome symptoms (cognitive (e.g. fear avoidance beliefs), emotional (e.g. Hospital Anxiety and Depression Scale), behavioural (e.g. obsessional behaviours questionnaire) or physical (e.g. daily calorie intake)
- and/or HR-QoL outcomes (e.g. EQ-5D).

If a review uses recovery rates then we will use the symptoms from which the patients are recovering from. We will also include reviews which synthesis RCTs exploring proposed mechanisms of action such as brain imaging, cognitive changes or cortisol levels.

We will also extract data on secondary outcomes including patient satisfaction, adverse events and economic analyses.

We will only include reviews with outcomes of in medium (6-12 months) to long term (>12 months) effectiveness. We will exclude those reporting <6 months duration.

Restrictions

There will be no language restrictions and we will search and include grey literature (unpublished). We will also identify ongoing reviews by searching Prospero and if necessary contacting the relevant authors.

Information sources

Our method of identifying systematic reviews will be conducted in line with the Cochrane Handbook for Systematic Reviews of Interventions [13] and recommendations for conducting Overviews of Systematic Reviews [14].

A comprehensive search strategy will be designed by an experienced information specialist (Kirtley) and will be informed by the consultation group, and index terms identified in key papers from our preliminary scoping searches of systematic reviews on CBT. Our preliminary search plan is attached in Appendix A. Once the search has been run as a sample search we will perform a sensitivity check to see if the key papers (10-20 papers), as identified by the consensus group, have all been picked up, if not the search will be revised accordingly.

The finalised search strategy will be adapted for use across our selected databases and tested for sensitivity with any required final revisions made. The search strategies will be available in the final protocol. We will explore the different systematic review search filters available from the McMaster Hedges Project Group and those available through the InterTASC Information Specialists' Sub-Group (ISSG) to enable us to select the best and most appropriate filter to specifically identify reports of systematic reviews. Many of these search filters have been adapted for use across multiple electronic databases [15] .

The final robust set of individual search strategies will be run across the Database of Abstracts of Reviews of Effects (DARE: up to March 2015) the Cochrane Library of Systematic Reviews, MEDLINE, EMBASE, PsychInfo and CINAHL from 1970 (inception of CBT) to present day. We plan to also explore (and then consult our consultation group) whether to additionally search the following databases (depending upon database scope/relevancy, range of journal titles indexed, inclusion of systematic reviews): Allied and Complementary Medicine (AMED); SCOPUS; Applied Social Sciences Index & Abstracts (ASSIA); Child Development & Adolescent Studies; Sociological Abstracts; Web of Science; Educational Resources Information Center (ERIC); Physiotherapy Evidence Database (PEDro); Global Health; LILACS (Latin America); OpenGrey.

We would like to capture reviews of CBT published at any time point, however, we will note if a review was published more than 5 years ago as this might need updating. We will also search PROSPERO to identify upcoming reviews which are due to be published within our study timeframe. There will be no language restrictions or publication status restrictions. We will check reference lists of relevant systematic reviews. We will also identify ongoing and unpublished reviews by contacting researchers in the field.

We will perform an update search 12 months after the initial searches have been run to check for any additional systematic reviews which have been published in the intervening months. We will also search ClinicalTrials.gov and the International Clinical Trials Registry Platform (ICTRP) for ongoing trials and reviews which may be exploring the areas we recommend for future research.

Study records

Search results will be exported into Endnote for de-duplication. Once any duplicate records have been excluded, the search results (title and abstract) will be exported into Covidence, which is a web-based software tool for study screening and data extraction of systematic reviews and is recommended by Cochrane [16].

The team will develop and test screening forms for the initial abstract sift and data extraction based on the inclusion criteria. Citation abstracts and full text articles will be uploaded together with their screening questions onto Coevidence. All members of the review team will be trained in how to use Coevidence prior to starting work on the review.

