This is an author produced version of Efficacy of Cognitive Behavioral Therapy for Generalized Anxiety Disorder in Older Adults: Systematic Review, Meta-Analysis and Meta-Regression.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/102819/

Article:

https://doi.org/10.1016/j.jagp.2016.06.006

Article available under the terms of the CC-BY-NC-ND licence (https://creativecommons.org/licenses/by-nc-nd/4.0/)
Efficacy of Cognitive Behavioral Therapy for Generalized Anxiety Disorder in Older Adults: Systematic Review, Meta-Analysis and Meta-Regression

Jo Hall MSc¹, Stephen Kellett DClin²,³, Raul Espinoza Berrios Phd⁴, Manreesh Kaur Bains DClin², and Shonagh Scott DClin²

¹Clinical Psychology Unit, Department of Psychology, University of Sheffield, Sheffield, UK
²Sheffield Health and Social Care NHS Foundation Trust, Sheffield, UK
³Centre for Psychological Services Research, University of Sheffield, Sheffield, UK
⁴Departamento de Administracion, Facultad de Administracion y Economia, Universidad de Santiago de Chile, Chile

Correspondence to:
Jo Hall
Clinical Psychology Unit
Department of Psychology
309 Western Bank
University of Sheffield
Sheffield, S10 2TP, UK
Tel: 0114 222 6537
Fax: 0114 222 6610
j.hall7@sheffield.ac.uk

Conflicts of interest: None declared.

Key words:
Generalized anxiety disorder
Older adults
Cognitive behavioral therapy
Meta-analysis
Review
Objective: Generalized anxiety disorder (GAD) is a common disorder in older adults creating functional impairment, and psychotherapy is the preferred treatment option. Meta-analytic methods sought to determine the efficacy of outpatient cognitive behavioral therapy (CBT) with respect to the hallmark feature of GAD - uncontrolled and excessive worry. In order to optimize clinical applicability, variables associated with GAD treatment outcomes were also examined. Method: Systematic search of relevant databases and iterative searches of references from articles retrieved. All studies were required to have been a randomized control trial (RCT), to have used the Penn State Worry Questionnaire (PSWQ) or its abbreviated version (PSWQ-A) as an outcome measure, and to have conducted CBT with outpatient older adults. Fourteen RCTs (N = 985) were suitable and random-effects meta-analyses and univariate meta-regressions were conducted. Results: At the end of treatment, and six-month follow-up, significant treatment effects favoring CBT were found in comparison to a waitlist or treatment-as-usual. When CBT was compared with active controls, a small non-significant treatment advantage was found for CBT at the end of treatment, with equivalence of outcomes at follow-up. Treatment effect size of CBT for GAD was significantly associated with attrition rates and depression outcomes. Conclusions: CBT is more helpful than having no treatment for GAD in later life. However, whether CBT shows long-term durability, or is superior to other commonly available treatments (such as supportive psychotherapy), remains to be tested. The relationship between treatment effects for GAD and depression following CBT warrants further research.
INTRODUCTION

Generalized anxiety disorder (GAD) is the most common anxiety disorder in older adults with reported prevalence rates of between 2.4% and 6.3%.\textsuperscript{1, 2} The numbers affected by GAD may actually exceed these figures, given that subthreshold GAD is associated with significant disability.\textsuperscript{3} GAD is a chronic and disabling condition regardless of age, and in older adults it is associated with increased disability, cognitive impairment, reduced quality of life, and increased service use.\textsuperscript{4-8} High rates of comorbidity occur, with depression comorbidity rates of up to 60% reported.\textsuperscript{9} Numerous differences have been found in the functional connectivity of emotion-focused brain networks amongst older adults with GAD, illustrating abnormalities in both worry generation and worry reappraisal.\textsuperscript{10} Cognitive behavioral therapy (CBT) uses this neuroanatomical evidence to justify targeting uncontrolled and excessive worry during treatment.\textsuperscript{11} Services are frequently faced with the challenge of treating older adult GAD, with patients preferring psychotherapy when offered treatment choice.\textsuperscript{12} This review sought to quantify and synthesize the older adult evidence for the treatment of GAD with CBT in order to provide contemporary guidance to clinicians concerning effective treatment options.

