clinician. Standard care, usual care or no treatment conditions were also included in this category.

2. Waiting list (WL)

A commonly used 'treatment as usual' is to randomise participants to active intervention groups and control group, to provide the active intervention to both groups but to delay delivery of the intervention to the control group until after those in the intervention group have completed treatment. As in TAU, patients in the WL condition could receive any appropriate medical care during the course of the study on a naturalistic basis.

3. Attention placebo (AP)

This was defined as a control condition that is regarded as inactive both by researchers and by participants in a trial

4. Psychological placebo (PP)

This was defined as a control condition that is regarded as inactive by researchers in a trial but is regarded as active by the participants

We planned to document additional naturalistic treatment(s) received by participants in both the control and active comparisons for each included study.

5. Pill placebo

Excluded interventions

Psychological therapy models based on social constructionist principles (that focus on the ways in which individuals and groups participate in the construction of their perceived social reality) including couples therapy (Barbato 2006), family therapy (Henken 2007), solution-focused therapy, narrative therapy, personal construct therapy, neuro-linguistic programming and brief problem-solving (Watzlavick 1974) will be excluded. These psychological therapies work with patterns and dynamics of relating within and between family, social and cultural systems in order to create a socially constructed framework of ideas (O'Connell 2007), rather than focusing on one individual's reality.

Guided self-help, in which the practitioner provides brief face-to-face non-therapeutic support to patients who are using a self-help psychological therapy intervention, were excluded, as were bibliotherapy and writing therapies.

Psychological therapies provided wholly by telephone or over the internet were also excluded.

Studies of dual modality treatments, in which patients are randomised to receive a psychological therapy intervention combined with pharmacological treatment, were excluded from the current review.

Component or dismantling studies (in which the effectiveness of individual components of a cognitive-behavioural therapeutic approach are investigated) will not be included unless there were viable comparisons to be made with another psychological therapy approach.

Outcome measures

Primary outcomes

(1) Treatment efficacy: the number of patients who respond to treatment, based on changes on Hamilton Rating Scale for Depression (HAM-D) (Hamilton 1960) or Montgomery-Asberg Depression Rating Scale (MADRS) (Montgomery and Asberg 1979) or Beck Depression Inventory (Beck et al. 1961), or any other validated depression scale at the end of acute phase treatment (8 weeks, range 4-16 weeks). Many studies define response by 50% or greater reduction on the rating scale but some studies define response using Jacobson's Reliable Change Index (Jacobson and Truax 1991). We will focus on 50% or greater reduction in depression severity. If the original authors report several outcomes corresponding with our definition of response, we will give preference to HAM-D. Any version of HAM-D will be accepted.

(2) Treatment acceptability: the number of participants who drop out of treatment for any reasons during the first 8 weeks of treatment (range: 4-16 weeks).

Secondary outcomes

(3) Remission: The number of patients who remit on treatment, based on the endpoint absolute status of the patients, as measured by HAM-D, MADRS, or any other validated depression scale. Examples of definitions of remission include 7 or less on 17-item HAM-D (Furukawa et al. 2007) or 11 or less on MADRS (Bandelow et al. 2006); we will accept the study authors' original definition. If the original authors report several outcomes corresponding with our definition of response, we will give preference to HAM-D.

(4) Severity of depression symptoms, based on a continuous outcome of group mean scores at the end of treatment using HAM-D, MADRS, BDI, or any other validated depression scale.

Search methods for identification of studies

Electronic searches

1. CCDANCTR Specialised Registers
We will search two clinical trials registers created and maintained by the Cochrane Depression, Anxiety and Neurosis Group (CCDAN), the CCDANCTR-Studies Register and the CCDANCTR-References Register. References to trials for inclusion in the Group's registers are collated from routine (weekly) searches of MEDLINE, EMBASE and PsycINFO, quarterly searches of the Cochrane Central Register of Controlled Trials (CENTRAL) and additional ad hoc searches of other databases (PSYNDEX, LILACS, AMED, CINAHL). These searches employ generic terms for depression, anxiety and neuroses; together with sensitive (database specific) RCT filters. Details of the generic search strategies can be found in the 'Specialized Register' section of the Cochrane Depression, Anxiety and Neurosis Group's module text.
References to trials are also sourced from international trials registers via the World Health Organisation's trials portal (http://apps.who.int/trialsearch/); drug companies; the hand-searching of key journals, conference proceedings and other (non-Cochrane) systematic reviews and meta-analyses. Trial databases of the following drug-approving agencies (the Food and Drug Administration in the USA, the Medicines and Healthcare products Regulatory Agency in the UK, the European Medicines Agency in the EU, the Medicines Evaluation Board in the Netherlands, the Medical Products Agency in Sweden, the Pharmaceuticals and Medical Devices Agency (PMDA) in Japan, the Therapeutic Goods Administration (TGA) in Australia) will be hand-searched for published, unpublished and ongoing controlled trials. The National Institute for Clinical Excellence (UK) and the Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Germany) will also be contacted for additional information. No language restriction will be applied.

The CCDANCTR-Studies Register
The CCDANCTR-Studies Register contains over 11,000 trials for the treatment or prevention of depression, anxiety and neurosis. Each trial has been coded using the EU-Psi coding manual (as a guide) and includes information on intervention, condition, comorbidities, age, treatment setting etc.
In order to identify all the studies that compared among the psychotherapies in question, the studies register will be searched using the following search terms:
Condition = (depress* or dysthymi*) and Intervention = (*therap* or training)
In order to identify all the studies that compared the psychotherapies in question against the included antidepressants, the studies register will be searched using the following search terms:
Condition=depress* AND Intervention = (cognitive* or behavi* or *therap* or training or treatment or counsel* or psycho* or humanistic or mindfulness) AND Intervention=(Amitriptylin* or Chlorimipramin* or Clomipramin* or Chlomipramin* or Clorimipramine or or Citalopram or Escitalopram or Fluoxetine or Fluvoxamine or Paroxetine or Sertraline or Desvenlafaxine or Duloxetine or Milnacipran or Venlafaxine or Agomelatine or Bupropion or Mirtazapine or Reboxetine or Trazodone or Vilazdone or tricyclic drugs" or (serotonin or selective) and "reuptake inhibitors"))

The CCDANCTR-References Register
The CCDANCTR-References Register contains bibliographic records of reports of trials coded in the CCDANCTR-Studies Register together with several other uncoded references (total number of records=24,500). This register will be searched using a comprehensive list of terms for 'psychotherapies' as indicated in below. Records already retrieved from the search of the CCDANCTR-Studies Register will be de-duplicated.
1:Title/Abstract=therap* or psychotherap*
2:Keywords=psychotherapy
3:Free-Text=acceptance* or commitment* or "activity scheduling" or adlerian or art or aversion or brief or "client cent*" or cognitive or color or colour or compassion-focused or "compassion* focus*" or compassionate or conjoint or conversion or conversational or couples or dance or dialectic* or diffusion or distraction or eclectic or (emotion and focus*) or emotion-focus* or existential or experiential or exposure or expressive or family or focus-oriented or "focus oriented" or freudian or gestalt or "group" or humanistic or

implosive or insight or integrative or interpersonal or jungian or kleinian or logo or marital or metacognitive or meta-cognitive or milieu or morita or multimodal or multi-modal or music or narrative or nondirective or non-directive or "non directive" or nonspecific or non-specific or "non specific" or "object relations" or "personal construct" or "person cent*" or person-cent* or persuasion or play or ((pleasant or pleasing) and event*) or primal or problem-focused or "problem focused" or problem-solving or "problem solving" or process-experiential or "process experiential" or psychodynamic or "rational emotive" or reality or "reciprocal inhibition" or relationship* or reminiscence or restructuring or rogerian or schema* or self-control* or "self control*" or "short term" or short-term or sex or "social effectiveness" or "social skill*" or socio-environment* or "socio environment*" or "solution focused" or solution-focused or "stress management" or supportive or time-limited or "time limited" or "third wave" or transference or transtheoretical or validation
4:Free-Text=(abreaction or "acting out" or "age regression" or ((assertive* or autogenic or mind or sensitivity) and train*) or autosuggestion or "balint group" or ((behavior* or behaviour*) and (activation or therap* or treatment or contracting or modification)) or biofeedback or catharsis or cognitive or "mind training" or counsel* or "contingency management" or countertransference or "covert sensitization" or "eye movement desensiti*" or "crisis intervention" or "dream analysis" or "emotional freedom" or "free association" or "functional analys*" or griefwork or "guided imagery" or hypno* or imagery or meditation* or "mental healing" or mindfulness* or psychoanaly* or psychodrama or psychoeducat* or "psycho* support*" or psychotherap* or relaxation or "role play*" or "self analysis" or "self esteem" or "sensitivity training" or "support* group*" or therapist or "therapeutic technique*" or "transactional analysis")
5:((1 or 2) and 3) or 4

2. In order to identify all published and unpublished RCTs that compared one antidepressant with another or placebo in the treatment of major depression, all the original databases will be searched using the following phrase: [depress* or dysthymi* or adjustment disorder* or mood disorder* or affective disorder or affective symptoms] and combined with a list of all included antidepressants.