Selection process

Two reviewers will independently screen all titles and abstracts for potentially eligible reviews, using the finalised screening questionnaire which will be based on the predefined eligibility criteria. We will obtain full-text reports of those selected for inclusion or for any uncertain cases. The same two reviewers will independently perform study selection. We will seek additional information from study authors where necessary to resolve questions about eligibility. We will resolve any disagreements regarding the inclusion or exclusion of individual studies by discussion or, if necessary, will consult a third reviewer. The search process and study identification will be documented in a figure as recommended by PRISMA statement [17]. This will results in a final list of included and excluded systematic reviews along with reasons for exclusion. This process will not be blinded so all reviewers will be able to see the authors and their affiliated institutions.

Data collection process

Data on the quality of the systematic review, the evidence base included in the review, overall estimates of effect and contextual information will be extracted by two independent reviewers using a bespoke data extraction form. The duplicate extractions will be compared and the anomalies discussed with the third reviewer (BF) until a decision is reached. Authors of the papers will be contacted if there are unresolved uncertainties.

Data items

The information extracted for each review will include (but is not necessarily limited to):

- Review author and date of publication
- Total number of included RCTs
- Number of participants in the CBT and control intervention arms
- Characteristics of the participants (e.g. gender, ethnicity, residential status)
- Clinical symptom outcomes and HR-QoL measure.
- If meta-analysis has been conducted we will extract, for each relevant outcome;
- The number of RCTs the result was based on
- The overall pooled treatment effect estimate with 95% confidence intervals
- The amount of heterogeneity (e.g. the 1^2 statistic) and GRADE assessment if completed
- We will also extract information on the methods of assessing risk of bias (e.g. Cochrane Risk of Bias) and a summary of the overall assessment.
- If reported, we will also record pooled estimates for proposed mechanisms of action.
- Detailed information on the CBT intervention will include for
    - 'which indication': ICD-10 classification, co-morbidities, severity of indication;
    - 'who' age-group;
    - 'when' if CBT is used as prevention, early intervention or standard treatment option;
    - 'where' community setting, inpatient setting or outpatient setting; and
    - 'how' CBT was delivered high intensity (e.g. individual or group based) or low intensity (e.g. computerised, phone, book).

If reported we will extract secondary outcome information such as acceptability, satisfaction, adverse events and economic analyses. We will also flag if this review was published within the last 5 years. This will enable us to identify areas where possible update reviews are needed.

Outcomes and prioritisation

This overview will prioritise mid to long term changes in health related quality of life (e.g. HR-QoL) and patient reported symptoms. The symptom outcomes will be categorised into cognitive (e.g. fear avoidance beliefs), emotional (e.g. Hospital Anxiety and Depression Scale), behavioural (e.g. obsessional behaviours questionnaire) or physical (e.g. daily calorie intake). If a review uses recovery rates then we will use the patient reported outcome symptoms from which the patients are recovering from.

As a secondary focus we will examine mechanisms of action, patient satisfaction, adverse events and economic outcomes.

Assessment of methodological quality of included reviews

Each systematic review will be assessed independently by the two reviewers using the AMSTAR [18] tool for the methodological quality assessment for systematic reviews. We will not reassess the quality of the individual included studies but rely on the author's assessment. We will resolve any discrepancies by consensus and when agreement cannot be reached a third overview author (BF) will consider the paper and make a majority decision. The tool consists of 11 items and has good face and content validity for measuring the methodological quality of systematic reviews. The answer to each question plus the overall score out of 11 will be recorded on the data extraction sheet.

Criteria	Specific requirements
1. Was an 'a priori' design provided?	The research question and inclusion criteria should be established before the conduct of the review.
2. Was there duplicate study selection and data extraction?	There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.
3. Was a comprehensive literature search performed?	At least two electronic sources should be searched. The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words or MESH terms, or both, must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers or experts in the particular field of study, and by reviewing the references in the studies found.
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?	The review authors should state that they searched for reports regardless of their publication type. The review authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.
5. Was a list of studies (included and excluded) provided?	A list of included and excluded studies should be provided.
6. Were the characteristics of the included studies provided?	In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analysed e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity or other diseases should be reported.
7. Was the scientific quality of the included studies assessed and documented?	'A priori' methods of assessment should be provided (e.g. for effectiveness studies if the review author(s) chose to include only randomised, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.