Prior reviews of the treatment effects of CBT for GAD in older adults have given inconsistent conclusions.\textsuperscript{13-16} Reviews have also suggested that CBT may be less effective for older adults than it is for younger adults, because of the effect of cognitive decline due to aging and high rates of psychiatric comorbidity.\textsuperscript{17-19} However, there are two key weaknesses of the evidence base for talking treatments for older adults with GAD: lack of specificity and measurement issues. Firstly, existing reviews have tended to cover a wide range of psychotherapeutic treatment options or late-life anxiety disorders and have therefore unwittingly masked potential differences between specific psychotherapies.\textsuperscript{13-16, 20} Secondly,
previous reviews have measured effect sizes using a pooled anxiety composite, which has
diluted and obscured treatment effects with respect to the defining feature of GAD:
uncontrolled and excessive worry.\textsuperscript{21} For this reason, in the measurement of GAD outcomes,
researchers have been strongly encouraged to use the Penn State Worry Questionnaire
(PSWQ).\textsuperscript{21-23} The PSWQ is a validated measure of worry appropriate for use in older adults,
as is its abbreviated version, the PSWQ-A.\textsuperscript{24-27}

The present study has therefore been prompted by identified methodological
weaknesses of the existing evidence base for talking treatments for GAD in older adults.
To improve the quality of the evidence base of CBT for GAD, the current meta-analysis
included a greater number of older adult trials, did not use an anxiety composite outcome,
performed pre-planned subgroup meta-analyses on the basis of control group subtype, and
included a number-needed-to-treat analysis. This review focused (a) exclusively on
standardized trials of CBT for GAD in older adults, and (b) assessed treatment effects if, and
only if, the trial used the PSWQ or PSWQ-A as an outcome measure. In summary, the main
purpose of this meta-analysis was to test the efficacy of outpatient CBT for uncontrolled and
excessive worry in older adults with GAD.

METHODS

In conducting and reporting results, PRISMA\textsuperscript{28} guidelines for preferred reporting
items for meta-analyses are followed.

Search Strategy

Three electronic databases (PsychInfo, Web of Science, and ProQuest Dissertation
and Theses) were searched from Jan 1987 to Nov 2015. The date that the DSM-III-R\textsuperscript{29} was
published (1987) was the start date, as this was the first diagnostic manual to recognize GAD
as a distinct disorder, characterized by excessive worrying. The following title search string was used based on search terms used in related reviews\textsuperscript{14, 15}: \([\text{GAD OR generalized anxiety disorder OR generalised anxiety disorder OR anxious OR anxiety OR worry}] \text{ AND [older OR elder* OR geriat* OR late life OR late-life]} \text{ AND [CBT OR cognitive behavioural therapy OR cognitive behavioral therapy OR treatment OR therapy]}.\) Reference lists of retrieved articles, and prior reviews on the psychological treatment of late-life anxiety published in the last 10 years, were also searched manually to identify potentially eligible studies.

**Eligibility Criteria**

Firstly, participants needed to have been at least 55 years old, with a mean age of $\geq 65$ years, and to have a principal or co-principal diagnosis of GAD. In mixed anxiety studies, 75\% of participants were required to have a principal or co-principal diagnosis of GAD.\textsuperscript{14} Secondly, studies needed to have been a randomized controlled trial (RCT). Thirdly, the CBT arm needed to have included psycho-education, cognitive restructuring, and exposure as treatment components.\textsuperscript{15} Finally, studies needed to have used the PSWQ or the PSWQ-A as an outcome measure.\textsuperscript{21, 23}

**Data Extraction**

An a priori data extraction coding frame was developed. Studies were coded for trial and practice factors including control type (waitlist, TAU, or active treatment) and treatment mode (individual or group). Clinical variables extracted included depression outcomes. Follow-up data was extracted in order to conduct treatment durability analyses. The percentage of treatment responders was calculated using an intention-to-treat analysis; drop-outs were classified as non-responders.\textsuperscript{30}
Within-Study Quality and Risk of Bias

The Cochrane Common Mental Disorders Anxiety and Neurosis Group (CCDAN) quality assessment tool\textsuperscript{31} was used to assess methodological quality; higher scores indicated greater methodological quality (possible scores ranged from 0 to 46). Three raters (all clinical psychologists) rated each study blind and independently; interrater reliability was calculated using Fleiss’ kappa.\textsuperscript{32} To assess within-study bias, the Cochrane Risk of Bias Tool was used.\textsuperscript{33}