We are aware that there are many trials carried out in China (Chakrabarti et al. 2007). However, for many of these studies only incomplete or conflicting information is available and it has been reported many of them do not use appropriate randomisation procedures (Wu et al. 2006). In an effort to avoid the potential biases that may be introduced by including these trials without further information, we will exclude these studies.

Searching other resources

1. Reference lists
The references of all selected studies will be searched for more published reports and citations of unpublished studies. Relevant review papers will be checked.
2. Personal communication
Subject experts will be contacted to check that all relevant studies, either published or unpublished, have been considered for inclusion.

Data collection

Selection of studies

Two review authors will examine the titles and abstracts of all publications obtained through the search strategy. Full articles of all the studies identified by either of the review authors will then be obtained and inspected by the same two review authors to identify trials meeting the following criteria:
1. Randomised controlled trial;
2. Participants have depression diagnosed by operationalised criteria
3. And compared any of the above-listed psychotherapies or drug therapies

Conflicts of opinion regarding eligibility of a study will be discussed with a third review author, having retrieved the full paper and consulted the authors if necessary, until consensus is reached.

Data extraction and management

Data from each study will be extracted independently by two review authors. Any disagreement will be resolved through discussion and in consultation with the principal investigators. Where necessary, the authors of the studies will be contacted for further information.

Information relating to study population, sample size, interventions, comparators, potential biases in the conduct of the trial, and outcomes will be abstracted from the original reports into specially designed paper forms then entered into a spreadsheet.

Management of time points

It is a problem of systematic reviews that usually trials have different durations. Clinically, the assessment of efficacy after 8 weeks of treatment or after 24 weeks or more may lead to differences in terms of treatment outcome. Clinicians need to know whether (and to what extent) treatments work within a clinically reasonable period of time. Unfortunately, there is no consensus on what the appropriate duration of an acute phase trial is. In the present review, acute treatment will be defined as an 8-week treatment in both the efficacy and acceptability analyses. If 8-week data are not available, we will use data ranging between 4 to 16 weeks and the time point given in the original study as the study endpoint will be given the preference. Longer-term studies will be excluded if they do not provide data for the 4-16 weeks period.

Assessment of risk of bias in included studies

Risk of bias will be assessed for each included study using the Cochrane Collaboration 'risk of bias' tool (Higgins and Green 2011). The following 10 domains will be considered:

1. Sequence generation: was the allocation sequence adequately generated?
2. Allocation concealment: was allocation adequately concealed?
3. Blinding of participants, personnel and outcome assessors for the primary outcome: was knowledge of the allocated treatment adequately prevented during the study?
4. Incomplete outcome data for the primary outcome: were incomplete outcome data adequately addressed?
5. Selective outcome reporting: are reports of the study free of suggestion of selective outcome reporting?

We also assessed

6. Researcher allegiance
7. Therapist allegiance
8. Therapist qualification
9. Treatment fidelity

in the case of psychotherapy interventions, and

10. Sponsorship

in the case of drug trials. We did not classify a study as industry-sponsored when only the medication was provided by the pharmaceutical company.

Other sources of bias included but are not limited to:

- Suboptimal randomization, such as recruiting additional patients to one arm which had a large number of dropouts, or reporting only a subset of the patients randomised without stratification (e.g. only those with major depression are reported when the authors originally randomised all patients who wanted to take the course)
- Stopped early due to some data-dependent process (including a formal-stopping rule)
- Had extreme baseline imbalance
- Differential treatment duration among the arms
- Insufficient delivery of treatment or insensitive scales to measure outcomes, leading to null results

A judgment on the risk of bias will be made for each domain, based on the following three categories:

High risk of bias
Low risk of bias
Unclear risk of bias.

Two independent review authors will assess the risk of bias in selected studies. Any disagreement will be resolved through discussion and in consultation with the principal investigators. Where necessary, the authors of the studies will be contacted for further information.

Statistical analyses

Measures of treatment effect

Considering that depression trials are usually small and that data distribution is difficult to assess for studies with small samples, in this review priority will be given to the use and analysis of dichotomous variables both for efficacy and acceptability.

Dichotomous outcomes: these outcomes will be analysed by calculating a pooled odds ratio (OR) and 95%

confidence intervals for each comparison.
Continuous outcomes: Where different measures are used to assess the same outcome, data will be pooled with standardised mean difference (SMD) and 95% confidence intervals calculated.

Dealing with missing data

Missing dichotomous data will be managed according to the intention to treat (ITT) principle, and it will be assumed that patients who dropped out after randomisation had a negative outcome. When dichotomous efficacy outcomes are not reported but baseline mean and endpoint mean and standard deviation of the depression rating scales are provided, we will calculate the number of responding patients at 8 weeks (range 4 to 16 weeks) employing a validated imputation method (Furukawa et al. 2005). We are aware that other methods to impute response rate are available and have been investigated (Anzures-Cabrera et al. 2011). Even though these imputation methods are valid and may give odds ratios (ORs) with narrower CIs, they only produce logORs and their variances rather than raw data. As we opt for a model based on 2x2 tables using the binomial likelihood, the Fukurawa method will be used in our review.
Missing continuous data will either be analysed on an endpoint basis, including only participants with a final assessment, or analysed using the last observation carried forward to the final assessment (LOCF) if LOCF data were reported by the trial authors. Where possible, exact SDs will be calculated from P values, t-values, confidence intervals or standard errors, when these are reported in articles (Altman and Bland 1996). Where no such information is available, attempts will be made to obtain these data through contacting trial authors. Where the vast majority of actual SDs are available and only a minority of SDs are unavailable or unobtainable, a method used for imputing SDs and calculating percentage responders devised by Furukawa and colleagues (Furukawa et al. 2006) will be used. We will check that the original standard deviations are properly distributed, so that the imputed standard deviation represents the average. Where this method is employed, data will be interpreted with caution, taking account of the degree of heterogeneity observed. A sensitivity analysis will also be undertaken to examine the effect of the decision to use imputed data.
Where additional figures are not available or obtainable, and it is not deemed appropriate to use the Furukawa method described above, the study data will not be included in the comparison of interest.

Data synthesis

We will generate descriptive statistics for trial and study population characteristics across all eligible trials, describing the types of comparisons and some important variables, either clinical or methodological (such as year of publication, age, severity of illness, clinical setting).

Pairwise meta-analysis

For each pair-wise comparison between treatments, the odds ratio will be calculated with a 95% CI. A standard, pairwise meta-analysis will be conducted for each pairwise comparison of treatments using RevMan. We anticipate some clinical heterogeneity between studies and so where there are ≥3 studies we plan to use a random-effects model to incorporate the assumption that the different studies are estimating different, yet related, treatment effects (DerSimonian and Laird 1986). Where there are <3 studies we will combine in a fixed effect analysis (Mantel and Haenszel 1959).
As the correct interpretation of the pooled summary from a random effects analysis is an "average" effect across the studies, it can be difficult to apply to individual study settings (Riley et al. 2011). A prediction interval, which captures the uncertainty in the summary estimate, the estimate of the between study standard deviation (Tau) and the uncertainty in Tau (Higgins et al. 2009), will therefore also be estimated.