8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.
9. Were the methods used to combine the findings of studies appropriate?	For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi² test for homogeneity, I² statistic). If heterogeneity exists a random-effects model should be used or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?), or both.
10. Was the likelihood of publication bias assessed?	An assessment of publication bias should include a combination of graphical aids (e.g. funnel plot, other available tests) or statistical tests (e.g. Egger regression test), or both.
11. Was the conflict of interest stated?	Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

AMSTAR review quality assessment tool

Data synthesis

Our proposed method of data synthesis presented here may be altered in response to our first expert consultation group meeting held in March 2018. Specifically they will advise on

(1) The most user friendly method to display the data synthesis for user, clinical and research application
(2) Developing the combined threshold rating system (see '**mapping the evidence by subgroup'**)
(3) Populating the framework for generalisation (see '**generalisation framework'**)

Summary tables

The information from each eligible systematic review will be entered into summary tables detailing the characteristics on the included studies and assessment of the overall quality. The descriptive table will include: indication type as defined by the ICD-10 (e.g. recurrent depressive episode), population type (e.g. adults with no co-morbidities specified), CBT type/approach (e.g. computerised CBT or CBT for psychosis with causal interventionist approach), the number of RCTs included in the review, the number of participants included, the pooled effect size or relative risk ratio, the AMSTAR rating, the original method of quality assessment for RCTs (e.g. Cochrane Risk of bias) and the score/ rating from the combined threshold rating system(e.g. Unclear, Low, Moderate, High). This will be accompanied with a written description of the included reviews.

During our scoping reviews we identified a few systematic reviews of the mechanisms of CBT action. We will provide a written section detailing the mechanisms which we find. These can be used to inform the generalisation process.

Mapping the evidence by indication type

First we will produce an evidence map and present all of the indications, as classified by the ICD-10, which are covered by the systematic reviews we have included. We will use a Bubble map [19] as this can represent four dimensions: effect size, mental/physical health indication, number of RCTs and number of participants. We will use the y axis to represent the number of participants included across all of the systematic reviews, the x axis to represent the trend or average (if pooled) effect size, the size of the bubble to represent the number of systematic reviews identified and the colour to represent if the indication is physical or mental health indication.

Mapping the evidence by subgroups

Next, we will then list all of the indications (ICD-10 classified) which are presented in the Bubble map into the left column of a table. The top row will categorise the sub-groups under which CBT has been evaluated in the systematic reviews included in our review.

From our scoping review we suggest the following categories to be included: 'severity' mild, moderate and severe, 'who' adults, children and adolescents, older adults, adults with intellectual disabilities, 'how' high intensity (e.g. individual or group), low intensity (e.g. phone, internet, book etc). This map will enable us to see in which subgroups the research has been saturated, where there are few reviews and where there is a lack of reviews.

We will use a combined threshold rating system, developed with the consultation group, based on reported effect sizes and confidence intervals, quality of the systematic reviews (AMSTAR) and size of the evidence base (n. of participants included).

Mapping the evidence by outcome measure

We will develop a table which can show where there is strong, moderate or weak evidence for a benefit of CBT across symptoms which are common across more than one indication.

From examining a selection of reviews in our scoping search we identified outcomes from all five of our proposed categories: emotional, cognitive, behavioural, physical and HR-QoL. We propose to remain at a low level of symptom granularity in order to identify generic effects across the indications.

We will only include an outcome if it is reported in more than one of the indications from our included systematic reviews. If a systematic review reports recovery rates then we will include whatever the primary outcome was for 'recovery' to be reached. We propose to use the same traffic light system of evidence strength (green=strong, amber=moderate and red=weak), as used in the subgroup map (Figure 5). However, our outcome map will only include evidence which demonstrates a beneficial effect upon outcomes, consequently, we will not need to use the '+, 0, -' symbols. If there is no evidence then the cell will be left blank. If the cell is not applicable i.e. disease specific physical markers in mental health indications then the cell will be marked 'N/A.'