Between-Group Effect Sizes

Effect sizes corresponded to the standardized difference between the CBT and controls.\textsuperscript{34} Between-group end of treatment effect sizes were calculated as: (CBT group end of treatment score – control group end of treatment score)/Pooled SD.\textsuperscript{34} Effect sizes were based on completers-only data, as intention-to-treat outcome data was not available for all studies. As a number of trials had small samples, effect sizes were corrected using an adjustment, J, to convert effect sizes to Hedges’ g.\textsuperscript{35} For studies in which multiple treatment arms received CBT, data was collapsed to form one group where treatment was comparable,\textsuperscript{36,37} and if not, data from the most relevant CBT group was extracted.\textsuperscript{38} In studies in which multiple comparison groups did not receive CBT, data was extracted from the most active comparison condition.\textsuperscript{38,39} This enabled a more conservative estimate of population effect size, given that passive controls often result in larger effect sizes than active controls.\textsuperscript{40}

Meta-Analysis

A random-effects meta-analysis was conducted to provide a more realistic estimate of pooled mean effect size and to increase the generalizability of overall findings, given
between-study heterogeneity was anticipated.\textsuperscript{14, 34, 41} Weighted average Hedges’ g effect sizes\textsuperscript{35} were calculated from the sum of the inverse within-study variance ($W = 1/V_g$).\textsuperscript{34} and the between-study variance was calculated based on the restricted maximum likelihood effect size method (REML).\textsuperscript{42} REML is more sensitive in meta-analyses including smaller studies.\textsuperscript{43}

Mean effect sizes obtained were reversed and a positive effect of CBT was represented by a positive effect size, and vice-versa. The threshold for statistical significance was an alpha value of 0.05. Effect sizes were classified as follows: 0.20-0.49 = small, 0.50-0.79 = medium, and \geq0.80 = large.\textsuperscript{44} Pooled mean effects sizes for end of treatment and six-month follow-up data were calculated and subgroup meta-analyses were pre-planned on the basis of anticipated heterogeneity between control groups (i.e. waitlist, TAU, and active).\textsuperscript{14} Effect sizes were then translated into the expected number of patients needed to be treated for one additional beneficial outcome (NNTB)\textsuperscript{45} using the following formula: $1/(2 \times \text{AUC} - 1)$.\textsuperscript{46}

To assess variables associated with CBT effect size, two methods were used. First, to assess categorical variables (control type and treatment mode) the analog to a one-way ANOVA was computed using the METAF macro,\textsuperscript{41} significant variables were indicated by a significant homogeneity Q statistic.\textsuperscript{47} Second, continuous variables were assessed using the METAREG macro\textsuperscript{41} which computes random-effects univariate meta-regressions; significance was indicated by a beta value of $p <0.006$ based on a Bonferroni adjustment to the significance level due to multiple testing.\textsuperscript{47} Eight pre-specified variables were assessed: age (mean), attrition rate (%), number of CBT sessions, baseline co-morbid psychiatric diagnoses (%), baseline depression diagnoses (%), CBT vs. any control post-treatment depression effect size, mean baseline pathological worry and depression scores (both based on standardized z scores).

\textbf{Analysis of Statistical Heterogeneity}
The Q-statistic was used to detect unexplained statistical heterogeneity between studies. Due to the small number of trials (k<10) included in sub-group and six-month follow-up analyses, a p-value of 0.1 was adopted. The I² statistic was an indicator of statistical inconsistency within meta-analyses; when 0-40% might not be important, 30-60% may represent moderate heterogeneity, 50-90% may represent substantial heterogeneity, 75-100% considerable heterogeneity. The interpretation of I² values was based on the magnitude and direction of effect size and evidence for heterogeneity.

Publication Bias

A funnel plot provided a graphical representation of the relationship between the standard error of included trials and associated effect sizes; the presence of asymmetry was considered potentially indicative of publication bias. As recommended, supplementary tests were then used to assess publication bias: Macaskill’s funnel plot regression method and Begg’s rank correlation method.

RESULTS

Study Selection

The initial search resulted in 428 potentially relevant titles (Figure 1), of which 273 titles remained after duplicate removal. On the basis of study abstract 132 papers were excluded, and of 141 papers retrieved for detailed consideration a further 124 papers were then excluded (reasons specified in Figure 1). Two of the remaining 17 studies were excluded, due to duplicate data from more appropriate eligible articles. A final study was excluded because outcome data had been reported as an anxiety composite. Fourteen RCTs met all inclusion criteria and so were included in this review. The total
sample for the meta-analysis was N = 985 with an average age of 68.16 years (SD = 2.52). On average, over half of participants (59.6%) met criteria for at least one other psychiatric diagnosis, with around a third (31.4%) having a diagnosis of a depressive disorder.