MTM

To ensure that the network is connected, a network diagram will be constructed for the primary outcome only. Note that MTM is only possible for a connected set of treatments.
A random-effects MTM, taking into account the correlations induced by multi-arm trials, will be conducted in a Bayesian framework and implemented using WinBUGS 1.4.1 (http://www.mrc- bsu.cam.ac.uk/bugs/). The probability that each treatment is the most effective at improving response will also be calculated. The goodness of fit of the model to the data will be measured by the posterior mean of the residual deviance. This is defined as the difference between the deviance for the fitted model and the deviance for the saturated model, where deviance measures the fit of the model to the data points using the likelihood function. We will examine

leverage plots to help identify any specific data points (trial arms) that were fitting poorly in each model. A leverage plot displays the leverage (a measure of influence equal to the contribution of each trial arm to P_D, the effective number of parameters) versus the signed, square root of the residual deviance (a measure of fit) for each data point. Points with a high leverage are influential, which means that they have a strong influence on the model parameters that generate their fitted values. Convergence will be assessed using two chains and based on the Brooks-Gelman-Rubin diagnostic tool in WinBUGS.

Unit of analysis issues

Studies with more than two intervention arms can pose analytical problems in pair-wise meta-analysis. Where studies have two or more active treatment arms to be compared against treatment as usual, data will be managed in this review as follows:
Continuous data: Means, SDs and number of participants for each active treatment group will be pooled across treatment arms as a function of the number of participants in each arm to be compared against the control group (Higgins 2008; Higgins 2008b; Law 2003).
Dichotomous data: Active treatment groups will be collapsed into a single arm for comparison against the control group, or the control group will be split equally between the treatment groups.

Assessment of heterogeneity

Pairwise meta-analyses

Visual inspection of the forest plots will be used to investigate the possibility of statistical heterogeneity. This will be supplemented using the I-squared statistic. This provides an estimate of the percentage of variability due to heterogeneity rather than a sampling error (Higgins et al. 2003). 95% confidence intervals will be calculated for I-squared, and a P value from a χ^2 test for heterogeneity will be used to assess evidence of its presence. We will also report Tau^2 (the between study variance). We consider a degree of heterogeneity inevitable and therefore only I^2 values $\geq$50% will be explored further using subgroup analyses for the primary outcome only.

MTM

Inconsistency can be considered an additional layer of heterogeneity which can occur in networks of evidence. It can occur when there is a discrepancy between a direct and indirect estimate of treatment effect. As such inconsistency is considered a property of 'closed loops' of evidence. As a first step, we will calculate the difference between indirect and direct estimates in each closed loop formed by the network of trials as a measure of inconsistency and we will subsequently examine whether there are any material discrepancies. We will also use model fit statistics as an informal check of inconsistency. In case of considerable inconsistency we will investigate possible sources of it.
In the network meta-analysis we will assume homogeneous between study variability across studies (Lu and Ades 2004). We will report Tau (the standard deviation of underlying effects across studies) as our estimate of heterogeneity. We will also report the effective number of parameters, pD, which increases with the degree of heterogeneity in the random effect models, and so can also be viewed as a measure of heterogeneity.

Assessment of reporting biases

As far as possible, the impact of reporting biases will be minimised by undertaking comprehensive searches of multiple sources (including trial registries), increasing efforts to identify unpublished material including contacts with the original study authors, and including non-English language publications.
We will also try and identify outcome reporting bias in trials by focusing on our predefined primary and secondary outcomes and using the imputation to increase the number of studies included in the review (Furukawa et al. 2007).

Pairwise meta-analyses

Only if there are greater than 10 studies per pairwise meta-analysis will we consider using funnel plots to assess the impact of reporting bias on the estimates of treatment effect. We will investigate the presence of small study effects for the primary outcomes only. For dichotomous outcomes several tests are available. Note that our choice of test for funnel plot asymmetry will depend on the degree of heterogeneity observed.

MTM

The assessment of reporting biases is a new area in network meta-analysis and no agreed methodology

currently exists. As such, we propose a simple approach to the evaluation of small study effects, and will conduct a sensitivity analysis excluding studies with sample ≤50 participants per arm. We will also use Tanaka's method [ref] and Chaimani's method (Chaimani and Salanti 2012) to assess small study effects in MTM.

Effect modifiers and investigation of heterogeneity

The following sources of possible clinical heterogeneity are listed a priori and will be examined as effect modifiers in the MTM.

1. Year of publication
2. Baseline depression severity: 17-item HAMD scores at baseline will be entered as a covariate.
3. Researcher allegiance in psychotherapy arms and sponsorship in pharmacotherapy arms

Sensitivity analysis

In order to examine if the obtained results are preserved when we limit the included studies to high quality ones only, we will examine the following variables.

1. Allocation concealment: Allocation concealment will be used as a marker of trial quality (Wood et al. 2008). Studies that were rated at high or unclear risk of bias for allocation concealment will be excluded.
2. Blinding: The influence of including open studies and single-blind studies will be examined.
3. Length of treatment: Four weeks may be too short to differentiate psychotherapies and pharmacotherapies. Studies will be limited to those whose outcomes are reported at 6 weeks or later.
4. Psychotherapy fidelity: Studies that were rated at high or unclear risk of bias for psychotherapy fidelity will be excluded.
5. If pharmacotherapy/GP visits allowed as a cointervention in psychotherapy arms
6. Imputation: Trials where missing data have been imputed will be excluded.

Results

References

Altman, D. G. and J. M. Bland (1996). "Detecting skewness from summary information." Bmj **313**(7066): 1200.

American Psychiatric Association (2010). "Practice guideline for the treatment of patients with major depressive disorder (third edition). American Psychiatric Association." Am J Psychiatry **167**(Suppl): 1-152.

Anzures-Cabrera, J., A. Sarpatwari, et al. (2011). "Expressing findings from meta-analyses of continuous outcomes in terms of risks." Statistics in medicine **30**(25): 2967-2985.

Bandelow, B., D. S. Baldwin, et al. (2006). "What is the threshold for symptomatic response and remission for major depressive disorder, panic disorder, social anxiety disorder, and generalized anxiety disorder?" J Clin Psychiatry **67**(9): 1428-1434.

Beck, A. T., A. J. Rush, et al. (1979). Cognitive Therapy of Depression. New York, Guilford Press.

Beck, A. T., C. H. Ward, et al. (1961). "An inventory for measuring depression." Arch Gen Psychiatry **4**: 561-571.

Bromet, E., L. H. Andrade, et al. (2011). "Cross-national epidemiology of DSM-IV major depressive episode." BMC Medicine **9**: 90.

Butler, A. C., J. E. Chapman, et al. (2006). "The empirical status of cognitive-behavioral therapy: a review of meta-analyses." Clin Psychol Rev **26**(1): 17-31.

Chaimani, A. and G. Salanti (2012). "Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions." Research Synthesis Methods.

Chakrabarti, A., C. E. Adams, et al. (2007). "Schizophrenia trials in China: a survey." Acta psychiatrica Scandinavica **116**(1): 6-9.

Churchill, R., M. Khaira, et al. (2000). "Treating depression in general practice: factors affecting patients' treatment preferences." The British journal of general practice : the journal of the Royal College of General Practitioners **50**(460): 905-906.

Cipriani, A., C. Barbui, et al. (2011). "Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis." Lancet.

Cipriani, A., T. A. Furukawa, et al. (2009). "Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis." Lancet **373**: 746-758.

Cooper, P. J. and L. Murray (1998). "Postnatal depression." BMJ **316**(7148): 1884-1886.

Cuijpers, P., A. van Straten, et al. (2008). "Are psychological and pharmacologic interventions equally effective in the treatment of adult depressive disorders? A meta-analysis of comparative studies." J Clin Psychiatry **69**(11): 1675-1685; quiz 1839-1641.

DerSimonian, R. and N. Laird (1986). "Meta-analysis in clinical trials." Control Clin Trials **7**(3): 177-188.

DeRubeis, R. J., L. A. Gelfand, et al. (1999). "Medications versus cognitive behavior therapy for severely depressed outpatients: mega-analysis of four randomized comparisons." The American journal of psychiatry **156**(7): 1007-1013.

Dobscha, S. K., K. Corson, et al. (2007). "Depression treatment preferences of VA primary care patients." Psychosomatics **48**(6): 482-488.