Forest plot synthesis

We will then display results from systematic reviews into forest plots using the lowest level of health indication granularity. For physical health we shall use the ICD-10 primary categories e.g. 'IV: Endocrine, nutritional and metabolic diseases' and for mental health we shall use the secondary categories e.g. 'V: [F30-F39] Mood disorders.'

Within these forest plots we shall list all of the relevant systematic reviews, separated with sub-headings to describe their specific population. The forest plots will detail either the effect size

(standardised mean difference) or the odds ratio with 95% confidence intervals, AMSTAR rating, number of studies included in the review and number of participants included overall.

Panoramic meta-analysis (PMA)

If a subset of systematic reviews is deemed to have moderate clinical, design and statistical homogeneity (I^2 less than 75%) we will synthesise these reviews and provide pooled (across indications) treatment effects.

Statistical heterogeneity in treatment effect estimates between indications will be explored using the I-squared statistic; clinical heterogeneity will be explored through discussion with the consultation group (see 'Section one: stage 1') using the generalisation framework (see 'Section one: generalisation framework'); and design heterogeneity explored using AMSTAR scores. Where reviews exhibit considerable statistical heterogeneity across indications (I^2>75%) results will not be pooled. For indications that have moderate to low statistical and design heterogeneity (I^2 less than 75%); and clinical plausibility for a common mechanism for action (generalisation), we will pool across indications. If these requirements are met, we will make the realistic assumption that the treatment effects are exchangeable.

Once we have a set of systematic reviews which meet out requirements for PMA we will assess if they include the same primary studies. If two reviews include the same primary studies, we will choose to include based on ordered preference criteria:

- The availability of numerical data
- Highest AMSTAR score
- Most recent date
- Larger number of studies included

A formal quantitative data synthesis will be undertaken using a two-step frequentist approach to a PMA. This method provided a single pooled estimate of the treatment effect, over all reviews, along with estimates of degree of heterogeneity between reviews. This allows for both between study variability (if random effects meta-analysis was used in the original indication review) and between indication variability (using random effects), but does assume exchangeability of treatment effects. Evidence of funnel plot asymmetry will be assessed using both the funnel plot and the Egger test using a conservative P-value of 0.1 to acknowledge the low power of this test.

Generalisation framework

To explore the question of whether generic CBT effects (established for certain populations/indications) could be generalised across to other populations/indications we will generate a generalisation framework. We will use a generalisation framework based on that developed by Howick, Glaziou and Aronson but adapted for CBT with our consultation group and lead by Howick. The framework uses a set of requirements for the clinical homogeneity necessary to justify proposed generalisation across:

- populations
- indications
- contexts

We will highlight that the recommendations from generalisation will carry less weight than recommendations supported by direct systematic review evidence.

Syntheses we will not conduct

From the evidence identification process, we recognise that additional analyses could be performed. For example, if we have reviews of high intensity CBT versus control and others of low intensity versus

control then we could use multiple treatment comparison analyses to generate an estimate of high versus low intensity CBT. Similarly, with reviews of CBT versus pharmaceutical active treatment and CBT versus other psychotherapy active treatment, we could identify which out of the three options was the most beneficial. We felt these types of analyses were beyond the scope of the call and that the analysis plan which we propose is ambitious yet feasible to complete within the 24-month project timeframe. However, if these further analyses were required then this project could provide the identification of the relevant systematic reviews on which it could be performed.

Economic evaluation

We do not plan to specifically search for economic systematic reviews. However, if there are economic analyses embedded within the reviews which we include then we will extract the basic descriptive information. We will provide a written summary of this evidence but no further analysis. We will highlight evidence from the main overview synthesis which could impact future economic analyses.