Study Characteristics

Table 1 organizes studies by control subtype category and quality ratings. Nine trials compared CBT against a passive control condition. Participants in all three CBT vs. wait-list trials were recruited via advertising and assumed not to be in contact with services during the wait period. TAU participants received contact of varying intensity, and in 3/6 studies this included weekly contact. In the five active control trials, the following controls were used: non-directive psychotherapy (either face-to-face or telephone-delivered), discussion group, acceptance and commitment therapy, and escitalopram. CBT dropout rates ranged from zero to 44.4%.

Individual CBT was the most common delivery method, although 3/14 studies delivered group CBT. The duration of CBT ranged from 8 to 16 weeks. CBT was typically delivered face-to-face (12/14 studies); two studies had telephone delivery. Follow-up data was sparse, with 6/14 studies presenting (treatment-free) six-month follow-up data for both CBT and comparison groups. Three trials had no definition of treatment response, with the remainder providing inconsistent definitions. CBT response rate ranged from 19.2% to 83.3% (M = 44.7, SD = 19.9).

Study quality varied, and ratings ranged from 19 to 41 (out of 46). However, average quality ratings within the three control subgroups were similar; between 33 and 34 (out of 46). Excellent inter-rater reliability for quality ratings was observed (κ = 0.99; 95% CI[0.94,
Efficacy of CBT for GAD in Older Adults: A Review

Ten of the trials were considered at low/low-medium risk of study bias and one trial\(^\text{60}\) was at high risk of bias. However, seven trials did not provide adequate details of the process of random sequence generation, indicating risk of selection bias. Furthermore, seven trials did not report blinding of outcome, indicating risk of detection bias.

Meta-Analyses

Between-group random effects meta-analyses were conducted for end of treatment and six-month follow-up PSWQ/PSWQ-A data. Control subgroup meta-analyses were conducted at each time-point.

End of Treatment Analysis

For all 14 trials (completer \(n = 772\)), the end of treatment population effect size estimate for CBT compared to any control was medium, and in favor of CBT, \(g = 0.66\) (95% CI: 0.42–0.90; \(z = 5.48, p <0.001\)) (Figure 2 plot d). Significant statistical heterogeneity was found between studies (\(Q_{(13)} = 28.67, p = 0.001, v = 0.10, I^2 = 54.7\%\)). The population effect size estimate for CBT compared to waitlist controls (\(k = 3, n = 86\)) was large, and in favor of CBT, \(g = 1.10\) (95% CI: 0.38–1.82; \(z = 3.01, p <0.001\)) (Figure 2 plot a). Between-study heterogeneity was substantial (\(Q_{(2)} = 5.38, p = 0.07, I^2 = 62.8\%, v = 0.25\)). For CBT compared to TAU (\(k = 6, n = 444\)) the population effect size was medium, and in favor of CBT, \(g = 0.67\) (95% CI: 0.36–0.98; \(z = 4.22, p<0.001\)) (Figure 2 plot b). The corresponding NNTB suggested that one out of every three patients would be expected find additional benefit from CBT when compared to TAU at the end of treatment. Between-study heterogeneity was
moderate ($Q_{(5)} = 9.67, p = 0.09, \Gamma^2 = 48.3\%, \nu = 0.07$). When CBT was compared to active controls ($k = 5, n = 242$) the population effect size estimate was small, $g = 0.42$ (95% CI: -0.05–0.89), and non-significant ($z = 1.75, p = 0.08$) (Figure 2 plot c). Thus, CBT was not found to be significantly superior to active treatments. The corresponding NNTB indicated that one out of every four patients would be expected to find additional benefit from CBT in comparison an active intervention at the end of treatment. Substantial between-study heterogeneity was found ($Q_{(4)} = 11.53, p = 0.02, \Gamma^2 = 65.3\%, \nu = 0.18$).