Dobson, K. S. (1989). "A meta-analysis of the efficacy of cognitive therapy for depression." Journal of consulting and clinical psychology **57**(3): 414-419.

Driessen, E., P. Cuijpers, et al. (2010). "Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis." Journal of Consulting and Clinical Psychology **78**(5): 668-680.

Ellis, A. (1979). Theoretical and Empirical Foundations of Rational-Emotive Therapy. Monterey, Brooks/Cole.

Freud, S. (1900). Die Traumdeutung [Interpretation of Dreams]. Leipzig & Wien, Franz Deuticke.

Furukawa, T. A., T. Akechi, et al. (2007). "Evidence-based guidelines for interpretation of the Hamilton Rating Scale for Depression." J Clin Psychopharmacol **27**(5): 531-534.

Furukawa, T. A., C. Barbui, et al. (2006). "Imputing missing standard deviations in meta-analyses can provide accurate results." J Clin Epidemiol **59**(1): 7-10.

Furukawa, T. A., A. Cipriani, et al. (2005). "Imputing response rates from means and standard deviations in meta-analyses." Int Clin Psychopharmacol **20**(1): 49-52.

Furukawa, T. A., N. Watanabe, et al. (2007). "Association between unreported outcomes and effect size

estimates in Cochrane meta-analyses." JAMA **297**(5): 468-470.
Gartlehner, G., R. A. Hansen, et al. (2011). "Comparative benfits and harms of second-generation antidepressants for treating major depressive disorder: an updated meta-analysis." Annals of Internal Medicine **155**: 772-785.
Gibbons, R. D., K. Hur, et al. (2012). "Benefits from antidepressants: Synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine." Archives of General Psychiatry.
Gloaguen, V., J. Cottraux, et al. (1998). "A meta-analysis of the effects of cognitive therapy in depressed patients." J Affect Disord **49**(1): 59-72.
Hamilton, M. (1960). "A rating scale for depression." J Neurol Neurosurg Psychiatry **23**: 56-62.
Higgins, J. P. and S. Green, Eds. (2011). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated September 2011] Available from www.cochrane-handbook.org.
Higgins, J. P., S. G. Thompson, et al. (2003). "Measuring inconsistency in meta-analyses." Bmj **327**(7414): 557-560.
Higgins, J. P., S. G. Thompson, et al. (2009). "A re-evaluation of random-effects meta-analysis." Journal of the Royal Statistical Society. Series A **172**(1): 137-159.
Higgins, J. P. and A. Whitehead (1996). "Borrowing strength from external trials in a meta-analysis." Stat Med **15**(24): 2733-2749.
Hollon, S. D. and K. Ponniah (2010). "A review of empirically supported psychological therapies for mood disorders in adults." Depression and anxiety **27**(10): 891-932.
Imel, Z. E., M. B. Malterer, et al. (2008). "A meta-analysis of psychotherapy and medication in unipolar depression and dysthymia." Journal of affective disorders **110**(3): 197-206.
Jacobson, N. S. and P. Truax (1991). "Clinical significance: a statistical approach to defining meaningful change in psychotherapy research." J Consult Clin Psychol **59**(1): 12-19.
Jung, C. G. (1921). Psychologische Typen [Psychological Types]. Zuerich, Rascher Verlog.
Klein, M. (1932). Die Psychoanalyse des Kindes [Psychoanalysis of Children]. Wien, Internationaler Psychoanalytischer Verlag.
Klerman, G. L., M. M. Weissman, et al. (1984). Interpersonal Psychotherapy for Depression. New York, Basic Books.
Lu, G. and A. E. Ades (2004). "Combination of direct and indirect evidence in mixed treatment comparisons." Stat Med **23**(20): 3105-3124.
Lumley, T. (2002). "Network meta-analysis for indirect treatment comparisons." Stat Med **21**(16): 2313-2324.
Luppa, M., S. Heinrich, et al. (2007). "Cost-of-illness studies of depression: a systematic review." Journal of affective disorders **98**(1-2): 29-43.
Lynch, D., K. R. Laws, et al. (2010). "Cognitive behavioural therapy for major psychiatric disorder: does it really work? A meta-analytical review of well-controlled trials." Psychol Med **40**(1): 9-24.
Mantel, N. and W. Haenszel (1959). "Statistical aspects of the analysis of data from retrospective studies of disease." Journal of the National Cancer Institute **22**: 719-748.
Marks, I. M. (1981). Cure and Care of Neuroses: Theory and Practice of Behavioural Psychotherapy. New York, Wiley.
Maslow, A. H. (1943). "A theory of human motivation." Psychological Review **50**: 270-296.
Montgomery, S. A. and M. Asberg (1979). "A new depression scale designed to be sensitive to change." Br J Psychiatry **134**: 382-389.
NICE (2009). Depression: the treatment and management of depression in adults (partial update of NICE clinical guideline 23). London, National Institute for Clinical Excellence.
Oei, T. P. and T. Kazmierczak (1997). "Factors associated with dropout in a group cognitive behaviour therapy for mood disorders." Behaviour research and therapy **35**(11): 1025-1030.
Pinquart, M., P. R. Duberstein, et al. (2006). "Treatments for later-life depressive conditions: a meta-analytic comparison of pharmacotherapy and psychotherapy." The American journal of psychiatry **163**(9): 1493-1501.
Psaty, B. M., T. Lumley, et al. (2003). "Health outcomes associated with various antihypertensive therapies used as first-line agents: a network meta-analysis." JAMA : the journal of the American Medical Association **289**(19): 2534-2544.
Riedel-Heller, S. G., H. Matschinger, et al. (2005). "Mental disorders--who and what might help?

Help-seeking and treatment preferences of the lay public." Social psychiatry and psychiatric epidemiology **40**(2): 167-174.
Riley, R. D., J. P. Higgins, et al. (2011). "Interpretation of random effects meta-analyses." BMJ **342**: d549.
Rogers, C. (1951). Client-Centered Therapy: Its Current Practice, Implications and Theory. London, Constable.
Roth, A. D. and S. Pilling (2008). "Using an evidence-based methodology to identify the competences required to deliver effective congitive and behavioural therapy for depression and anxiety disorders." Behavioural and Cognitive Psychotherapy **36**(2): 129-147.
Skinner, B. F. (1953). Science and Human Behavior. New York, Free Press.
Stettler, C., S. Wandel, et al. (2007). "Outcomes associated with drug-eluting and bare-metal stents: a collaborative network meta-analysis." Lancet **370**(9591): 937-948.
Turner, E. H., A. M. Matthews, et al. (2008). "Selective publication of antidepressant trials and its influence on apparent efficacy." New England Journal of Medicine **358**: 252-260.
van Geffen, E. C., H. Gardarsdottir, et al. (2009). "Initiation of antidepressant therapy: do patients follow the GP's prescription?" The British journal of general practice : the journal of the Royal College of General Practitioners **59**(559): 81-87.
van Schaik, D. J., A. F. Klijn, et al. (2004). "Patients' preferences in the treatment of depressive disorder in primary care." General hospital psychiatry **26**(3): 184-189.
Vergouwen, A. C., A. Bakker, et al. (2003). "Improving adherence to antidepressants: a systematic review of interventions." The Journal of clinical psychiatry **64**(12): 1415-1420.
Watson, J. B. (1924). Behaviorism. New York, Norton.
WHO (2008). The global burden of disease: 2004 update. http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/index.html.
Wood, L., M. Egger, et al. (2008). "Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study." BMJ **336**(7644): 601-605.
Wu, T. X., Y. P. Li, et al. (2006). Investigation of authenticity of 'claimed' randomized controlled trials (RCTs) and quality assessment of RCT reports published in China. XIV Cochrane Colloquium. Dublin, Irreland.

MAP-P: Meta-Analyses of Psychotherapies for Panic disorder

[PROTOCOL]

Pompoli A, Furukawa TA, Imai H, Tajika A, Salanti G, Efthimiou O

Background

Description of the condition

A *panic attack* is a discrete period of fear or anxiety that has a rapid onset, reaches a peak within 10 minutes and in which at least four of thirteen characteristic symptoms are experienced. Many of these symptoms involve bodily systems, such as racing heart, chest pain, sweating, shaking, dizziness, flushing, stomach churning, faintness and breathlessness. Further recognised panic attack symptoms involve fearful cognitions, such as the fear of collapse, going mad or dying, and derealization (APA 2000).