Protocol development process

To answer our study objectives we are continuing to develop our overview methodology with the guidance of the expert consultation group. This group consists of CBT service users, clinician and research experts from specialities including child and adolescent mental health, minority communities and severe mental health fields. The consultation group facilitators will use a modified nominal group technique (NGT) [20] to run the meeting in order to achieve a degree of agreement within the two full-day meetings. We are not aiming to achieve full consensus across all consultation group members but a workable position from which to achieve this project's outcomes. The expert consultation group will meet in February 2018 and will be involved in:

(1) Refine working definition of CBT
(2) Reviewing the search strategy plan (see **'Appendix A'**)
(3) Reviewing methods to map and synthesise the data
(4) Review the data extraction sheet
(5) Developing the combined threshold rating system (see **'Data synthesis: mapping the evidence by subgroup'**)
(6) Populating the framework for generalisation (see **'Data synthesis: generalisation framework'**)

After the protocol has been updated in alignment with the decisions from the consultation group, we will publish a finalised protocol.

<u>Summary points</u>

We are sensitive to the importance of not conclusively representing CBT as being effective or not and to accurately reflect where further research, whether exploratory or secondary analysis work is needed. We will caveat all summary statements and recommendations with the limitations of the methodology but treat this as a necessary step in addressing the current state of the CBT evidence base.

The indication Bubble plot could show us where CBT has been evaluated, thereby indicating the remaining ICD-10 indications have not been examined with a high quality systematic review.

By using the generalisation framework, we can justify where we could generalise evidence to fill some of these evidence gaps. For example, we may find moderate to strong evidence in support of the benefit of CBT upon outcomes in *a, b and c* and no evidence evaluating the effect if CBT in *x, y and z*.

If the consultation group agree that *a, b, and c* are homogenous with *x, y and z*, then we can justify a recommendation that CBT is beneficial for *a, b, c and x, y, z.* The evidence statement for CBTs clinical benefit is weaker for *x, y, z* than for *a, b, c.* This finding, however, could suggest that future primary research be directed to areas where there is an evidence gap and where there is no justification for generalisation.

We will search ClinicalTrials.gov and ICTRP to identify on-going trials or systematic reviews which have addressed the areas we recommend for further research.

Guided by the consultation group we will present the evidence to answer the questions from the original research question. We will produce a list of ICD-10 indications and simply indicate if there is

(a) Systematic reviews conducted within this indication within the past 5 years and what the effect is. These areas will be subdivided into who, what, where and when and the effect will be categorised as:
    a. Beneficial
    b. Harmful
    c. No effect
(b) Areas where we can justify proposing to generalise clinical effectiveness to.
(c) Identifying an evidence gap. Areas where we cannot justify generalisation and the evidence base is too weak to determine CBTs clinical effects because:
    a. Quality of the evidence is poor
    b. Reviews have not been updated in the last 5 years
    c. There is no evidence

This summary will lead to a set of recommendations regarding:

1. No further primary research into areas where the clinical effectiveness evidence base is strong or where we propose generalising the findings.
2. Prioritising primary research into areas where we cannot generalise the clinical effectiveness findings and the evidence base is weak.

The final evidence statements and recommendations will be determined by what we find and what the consultation group decide, these predictions are based on an ideal project where we can answer all of the call's research questions however we appreciate this may not be possible.

References

1. Grazebrook KaG, A. What are cognitive and/or behavioural psychotherapies. UKCP/BACP Mapping psychotherapy exercise. . 2005.
2. Mansell W. The seven C's of CBT: A consideration of the future challenges for cognitive behaviour therapy. *Behavioural and Cognitive Psychotherapy.* 2008;.36(6):pp.
3. NICE. Depression in adults: recognition and management. . 2009.
4. NICE. Diabetes (type 1 and type 2) in children and young people: diagnosis and management. http://www.nice.org.uk/guidance/ng18 2015.
5. Roth A, Pilling S. *The competencies required to deliver effective Cognitive and Behavioural Therapy for people with Depression and Anxiety disorders. .* London: Department of Health.; 2007.
6. Beck AT, Rush, A. J., Shaw, B. F. & Emery, G. *Cognitive Therapy of Depression. .* New York: Guilford Press. ; 1979.
7. Mehl S, Werner D, Lincoln TM. Does Cognitive Behavior Therapy for psychosis (CBTp) show a sustainable effect on delusions? A meta-analysis. *Frontiers in psychology.* 2015;6.