Insert Figure 2 here please

**Follow-up Analysis**

The population effect size estimate for CBT compared to any control group at six-month follow-up ($k = 5, n = 238$), was in the small-to-medium range in favor of CBT, $g = 0.46$ (95% CI: 0.07–0.85; $z = 2.28, p = 0.02$) (Figure 2 plot g). Studies violated the assumption of statistical homogeneity ($Q_{(4)} = 8.24, p = 0.08, \Gamma^2 = 51.5\%, \nu = 0.10$). Due to the paucity of passive control studies that had six-month control follow-up data, waitlist and TAU studies were considered as a single passive control subgroup. The follow-up population effect size estimate for CBT compared to passive controls ($k = 2, n = 170$) was large, and in favor of CBT, $g = 0.83$ (95% CI: 0.52–1.14; $z = 5.21, p < 0.001$) (Figure 2 plot e). Studies were statistically homogenous ($Q_{(1)} = 0.03, p = 0.86, \Gamma^2 = 0\%, \nu = 0.00$). The population effect size estimate for CBT compared to active controls at follow-up ($k = 3, n = 68$) was near zero, $g = 0.06$ (95% CI: -0.37–0.49) and non-significant ($z = 0.28, p = 0.78$) (Figure 2 plot f). Thus, no significant advantage was found for either CBT or active controls at follow-up. Between-study statistical homogeneity was observed ($Q_{(2)} = 0.19, p = 0.91; \Gamma^2 = 0\%, \nu = 0.00$).
Meta-Regression Analysis

Random-effects univariate meta-regression found depression effect size was significantly associated with PSWQ/PSWQ-A effect size ($\beta = 0.60$, $z = 2.76$, $p = 0.0057$). Therefore, trials with greater depression treatment effects in favor of CBT (when compared to any control), were associated with greater GAD treatment effects in favor of CBT (when compared to any control). Attrition rate was also significantly associated with PSWQ/PSWQ-A effect size ($\beta = -0.62$, $z = -2.89$, $p = 0.0039$), and trials with higher attrition rates were found to have worse GAD treatment effects following CBT (when compared to any control). No other variables (categorical or continuous) were significantly associated with PSWQ/PSWQ-A effect size.

Reporting Bias

Inspection of the funnel plot (Figure 3) suggested potential reporting bias, as the study distribution around the pooled mean effect size was slightly asymmetrical. However, the funnel plot regression method ($B = -0.001$, $t(13) = -0.63$, $p = 0.54$), and Begg’s rank correlation method based on 10,000 resamples ($\tau = -0.001$, $SE = 0.27$), did not indicate significant reporting bias. Therefore, the overall population effect size estimate was likely to be relatively robust.

Insert Figure 3 here please

DISCUSSION

This review has tested the efficacy of CBT for older adults with GAD in terms of its defining feature of uncontrolled and excessive worry. Inclusion of a greater number of trials, and increased specificity (in terms of trials reviewed and the measurement of treatment
Efficacy of CBT for GAD in Older Adults: A Review

Outcomes, has increased the validity of results compared to previous work. In comparison to a waitlist, CBT was found to produce a large effect with respect to reduced tendency to worry immediately following treatment. The associated NNTB value suggested that one out of every two patients receiving CBT would be expected to find additional benefit (in terms of reduced worry) when compared to a waitlist. Results of CBT in comparison to TAU found medium treatment effects in favor of CBT. At six-month follow-up, large effects in favor of CBT were observed in comparison to passive control conditions. The corresponding NNTB value indicated that one out of every two patients would be expected to gain additional benefit from CBT at six-month follow-up when compared to a passive control. Combined, these findings suggest that when compared to a waitlist or TAU, CBT is efficacious for older adults with GAD. When comparisons were made of CBT with active controls results were less convincing. Findings suggested a slight advantage for CBT over active treatment at the end of treatment, with equivalent outcomes at follow-up. Furthermore, the magnitude of CBT treatment effects when compared to a range of controls adds to evidence suggesting that CBT for GAD may be less effective for older adults than it is for younger adults.21,23

Limitations and Future Directions

The present review has a number of limitations, which usefully highlight how the evidence base could be further developed. A number of the studies were found to have significant risk of bias, such as the randomization process being only fully described in seven of the trials. Future trials need to report randomization processes in full. The use of passive control conditions in trials is fraught with limitations. For example, waitlist controls have been shown to inflate treatment effect sizes in comparison to ‘no treatment’ control conditions,65 and TAU conditions are often heterogeneous reducing the generalizability of findings.66 Only one trial62 compared CBT to another evidence-based psychotherapy (CBT
versus ACT), and so future studies certainly need to use valid active controls. There were also relatively few studies containing sufficient follow-up data, increasing the risk of positive selection bias and an inflated effect size estimate. Longer follow-up periods in future trials are required. The finding that depression treatment effects were associated with GAD treatment outcomes highlights potential areas for clinical innovation, particularly given current interest in transdiagnostic approaches for the treatment of comorbid anxiety and depression.19, 67