According to the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR), *panic disorder* is characterized by the presence of recurrent unexpected panic attacks, of which at least one has been followed by one month (or more) of persistent concern about having additional attacks, worry about the implications of the attack (or its consequences) or a significant change in behavior related to the attacks.

Panic disorder is common in the general population with a lifetime prevalence of 1 to 4% (Eaton 1994; Bijl 1998). In primary care settings panic syndromes have been reported to have a prevalence of around 10% (King 2008). Its aetiology is not fully understood and is probably heterogeneous. Biological theories incorporate the faulty triggering of an inbuilt anxiety response, possibly a suffocation alarm. Evidence for this comes from biological challenge tests (lactate and carbon dioxide trigger panic in those with the disorder) and from animal experiments and neuroimaging studies in humans that show activation of fear circuits, such as that involving the periaqueductal grey matter (Gorman 2000).

Agoraphobia is anxiety about being in places or situations from which escape might be difficult or embarrassing or in which help may not be available in the event of having a panic attack (APA 2000). Agoraphobia can occur with panic disorder: in the general population, about one-fourth of people suffering from panic disorder also have agoraphobia but this proportion is much higher in the clinical samples (Kessler 2006). The presence of agoraphobia is associated with increased severity and worse outcome (Kessler 2006). There are several risk factors that predict the development of agoraphobia in people suffering from panic disorder: female gender, more severe dizziness during panic attacks, cognitive factors, dependent personality traits and social anxiety disorder (Starcevic 2009).

Panic disorder, with or without agoraphobia, is highly comorbid with other psychiatric disorders, such as drug dependence, major depression, bipolar I disorder, social phobia, specific phobia, generalized anxiety disorder (Grant 2006). It is estimated that generalized anxiety disorder co-occurs in 68% of people with panic disorder, whilst major depression has a prevalence of 24 to 88% among people with panic disorder (Starcevic 2009).

Description of the intervention

Recent guidelines (NICE 2011) recommend three types of intervention in the care of individuals with panic disorder, any of which should be offered promptly and taking into account the preference of the patient. According to the NICE guidelines, the interventions that have evidence for the longest duration of effect, in descending order, are psychological therapy, pharmacological therapy (antidepressant medication) and self-help. Among psychological interventions, NICE guidelines further suggest that cognitive behavioural therapy (CBT) should be used, optimally in the form of 1-2 hours weekly sessions, for a total of 7-14 hours within a maximum of 4 months since commencement. A further recommendation is that CBT should be delivered only by suitably trained and supervised people who can demonstrate that they adhere closely to empirically grounded treatment protocols. A Cochrane meta-analysis comparing combined psychotherapy plus antidepressants vs psychotherapy alone or pharmacotherapy alone (Furukawa 2007) showed the superiority of combined therapy over either monotherapies in the short term, and that of combined therapy and psychotherapy alone over pharmacotherapy alone in the long term, thus suggesting that either combined therapy or psychotherapy alone can be chosen as first line treatment for panic disorder with or without agoraphobia. In particular, behavioural and cognitive-behavioural psychotherapy showed the strongest evidence. Another meta-analysis, aimed at analysing the efficacy of psychological interventions vs control conditions in the treatment of panic disorder with or without agoraphobia (Sánchez-Meca 2010), showed a general efficacy of psychological treatments over different clusters of symptoms, with the most consistent results in favour of the combination of exposure strategies with relaxation training and/or breathing retraining techniques.

Why it is important to do this review

The efficacy of psychological therapies in the treatment of panic disorder with or without agoraphobia has been repeatedly demonstrated in a growing number of trials and further confirmed in a recent meta-analysis (Sánchez-Meca 2010), with particular consistency in favour of behavioural and cognitive-behavioural strategies. The study conducted by Sánchez-Meca et al. revealed the presence of substantial heterogeneity among included studies ($I^2 = 70,4\%$; $p < .01$). The authors explored the observed heterogeneity with exploratory (i.e. uncorrected for multiple testing) secondary analyses, suggesting the role of different factors in accounting for the observed variance. Among these, the following seem to be particularly noteworthy: type of treatment, type of control group and percentage of patients with agoraphobia. The observed degree of heterogeneity due to differences in the psychological treatment suggests that some psychotherapies may be more effective than others in the treatment of the disorder. However, both the existence and the eventual magnitude of such differences remain unclear. This is partly due to the presence of methodological diversity among

available studies: as suggested by Sánchez-Meca et al., the type of control group may significantly influence the measured effect size, limiting the possibility of drawing conclusions. A further consideration is that only a few trials compared different psychological approaches with each other and, more generally, psychological therapies have not been all equally investigated.
In the attempt to overcome these issues, in this review we will perform a network meta-analysis (NMA), also known as multiple treatment meta-analysis. As a first step, we will conduct a systematic search of all relevant papers. Then, a treatment network will be constructed in order to clarify the extent to which each treatment and each comparison have been investigated so far in randomized controlled trials. In contrast to previous reviews, control conditions will be classified in four possible categories, each constituting an independent node of the network, in order to reduce the amount of heterogeneity among network comparisons and allow formal comparisons of control conditions among themselves. As a third step, for each available comparison, an independent meta-analysis will be performed in order to obtain direct estimates of the relative effect-sizes, and thus synthesize the results of the included studies. Subsequently, mixed estimates of all comparisons will be calculated synthesizing the available direct and indirect evidence via NMA, in order to obtain an overall effect-size estimate for each possible pair of treatments in the network. This will allow to disclose and assess differences in effects not only among experimental interventions, but also among control conditions. Finally, a probabilistic ranking will be calculated in order to help the identification of those interventions which are more likely to be more effective than others in the treatment of panic disorder.

Objectives

To assess the comparative efficacy and acceptability of different psychological therapies and different control conditions in the treatment of panic disorder with or without agoraphobia in adults.

Methods

Criteria for considering studies for this review

Types of studies
All randomized controlled trials (RCTs), comparing one type of psychotherapy against another or against a non-pharmacological control condition in the short and long term treatment of panic disorder. Randomized controlled trials with more than two arms will be included. Cluster-randomized trials will be included when effects of clustering are taken account of. For trials which had a crossover design, only results from the first randomization period will be considered.
Quasi-randomized controlled trials (in which treatment assignment was decided trough methods such as alternate days of the week) will be excluded.

Types of participants
Inclusion criteria
Patients, aged 18 or older, of both sexes, with a primary diagnosis of panic disorder with or without agoraphobia diagnosed according to any of the following criteria: Feighner criteria, Research Diagnostic Criteria, DSM-III, DSM-III-R, DSMIV or ICD-10. There is evidence that over 95% of patients with agoraphobia seen clinically suffer from panic disorder as well (Goisman 1995). According to this finding, studies focusing on agoraphobia, rather than panic disorder, will be included if operationally diagnosed according to the above-named criteria and when it can be safely assumed that at least 30% of the participants are suffering from panic disorder. The effect of the inclusion of trials with different percentages of patients suffering from agoraphobia will be explored in a meta-regression analysis.
Studies focusing on adolescents will not be included. However, studies including in the sample patients younger that 18 will be included as long as it can be safely assumed that at least 80% of patients are aged 18 or older.
Subjects must be outpatients at the time of enrolment. Both treatment-naive patients and patients who have already undergone some previous treatment (either psychological or pharmacological) will be included, as long as they satisfy the above mentioned inclusion criteria.
Studies including in the sample patients with other anxiety disorders (e.g. GAD, specific phobias) or with subthreshold panic disorder will be included if: 1) separate results for patients with PD are available and 2) randomization is stratified by specific diagnoses. Stratification by diagnosis will not be required if the total sample includes at least 40 patients with PD.
Exclusion criteria
Trials in which all participants had a concurrent primary diagnosis of Axis I or II disorders other than panic disorder or agoraphobia will be excluded.
Physical comorbidities in themselves do not constitute an exclusion criterion. However, trials which focus on panic disorder or agoraphobia among patients with a certain physical comorbidity will be excluded.