8. Economist. Expanding the Shrinks. *Economist.* July 2016, 2014.
9. HSCIC. Psychological therapies annual report on the use of IAPT services – England, 2013–14 reference data tables. . 2014. Accessed 05 January 2016.
10. Kirtley A. Preliminary Analysis of the UK Mental Health Research Funding Landscape 2008–2013. . Accessed 05 January 2016.
11. Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. *Cognitive therapy and research.* 2012;36(5):427-440.
12. CRD. *Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for Carrying Out or Commissioning Reviews (CRD Report 4) 2nd edn. .* York: NHS Centre for Reviews and Dissemination, University of York.; 2001.
13. Higgins JP, Green S. *Cochrane handbook for systematic reviews of interventions.* Vol 5: Wiley Online Library; 2008.
14. Cochrane. The Cochrane Comparing Multiple Interventions Group

Accessed 6th Jan 2016.
15. ISSG. Filters to Identify Systematic Reviews. https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/filters-to-identify-systematic-reviews. Accessed 3 August, 2016.
16. Covidence. Covidence: Faster systematic reviews. https://www.covidence.org/. Accessed June, 2016.
17. Moher D, Liberati A, Tetzlaff J, Altman DG, Grp P. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Plos Med.* 2009;6(7).
18. Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology.* 2009;62(10):1013-1020.
19. Miake-Lye IM, Hempel S, Shanman R, Shekelle PG. What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products. *Systematic reviews.* 2016;5(1):1.
20. Delbecq AL, Van de Ven AH. A Group Process Model for Problem Identification and Program Planning. *The Journal of Applied Behavioural Science.* 1971;7(4): 466-492.

Appendix A

Draft Literature Search Protocol

The following document sets out our current plans regarding searching the literature to identify relevant studies.

After the grant start date (January 2018) we plan to extensively seek advice from our consultation group regarding database inclusion and for help in compiling a comprehensive list of search terms to include term variants, specialist terminology e.g. names of mobile apps or computerised CBT packages, acronyms, and a range of relevant controlled vocabulary headings (e.g. Mesh, Emtree). Search terms will also be identified from the titles, abstracts and indexing headings of already known reviews and reviews retrieved from our preliminary scoping search. Individual search strategies will be developed for each of the databases to be searched.

This draft protocol therefore is very much subject to change up to Mid 2018, at which point we aim to have registered our formal study protocol.

Research question: Cognitive behavioural therapy: An overview of systematic reviews and meta-analyses

Proposed databases to be searched: DARE; Cochrane Library of Systematic Reviews; MEDLINE; EMBASE; PsychInfo; CINAHL.
[We plan to also explore (and then consult our consultation group) whether to additionally search the following databases (depending upon database scope/relevancy, range of journal titles indexed, inclusion of systematic reviews): Allied and Complementary Medicine (AMED); SCOPUS; Applied Social Sciences Index & Abstracts (ASSIA); Child Development & Adolescent Studies; Sociological Abstracts; Web of Science; Educational Resources Information Center (ERIC); Physiotherapy Evidence Database (PEDro); Global Health; LILACS (Latin America); OpenGrey].

Proposed additional sources to search: PROSPERO; ClinicalTrials.gov; the International Clinical Trials Registry Platform (ICTRP).

Proposed non-database search-related identification of papers: hand searching the reference lists of those reviews retrieved by the search; consultation group knowledge; contacting experts.

Search period: 1970 (inception of CBT) to present.

Proposed free-text search terms (variants all listed in full)
The following free-text search terms have been identified thus far, additional terms will be identified from further reading and suggestions from our consultation group.