A number of the meta-analyses were statistically heterogeneous, reducing the generalizability of conclusions. This may have reflected variation within control subgroups, such as differing definitions of TAU.66 The inclusion of a number of small studies may have also induced a ‘small-study effect’, whereby smaller studies show larger treatment effects and so positively bias meta-analytic findings.68 Future trials need to be sufficiently powered to detect differences between treatment arms. The per protocol analyses in some studies also meant that completers-only effect size estimates were available and so intention-to-treat analyses are desirable for future trials.30 This is important considering the present finding that attrition from CBT reduces GAD outcomes and interventions for ensuring treatment completion should also be tested. Trials need to report response rates using a consistent definition of recovery from GAD; the reliable and clinically significant change criteria appears useful.69 Future psychotherapy trials for older adult GAD could also usefully assess treatment effects of CBT based on functional neuroanatomical outcomes.10

**CONCLUSION**

In a meta-analysis of gold standard clinical trials, CBT has been found to be an efficacious treatment for uncontrolled and excessive worry in older adults with GAD. Findings suggest that CBT should be routinely offered to older adults presenting to services
with GAD. However, results do not provide evidence that CBT is durable or more efficacious than other psychological interventions. There is, therefore, a real need for further sufficiently powered ‘head-to-head’ RCTs (with longer follow-up periods) to be conducted, to enable the comparative efficacy and durability of CBT treatment to be firmly established. Preliminary findings regarding attrition from treatment, and of a relationship between treatment effects for symptoms of GAD and depression following CBT, are also important avenues for further examination.
References

generalized anxiety and options for DSM-V. Depress Anxiety 2010; 27:134-147


26. Wuthrich VM, Johnco C, Knight A: Comparison of the Penn State Worry Questionnaire (PSWQ) and abbreviated version (PSWQ-A) in a clinical and non-clinical population of older adults. J Anxiety Disord 2014; 28:657-663


43. Jackson D, Bowden J, Baker R: How does the Der Simonian and Laird procedure for random effects meta-analysis compare with its more efficient but harder to compute counterparts? J Stat Plan and Infer 2010; 140:961-970


46. Kraemer HC, Kupfer DJ: Size of treatment effects and their importance to clinical research and practice. Biol Psychiat 2006; 59:990-996