Types of interventions
Experimental interventions

The following psychological treatments will be included:

1. PE - Psychoeducation, intended as sessions in which patients are provided informations about their disease.
2. PS - Psychological support, with or without a psychoeducational component, intended as sessions in which no specific psychotherapy is employed.
3. PT - Physiological therapies that use some kind of physical training (e.g. breathing retraining, applied relaxation and other relaxation techniques).
4. BT - Behaviour therapy that uses some kind of exposure with or without physiological therapy elements such as the above mentioned.
5. CT - Cognitive therapy that uses some kind of cognitive restructuring without neither exposure nor physiological therapy elements.
6. CBT - Cognitive-behaviour therapy containing both cognitive and behavioural or physiological therapy elements, administered face-to-face.
7. SH CBT - Cognitive-behaviour therapy, described as above, in its book/computer/internet version, administered with or without minimal therapist contact
8. 3rd Wave - Acceptance and Commitment Therapy, mindfulness-based therapy and other so-called "3rd wave" therapies.
9. PD - Psychodynamic therapy focused on revealing and resolving intrapsychic or unconscious conflicts.

Therapies can be of any length so that those given in a single session will be accepted.
Both individual and group therapies will be included.
So-called component studies (e.g. dismantling studies) will be included as long as each arm can be regarded as any of the above-defined experimental interventions compared against another experimental or comparator treatment. Eventually, study arms may be regarded as giving informations about the same experimental intervention and thus be combined.
Interpersonal psychotherapy, emotion-focused therapy and any other approach not specifically designed for the treatment of panic disorder, such as Morita therapy, Eye Movement and Desensitization Reprocessing (EMDR), Music therapy and physical exercise will also be excluded.
With the only exception of Self Help CBT, therapies must be administered face-to-face. Therapies remotely administered (e.g. through phone calls or video calls) will therefore be excluded. In the case of Psychoeducation, the simple provision of informational material without any face-to-face session will not be considered an active intervention.
When psychoeducation and/or psychological support are accompanied by any other psychological intervention, study arms will be classified accordingly to the latter and psychoeducation and psychological support will be regarded as components of that intervention.
Family therapy, couple therapy and other psychosocial interventions whose intervention focus is not the individual but rather the family system or couple as a whole will also be excluded.
Study-arms in which a pharmacological co-administration is allowed will be included as long as there are no systematic differences in drug administration between them. The percentage of studies in which a drug co-administration is allowed, the percentage of studies that require a stabilization of therapy and, in this latter case, the time of required stabilization, will be reported.

Comparator interventions

1. NT - No psychological treatment (participants receive assessment only, with or without simple provision of informational material and/or minimal therapist contact, and they know that they will not receive the active treatment in question after the trial).
2. WL - Waiting List (participants receive assessment, with or without minimal therapist contact during the waiting phase, and they know that they will receive the active treatment in question after the trial).
3. APP - Attention or Psychological Placebo (participants receive a face-to-face inactive* intervention that can be perceived both as ineffective or, respectively, effective).

Given the general inconsistency of comparator interventions' definitions among different studies, the attribution of a control group to one of these prespecified categories will rely on its detailed description rather than on the name given by the authors. However, when enough information is not available neither from the paper nor by contacting the original authors, the attribution will solely rely on the given definition. Particular inconsistency exists in the definition of what is intended for treatment as usual (TAU): when TAU is intended as no treatment, waiting list or psychological support, groups will be classified accordingly.
*Attention placebo is defined as any form of inactive intervention designed by the original authors to be perceived as ineffective by patients; Psychological placebo is defined as any form of inactive intervention designed by the original authors to be perceived as effective by patients. The inclusion of an intervention among attention or psychological placebo groups requires that intervention to be inactive. Any form of active intervention will therefore be included among experimental interventions even if defined as a control condition by the original authors.
Study arms in which a Pharmacological placebo is either co-administered or used as control condition will be excluded.
In total we expect the network to have 12 nodes each one representing an intervention or control.

Types of outcome measures

Primary outcomes

- Short term* remission** of panic disorder with or without agoraphobia.
- Short term* response** of panic disorder with or without agoraphobia.
- Dropouts for any reason in the short term

(*) Short term, i.e. within 6 months from treatment commencement. When multiple time-point measures are available, preference will be given to measures at approximately 3 months after treatment commencement.
(**) "Remission" is intended as a dichotomous outcome expressing the number of patients who reached a satisfactory end-state as defined by global judgment by the original investigators. Examples would be "panic free" and "no or minimal symptom" according to the Clinical Global Impression Severity Scale.
"Response" is intended as a dichotomous outcome expressing the number of patients who had a substantial improvement from baseline as defined by the original investigators. Examples would be "very much or much improved" according to the Clinical Global Impression Change Scale, more than 40% reduction in the Panic Disorder Severity Scale score, and more than 50% reduction in the Fear Questionnaire Agoraphobia Subscale.

Secondary outcomes

- Short term improvement of panic disorder with or without agoraphobia as measured on a continuous scale*.
- Long term** remission/response*** of panic disorder with or without agoraphobia after treatment discontinuation.

(*) Examples would be Panic Disorder Severity Scale (total score 0 to 28), Panic and Agoraphobia Scale (total score 0 to 45), Clinical Global Impression Severity Scale (1 to 7), Clinical Global Impression Change Scale (1 to 7), etc. When more than one scale is available in the paper, preference will be given in the following order:

1. PDSS > PAS > ASI-R > ASI> ACQ > BSQ > other scales specific for panic disorder
2. CGI-S > CGI-I > GAS > GAF > other global scales
3. FQ-ag > FQ-global > MI-AAL > MI-AAC > other scales specific for agoraphobia only
4. Panic frequency > panic severity > other scales specific for panic attacks only
5. BAI > HAM-A > STAI > other general anxiety scales

Once the scale has been chosen, if both self and observer-rated assessments are available, preference will be given to the latter. The actual measure entered into meta-analysis is indicated at the top of the listings in the Table of Included Studies.
(**) Long term, i.e. 6 months or longer after treatment commencement, either on treatment discontinuation or on continued treatment (in the case of long-term therapies). When multiple time-point measures are available, preference will be given to measures at approximately 12-15 months after treatment commencement.
(***) "Response" and "Response" are intended as above. It is likely that not all studies will report both response and remission rates in the long term. When both remission and response rates are reported, we will consider the former. However, if remission rates are not reported but response rates are available, these will be used for the analyses.

Search methods for identification of studies

Electronic searches

(1) Cochrane Depression, Anxiety and Neurosis Group's Specialized Register (CCDANCTR)

We will search the Cochrane Collaboration Depression, Anxiety and Neurosis Controlled Trials Registers (CCDANCTR-Studies and CCDANCTR-References), study-based and reference-level registers, respectively, of randomized trials incorporating results of group searches of MEDLINE (1966 onwards), EMBASE (1980 onwards), CINAHL (1982 onwards), PsycINFO (1974 onwards), PSYNDEX (1977 onwards) and LILACS (1982 to 1999), and hand searches of major psychiatric and medical journals as well as trial registries and pharmaceutical company websites (see Collaborative Review Group Search Strategy).The following simple search terms will be used: "panic" AND "psychotherapy". No language restriction will be applied

(2) Supplementary searches

Complementary searches will be conducted on PubMed as well as on trials registers such as the WHO International Clinical Trials Registry Platform (http://apps.who.int/trialsearch/) and Clinicaltrials (http://clinicaltrials.gov/).

Searching other resources

(3) Citation indexes

All the selected studies will be sought as a citation in the Web of Science database in order to identify more studies.

(4) Gray literature

In order to detect references not formally published in books or journals, a search will be run in the OpenSIGLE database (http://www.opengrey.eu/).

(5) Reference checking

The references of all selected studies will be inspected for more published reports and citations of unpublished research. In addition, relevant review papers, both systematic and traditional ones, will be checked.

(6) Personal communication

Experts in the field will be contacted to identify any outstanding study.

Data collection and analysis

Selection of studies
At least two out of three review authors (AP, AT, HI) will examine titles and abstracts of references identified by the electronic search strategies described above to check whether the study is likely to be relevant. Each potentially relevant study located in the search will then be obtained as a full article and independently assessed for inclusion by the same two review authors and, in the case of discordance, resolution will be sought by discussion. When disagreement cannot be solved by discussion, arbitration will be provided by a fourth author (TAF). Agreement between review authors in the study selection will be reported. The discordance in the selection of studies will be evaluated quantifying both the percentage of agreement and Cohen's Kappa (k) (Cohen 1960). Where it is not possible to evaluate the study because of missing information, the study will be classified as "Study awaiting assessment" until further information can be obtained. The reasons for the exclusions of trials will be reported in the "Characteristics of excluded studies" table.