British spelling variants: **c**ognitive behaviour(al) therapy/therapies; cognitive-behaviour(al) therapy/therapies; cognitive behaviour(al) theory/theories; cognitive behaviour(al) intervention(s); cognitive behaviour(al) training; cognitive behaviour(al) treatment(s); cognitive behaviour(al) psychotherapy/psychotherapies; cognitive-behaviour(al) psychotherapy/psychotherapies; cognitive-behaviour(al) programme(s); cognitive behaviour(al) programme(s); cognitive behavioural analysis system of psychotherapy; cognitive-behaviour(al) group psychotherapy/psychotherapies; cognitive behaviour(al) group psychotherapy/psychotherapies; cognitive-behaviour(al) pretreatment intervention(s); cognitive behaviour(al) pretreatment intervention(s); cognitive-behaviour(al) psychotherapeutic treatment; cognitive behaviour(al) psychotherapeutic treatment; cognitive-behavioural oriented group psychotherapy/psychotherapies; cognitive behavioural oriented group

psychotherapy/psychotherapies; cognitive-behaviour(al) method(s); cognitive behaviour(al) method(s).

American spelling variants: cognitive behavior(al) therapy/therapies; cognitive-behavior(al) therapy/therapies; cognitive behavior(al) theory/theories; cognitive behavior(al) intervention(s); cognitive behavior(al) training; cognitive behavior(al) treatment(s); cognitive-behavior(al) psychotherapy/psychotherapies; cognitive behavior(al) psychotherapy/psychotherapies; cognitive-behavior(al) program(s); cognitive behavior(al) program(s); cognitive behavioral analysis system of psychotherapy; cognitive-behavior(al) group psychotherapy/psychotherapies; cognitive behavior(al) group psychotherapy/psychotherapies; cognitive behavior(al) pretreatment intervention(s); cognitive-behavioral pretreatment intervention(s); cognitive-behavior(al) psychotherapeutic treatment; cognitive behavior(al) psychotherapeutic treatment; cognitive-behavioral oriented group psychotherapy/psychotherapies; cognitive behavioral oriented group psychotherapy/psychotherapies; cognitive-behavioral method(s); cognitive behavior(al) method(s).

General terms: CBT; “Beating the Blues”; Y-CBT; CBASP; CCBT; FearFighter; Sleepio; BtB; "Overcoming Depression"; Cope; "BT Steps"; digital CBT.

Proposed keywords
The following keyword search terms (often author assigned) have been identified thus far, additional terms will be identified from further reading and advice from our consultation group.
Cognitive behavioral therapy; Cognitive behavioural therapy; Cognitive-behavioral therapy; Cognitive-behavioural therapy; CBT; Internet-delivered cognitive-behavioral therapy.

Proposed controlled vocabulary headings
The following controlled vocabulary heading terms have been identified thus far, additional terms will be identified from further reading and on advice from our consultation group.
MeSH headings: Cognitive Therapy; Behavior Therapy; Therapy, Computer-Assisted; Remote Consultation; Internet; Mobile Applications.

Emtree headings: Cognitive Behavioral Therapy; Cognitive Therapy; Cognitive Behavioral Stress Management; Information Technology; Mobile Application; Digital Cognitive Behavioral Therapy; Internet.

Proposed use of validated search filters: we intend to select an appropriate systematic review search filter to include in our search string. A decision will be made at a later stage on the best validated filter to use based on testing of the available filters from the McMaster Hedges Group (https://hiru.mcmaster.ca/hiru/HIRU_Hedges_home.aspx) and the collection compiled by the InterTASC Information Specialists’ Sub-Group (ISSG https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home).

Literature search update search: an update search across all of the databases will be conducted 12 months after the first search date to identify any systematic reviews published in the intervening months.

To note at this stage: in terms of the PRISMA for Protocols checklist items 9 and 10 that are relevant to the literature search, at this stage we cannot fully report these as this is still a draft protocol (written in advance of the grant start date) requiring much additional work from the information specialist (e.g. we have not designed a search strategy yet) and the advice, input and decision making of the consultation group (first meeting to be held on 13 February 2018).