<table>
<thead>
<tr>
<th>Author (Year)</th>
<th>N</th>
<th>CBT format (N sessions, min)</th>
<th>Control condition</th>
<th>CBT attrition % (Dropout CBT, Dropout control)</th>
<th>Follow-up period mths CBT (Control)</th>
<th>Response definition</th>
<th>ITT response rate % CBT (Control)</th>
<th>CCDAN score (Overall bias rating)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBT vs. Waitlist</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brenes et al. (2012)</td>
<td>60</td>
<td>Individual (12)</td>
<td>Information only</td>
<td>Not stated (13.3) (3.3)</td>
<td>6 (6)</td>
<td>Not defined</td>
<td>N/A</td>
<td>36</td>
</tr>
<tr>
<td>Mohlman &amp; Gorman (2005)</td>
<td>32</td>
<td>Individual (13) (13x50min)</td>
<td>Waitlist</td>
<td>Not stated (9.0) (0)</td>
<td>12 (0)</td>
<td>No GAD and 20% reduction in 75% of OMs</td>
<td>50.0 (0)</td>
<td>Low-medium</td>
</tr>
<tr>
<td>Mohlman (2008)</td>
<td>8</td>
<td>Individual (8) (8x90min)</td>
<td>Waitlist (CBT /APT group)</td>
<td>Not stated (0) (0)</td>
<td>6 (0)</td>
<td>No GAD and 20% reduction in 80% of OMs</td>
<td>50.0 (0)</td>
<td>Medium-high</td>
</tr>
<tr>
<td>Stanley et al. (2014)</td>
<td>223</td>
<td>Individual (10 plus booster)</td>
<td>Usual care</td>
<td>24.7 (13.4) (22.0)</td>
<td>0 (0)</td>
<td>20% reduction in 75% of OM</td>
<td>29.3 (17.8)</td>
<td>Low</td>
</tr>
<tr>
<td>Stanley et al. (2009)</td>
<td>134</td>
<td>Individual (10)</td>
<td>Enhanced usual care, biweekly calls</td>
<td>22.3 (7.0) (22.0)</td>
<td>15 (15)</td>
<td>Meaningful change in 50% of OM</td>
<td>54.3 (48.4)</td>
<td>Low</td>
</tr>
<tr>
<td>Wetherell et al. (2009)</td>
<td>31</td>
<td>Individual (12)</td>
<td>Enhanced community care</td>
<td>Not stated (20.0) (Not stated)</td>
<td>0 (0)</td>
<td>Not stated</td>
<td>N/A</td>
<td>36</td>
</tr>
<tr>
<td>Stanley et al. (2003b)</td>
<td>85</td>
<td>Group (15)</td>
<td>Minimal contact, weekly calls</td>
<td>22.4 (25.6) (14.6)</td>
<td>12 (0)</td>
<td>20% reduction in symptom severity</td>
<td>33.3 (7.3)</td>
<td>Low-medium</td>
</tr>
<tr>
<td>Gorenstein et al. (2005)</td>
<td>42</td>
<td>Individual (13 plus medication management) (13x50min)</td>
<td>Medication management, weekly contact 10-15 mins</td>
<td>34.9 (39.1) (26.3)</td>
<td>6 (0)</td>
<td>Improved or much improved</td>
<td>39.1 (26.3)</td>
<td>Low-medium</td>
</tr>
<tr>
<td>Stanley et al. (2003c)</td>
<td>12</td>
<td>Individual (8)</td>
<td>Usual care, weekly calls</td>
<td>Not stated (16.7) (33.3)</td>
<td>0 (0)</td>
<td>20% reduction in 67% of OM</td>
<td>83.3 (16.7)</td>
<td>High</td>
</tr>
<tr>
<td>Green et al. (2006)</td>
<td>12</td>
<td>Individual (8)</td>
<td>Usual care, weekly calls</td>
<td>Not stated (16.7) (33.3)</td>
<td>0 (0)</td>
<td>20% reduction in 67% of OM</td>
<td>83.3 (16.7)</td>
<td>High</td>
</tr>
<tr>
<td>Stanley et al. (2003c)</td>
<td>12</td>
<td>Individual (8)</td>
<td>Usual care, weekly calls</td>
<td>Not stated (16.7) (33.3)</td>
<td>0 (0)</td>
<td>20% reduction in 67% of OM</td>
<td>83.3 (16.7)</td>
<td>High</td>
</tr>
<tr>
<td>Green et al. (2006)</td>
<td>12</td>
<td>Individual (8)</td>
<td>Usual care, weekly calls</td>
<td>Not stated (16.7) (33.3)</td>
<td>0 (0)</td>
<td>20% reduction in 67% of OM</td>
<td>83.3 (16.7)</td>
<td>High</td>
</tr>
</tbody>
</table>
TABLE 1 (continued). Characteristics of the clinical trials included in the meta-analysis

<table>
<thead>
<tr>
<th>Author (Year)</th>
<th>N</th>
<th>CBT format (N sessions) (total length, min)</th>
<th>Control condition</th>
<th>Attritionb</th>
<th>Follow-up period mths</th>
<th>Response Definition</th>
<th>ITT response rate % CBT (Control)</th>
<th>CCBAN score (Overall bias ratingd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brener et al. (2015)</td>
<td>141</td>
<td>Individual telephone 9-11 (50 min p/session)</td>
<td>Telephone-directed, non-directive supportive therapy</td>
<td>13.10 (25.71)</td>
<td>0</td>
<td>5.5 point decrease in PSWQ-A scores</td>
<td>72.4 (42.9)</td>
<td>Low</td>
</tr>
<tr>
<td>Wetherell et al. (2003)</td>
<td>75 (36)</td>
<td>Group (12) (12x90min)</td>
<td>Discussion group</td>
<td>31.0 (31.0)</td>
<td>6</td>
<td>20% reduction in 75% of OM</td>
<td>23.1 (23.1)</td>
<td>Low-medium</td>
</tr>
<tr>
<td>Wetherell et al. (2013)</td>
<td>73 (34)</td>
<td>Individual (16 plus escitalopram)</td>
<td>Escitalopram</td>
<td>18.06 (0)</td>
<td>7h</td>
<td>HAM-A score ≤ 10 and decrease ≥ 8.5 points on the PSWQ</td>
<td>38.9 (15.8)</td>
<td>Low-medium</td>
</tr>
<tr>
<td>Stanley, Beck, Glassco (1996)</td>
<td>48</td>
<td>Group (14) (12x90min)</td>
<td>Supportive psychotherapy</td>
<td>Not stated (31.0)</td>
<td>6</td>
<td>20% reduction in 75% of OM</td>
<td>19.2 (35.0)</td>
<td>Medium-high</td>
</tr>
<tr>
<td>Wetherell et al. (2011)</td>
<td>21</td>
<td>Individual (12) (12x60min)</td>
<td>Acceptance and commitment Therapy</td>
<td>42.3 (44.4)</td>
<td>6</td>
<td>Not stated</td>
<td>N/A</td>
<td>Medium-high</td>
</tr>
</tbody>
</table>