Data extraction and management
At least two out of three review authors (AP, AT, HI) will use a structured, pilot-tested, Excel data collection form to independently extract the data from included studies. Extracted data will concern: study design, administered interventions (format and timing of psychotherapy and control condition, therapist training, intervention components), participants' characteristics (diagnostic criteria, percentage of agoraphobic patients), outcomes, risk of bias and publication. Again, any disagreement will be resolved either by discussion or by consultation of a fourth member of the review team (TAF). If necessary, authors of studies will be contacted to obtain further clarification. Agreement between data extractors will be reported.

Assessment of risk of bias in included studies
At least two out of three review authors (AP, AT, HI) will independently assess the risk of bias in included studies using the tool described in the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2011). The following domains will be assessed:
1. Random sequence generation and allocation concealment (Selection bias)
2. Therapist and researcher allegiance, treatment fidelity (Performance bias)
3. Blinding of outcome assessor (Detection bias)
4. Incomplete outcome data reporting (Attrition bias)
5. Selective outcome reporting (Reporting bias)

The risk of bias, in each domain and overall, are assessed and categorized into:
- Low risk of bias, plausible bias unlikely to seriously alter the results
- High risk of bias, plausible bias that seriously weakens confidence in the results
- Unclear risk of bias, plausible bias that raises some doubt about the results

Where inadequate details of randomization and other characteristics of trials are provided the risk of bias will be classified as unclear, unless further information can be obtained by contacting the authors. If the assessors disagree, the final rating will be made by discussion or with the involvement of another member of the review group (TAF) if necessary. Agreement between the two independent raters will be reported.
Risk of performance bias will not be evaluated (and assumed to be constantly at high risk) as in this kind of studies blinding of participants and therapist is not feasible.
The risk of detection bias will be evaluated for the first of the primary outcomes only. Studies will be classified as having a low risk of detection bias when the identification of a patient as a "remitter" requires at least one observer rating and the observer is blind to the treatment allocation.
Risk of attrition bias will be calculated separately for short-term remission (the first of the primary outcomes) and long-term remission/response. A study will be classified as having a low risk of attrition bias according to the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2011).
Whenever possible, study protocols will be retrieved in order to assess the risk of reporting bias. A study will be considered to be at low risk of reporting bias when the study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way.

Measures of treatment effect
Binary or dichotomous data
As measure of treatment effect for binary outcomes we will use the risk ratio (OR) and its 95% confidence interval (CI).
Continuous data
It is likely that different studies have used varied panic rating scales; therefore we will use standardized mean difference (SMD). If all included studies have used the same instrument, we will use mean differences (MD).
Endpoint versus change data

We prefer to use scale endpoint data, which typically cannot have negative values and are easier to interpret from a clinical point of view. However, if endpoint data are unavailable, we will use the change data. Although statistically imprecise, this approach seems less likely to affect the final results than just excluding change data from the analyses, that may result in an increased risk of selective outcome reporting.

Unit of analysis issues

Cluster-randomised trials

In cluster randomized trials, groups of individuals rather than individuals are randomized to different interventions (Higgins 2011). Cluster-randomized trials will be included when effects of clustering are taken account of.

Cross-over trials

Crossover trials are trials where all participants receive both the control and intervention treatment but in a different order. The major problem is a carry-over effect from the first phase to the second phase of the study, especially if the condition of interest is unstable (Elbourne 2002). As this is the case with panic disorder, randomized crossover studies will be eligible but only data up to the point of first cross over will be used.

Studies with multiple treatment groups

Where a study involves more than two treatment arms, especially in the case of dismantling studies, arms will be combined as long as they can be regarded as subtypes of the same psychotherapy under review. When arms cannot be regarded as if in each of them a different subtype of the same intervention is administered, we will compare each arm with the common comparator separately. If such a situation occurs, the common comparator arm will be subdivided for pairwise meta-analyses. For example, the sample size and the number of responders of that arm will be halved for dichotomous outcomes; for continuous outcomes, the mean and SD will remain the same but the number of patients included will be halved. In the NMA, multivariate meta-analysis methods will be used to synthesise the results (Higgins 2011, chapter 16.6.3).

Dealing with missing data

We will try to contact the authors for all relevant missing data.

Dichotomous outcomes

The proportion of remission and response will be calculated using an intention-to-treat analysis (ITT) following the principle "once randomised always analysed". To this end, all randomized patients for which outcome data are not available will be assumed to be non-responders (Spinelli et al, to appear in clinicaltrials). The same principle will be applied to short and long term outcomes. When dichotomous outcomes are not reported but the means and standard deviations on a panic disorder scale are reported, we will calculate the number of responding or remitted participants according to a validated imputation method (Furukawa 2005).

Continuous outcomes

The ITT approach is not feasible when studies only perform LOCF or endpoint analyses. Therefore, for continuous data, an "available cases analysis" will be performed in which outcomes will be analysed on an endpoint basis, including patients with either a final assessment or a last observation carried forward to the final assessment.

Missing statistics

When only P or standard error (SE) values are reported, we will calculate standard deviations (SDs) (Altman 1996). In the absence of supplementary data after requests to the authors, the SDs will be calculated according to a validated imputation method (Furukawa 2006).

Data synthesis

Pairwise meta-analyses

For each available comparison explored by at least two trials, we will perform a pair-wise meta-analysis in order to provide overall estimates of treatment effect. Since we expect some clinical heterogeneity between studies, we plan to use a random-effects model to incorporate the assumption that the different studies are estimating different, yet related, treatment effects (Higgins 2011). An "average" treatment effect across the studies will therefore be calculated for each available comparison. For dichotomous outcomes, the average risk ratio will be calculated with a 95% CI; for continuous outcomes the average SMD (or the MD if all trials use the same scale) will be calculated with a 95% CI.

For each direct comparison, X^2 and I^2 statistics will be calculated in order to detect the presence of heterogeneity and, respectively, assess its degree. I^2 provides an estimate of the percentage of variability in effect estimates that is due to heterogeneity rather than chance alone (Higgins 2003). We will also report Tau^2, the between study variance in random-effects meta-analysis.

Finally, visual inspection of the forest plots will also be used in order to investigate the presence and nature of statistical heterogeneity. We will investigate the presence of small study effects for the primary outcomes only: along with visual inspection of the plots, we will formally examine whether the association between estimated intervention effects and the study size is greater than might be expected to occur by chance.

Network meta-analysis

An indirect comparison allows to estimate the effect of treatment B relative to treatment A via a common comparator C, by statistically combining the summary effects from "A vs C" and "B vs C" studies (Glenny 2005, Caldwell 2005). A NMA combines direct and indirect evidence across a network of studies to make inferences regarding the relative effectiveness of multiple interventions.
A NMA is only possible for a connected set of treatments. A network diagram will be constructed for our primary outcomes in order to evaluate the extent to which treatments are connected. A visual inspection of network geometry and asymmetry will be conducted in order to investigate, respectively, the overall pattern of comparisons and the extent to which specific treatments or specific comparisons are represented more heavily than others in the network (Salanti 2008).
A random-effects NMA, taking into account the correlations induced by multi-arm trials, will be conducted (Lu & Ades 2004, Higgins Salanti Dias). For each comparison, an average effect estimate along with its 95% credible interval (CrI) will be reported.
Besides yielding relative treatment effects for each comparison, a NMA also allows to estimate the relative ranking of treatments. To rank the treatments according to each outcome accounting for the uncertainty in the treatment effects, we will use the surface under the cumulative ranking curve "SUCRA" (Salanti 2011). The absolute ranks of the treatments per outcome will be presented using"Rankograms" that visually show the distribution of ranking probabilities (Salanti 2011). NMA models typically employ a single heterogeneity parameter. We will report it and we will judge its magnitude against the distribution of values typically found in Cochrane reviews as presented in (Turner 2012).
An assumption underlying NMA is that effect modifiers are similarly distributed across comparisons in the network. That means that an effect modifier should be similar in AB and BC trials in order to obtain a valid AC estimate. Equivalent formulations of the transitivity assumption are presented in Salanti 2012. In order to verify this assumption, for each comparison we will compile a table of important trial and patient characteristics and visually inspect the similarity of factors we consider likely to modify treatment effect. We will also assess the inclusion/exclusion criteria of every trial in the network, to ensure that patients, trial protocols, etc are similar in those aspects which might modify the treatment effect.
Lack of transitivity can be manifested in the data as disagreement between direct and indirect evidence (Caldwell 2005, Lu & Ades 2004, Lumley 2002). This can be evaluated statistically by contrasting the direct and the indirect estimates and calculate a test within each closed loop (Bucher 1997, Salanti 2009). We will report the percentage of inconsistent loops in the network and will examine further the data of loops that appear particularly inconsistent. As this approach does not provide an omnibus test and is associated with multiple testing we will also employ other approaches to infer about the statistical inconsistency. First, we will compare the goodness of fit between models that assume consistency and models that do not (pairwise meta-analyses sharing the same heterogeneity parameter). Subsequently we will perform a design-by-treatment interaction test (Higgins 2012). In case that a small amount of inconsistency is found, we will incorporate this in the estimation by fitting inconsistency models (Higgins 2012, Lu & Ades 2004).

Subgroup analysis and investigation of heterogeneity
Subgroup and meta-regression analyses are often exploratory in nature and should be interpreted cautiously. Firstly, because these analyses often involve multiple analyses, they may yield false positive results; secondly, these analyses lack power and are more likely to result in false negative results. Keeping in mind the above reservations, we would perform meta-regression analyses to investigate, for the first of the primary outcome only (Short term rate of remission of panic disorder with or without agoraphobia), the following candidate explanatory variables:

- Time
- Mean number of treatment sessions: less than 4 sessions, from 4 to 12 sessions, more than 12 sessions
- Therapist training: therapist with/without formally recognized specific training in the type of psychotherapy administered
- Percentage of patients with agoraphobia: measured as a continuous variable
- Percentage of patients with depression: measured as a continuous variable
- Percentage of patients on drug treatment: measured as a continuous variable

Sensitivity analysis
The process of undertaking a systematic review and meta-analysis involves a sequence of decisions, some of which are somewhat arbitrary or unclear (Higgins 2011). A sensitivity analysis is a repeat of the primary analysis, substituting alternative decisions or range of values for decisions that were arbitrary or unclear. We plan to perform the following sensitivity analyses for the first of the primary outcome only (Short term rate of remission of panic disorder with or without agoraphobia):

- Restrict the inclusion in the analyses only to studies considered to be at low risk of selection and detection bias (i.e. adequate allocation sequence generation, adequate allocation concealment, blinding of assessors).
- Exclude from the analyses group therapy trials.
- Exclude from the analyses trials in which a concomitant pharmacotherapy is allowed.
- Exclude from the analyses trials in which drug therapy is not stabilized*.
- For pair-wise meta-analyses, use a fixed-effect model instead of a random-effects model

- For NMA, use a fixed-effect model or a "relaxed" random-effects model instead of a classic random-effects model

(*) Drug therapy will be considered stabilized when: 1) drug administration remains stable before randomization (for at least 4 weeks in the case of antidepressants and for at least 2 weeks in the case of benzodiazepine and other drugs), and 2) patients are asked to avoid any drug-therapy change for the whole duration of the study.

Other references

Additional references

Altman 1996

Altman DG, Bland MJ. Detecting skewness for summary information. BMJ 1996;313:1200.

APA 2000

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th (TR) edition. Washington DC, 2000.

Bijl 1998

Bijl RV, Ravelli A, van Zessen G. Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Social Psychiatry and Psychiatric Epidemiology 1998;33:587-595.

Bucher 1997

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 Epidemiol1997;50(6):683-91.

Caldwell 2005

Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005;331(7521):897-900.

Cohen 1960

Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 1960;20:37-46. [DOI: 10.1177/001316446002000104]

Eaton 1994

Eaton WW, Kessler RC, Wittchen H-U, Magee WJ. Panic and panic disorder in the United States. American Journal of Psychiatry 1994;151:413-420.

Elbourne 2002

Elbourne DR, Altman DG, Higgins JP, Curtin F, Worthington HV, Vail A. Meta-analyses involving cross-over trials: methodological issues. International Journal of Epidemiology 2002;31:140-9. [DOI: 10.1093/ije/31.1.140]

Furukawa 2005

Furukawa TA, Cipriani A, Barbui C, Brambilla P, Watanabe N. Imputing response rates from means and standard deviations in meta-analysis. International Clinical Psychopharmacology 2005;20:49-52.

Furukawa 2006

Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. Journal of Clinical Epidemiology 2000;59:7-10.

Furukawa 2007

Furukawa TA, Watanabe N, Churchill R. Combined psychotherapy plus antidepressants for panic disorder with or without agoraphobia. Cochrane Database of Systematic Reviews 2007, Issue 1. Art. No.: CD004364. DOI: 10.1002/14651858.CD004364.pub2.

Glenny 2005

Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, Bradburn M, Eastwood AJ. Indirect comparisons of competing interventions. Health Technology Assessment 2005;9(26):1-134.

Goisman 1995

Goisman RM, Warshaw MG, Steketee GS, Fierman EJ, Rogers MP, Goldenberg I, Weinshenker NJ, Vasile RG, Keller MB. DSM-IV and the disappearance of agoraphobia without a history of panic disorder: new data on a controversial diagnosis. American Journal of Psychiatry 1995;152:1438-1443.

Gorman 2000

Gorman JM, Kent JM, Sullivan GM, Coplan JD. Neuroanatomical hypothesis of panic disorder, revised. American Journal of Psychiatry 2000;157:493-505.

Grant 2006

Grant BF, Hasin DS, Stinson FS, Dawson DA, Goldstein RB, Smith S, Huang B, Saha TD. The epidemiology of DSM-IV panic disorder and agoraphobia in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Journal of Clinical Psychiatry 2006;67:363-374.

Higgins 2003

Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557-60.

Higgins 2011

Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions (version 5.1.0). The Cochrane Collaboration, 2011.

Higgins 2012

Higgins JPT,Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. Research Synthesis Methods 2012;3(2):98-110.

Kessler 2006

Kessler RC, Chiu WT, Jin R, Ruscio AM, Shear K, Walters EE. The epidemiology of panic attacks, panic disorder, and agoraphobia in the National Comorbidity Survey Replication. Archives of General Psychiatry 2006;63:415-424.

King 2008

King M, Nazareth I, Levy G, Walker C, Morris R, Weich S, Bellón-Saameño JA, Moreno B, Svab I, Rotar D, Rifel J, Maaroos HI, Aluoja A, Kalda R, Neeleman J, Geerlings MI, Xavier M, de Almeida MC, Correa B, Torres-Gonzalez F. Prevalence of common mental disorders in general practice attendees across Europe. British Journal of Psychiatry 2008;192:362-367.

Lu & Ades 2004

Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004;30(20):3105-24.

Lumley 2002

Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002;21(16):2313-24.

NICE 2011

NICE clinical guidelines: Generalised anxiety disorder and panic disorder (with or without agoraphobia) in adults: Management in primary, secondary and community care. 2011;CG113.

Salanti 2008

Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomized trials. Stat Methods Med Res 2008;17:279. [DOI: 10.1177/0962280207080643]

Salanti 2009

Salanti G, Marinho V, Higgins JP. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. J Clin Epidemiol 2009;62(8):857-64.

Salanti 2011

Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. Journal of Clinical Epidemiology 2011;64:163-171.

Salanti 2012

Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Research Synthesis Methods 2012;3(2):80-97.

Starcevic 2009

Starcevic V. Anxiety Disorders in Adults: a Clinical Guide. Oxford University Press, 2009.

Sánchez-Meca 2010

Sánchez-Meca J, Rosa-Alcázar AI, Marín-Martínez F, Gómez-Conesa A. Psychological treatment of panic disorder with or without agoraphobia: a meta-analysis. Clin Psychol Rev 2010;30(1):37-50.

Turner 2012

Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. Int J Epidemiol 2012;41(3):818-27.