Note: ITT: intention-to-treat; CCDAN: Cochrane Common Mental Disorders Anxiety and Neurosis Group quality assessment tool; TAU: treatment-as-usual; OM = outcome measures; PSWQ: Penn State Worry Questionnaire; PSWQ-A: Penn State Worry Questionnaire – Abbreviated; HAM-A: Hamilton Anxiety Rating Scale.

a CBT session duration (mins) is provided for those trials in which this was reported. b Attrition rate is based on total number of participants eligible for each trial pre-randomisation and is not reported for trials in which this was not explicitly stated. c Overall quality ratings from CCDAN tool out of 46, higher scores represent papers rated as higher quality. d Summary of assessed overall risk of bias for PSWQ/PSWQ-A outcome scores (low-high). e Medication management was comparable to control conditions described as TAU. f Trial contained two control groups therefore only data from the most active control (discussion group) was used. g Trial contained multiple treatment phases and control arms, therefore data from the most relevant were extracted for analyses (CBT plus Escitalopram vs. Escitalopram only), and participants that recovered in the acute phase, pre-randomisation, were excluded from analyses (n = 3). h Follow-up period was not treatment-free therefore data was excluded from follow-up analyses.
FIGURE 1. Flow chart of study selection

- Records identified through database searching (n = 426)
  - Duplicated abstracts excluded (n = 155)
- Records after duplicates removed (n = 273)
  - Abstracts excluded that did not meet broad criteria (n = 132):
    - Medication trials (n = 49), reviews (n = 28), younger samples (n = 17), other, e.g. editorial (n = 35)
  - Full-text articles excluded that did not meet specific inclusion criteria (n = 124):
    - 75% of the sample did not have a principal or co-principal diagnosis of GAD (n = 50), not an intervention study (n = 35), uncontrolled (n = 16), no CBT used (n = 14), mean age not ≥65 years (n = 6), PSWQ/PSWQ-A not used as an outcome measure (n = 3)
- Records screened (n = 273)
- Full-text articles assessed for eligibility (n = 141)
- Papers initially considered for the review (n = 17)
  - Papers excluded as data from same trial published twice (n = 2)
- Eligible trials (k = 15)
  - Trial excluded as PSWQ/PSWQ-A outcome data unavailable (k = 1)
- Trials included in meta-analysis (k = 14)
FIGURE 2. End of treatment, and six-month follow-up, forest plots of PSWQ/PSWQ-A Hedge’s g effect sizes (g), standard errors (S.E.), confidence intervals (95% C.I.) for CBT vs. control conditions (a-g); n: completer sample size, NNTB: number of patients needed to be treated for one additional beneficial outcome, I²: measure of inconsistency across findings
### Efficacy of CBT for GAD in Older Adults: A Review

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>g</th>
<th>S.E.</th>
<th>95% C.I.</th>
<th>Pooled Hedges' g effect size (Random), 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a. CBT vs. Waitlist</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breines et al. (2012)</td>
<td>95</td>
<td>0.60</td>
<td>0.28</td>
<td>0.06, 1.14</td>
<td></td>
</tr>
<tr>
<td>Mohrman et al. (2005)</td>
<td>23</td>
<td>1.07</td>
<td>0.40</td>
<td>0.29, 1.77</td>
<td></td>
</tr>
<tr>
<td>Mohrman (2008)</td>
<td>8</td>
<td>1.93</td>
<td>0.86</td>
<td>0.22, 3.14</td>
<td></td>
</tr>
<tr>
<td><strong>Subtotal (NNTB = 2)</strong></td>
<td>96</td>
<td>1.10</td>
<td>0.37</td>
<td>0.38, 1.82</td>
<td></td>
</tr>
<tr>
<td><strong>b. CBT vs. TAU</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goreinstein et al. (2006)</td>
<td>28</td>
<td>0.52</td>
<td>0.38</td>
<td>-0.24, 1.27</td>
<td></td>
</tr>
<tr>
<td>Stanley et al. (2003b)</td>
<td>70</td>
<td>1.06</td>
<td>0.27</td>
<td>0.54, 1.58</td>
<td></td>
</tr>
<tr>
<td>Stanley et al. (2003c)</td>
<td>9</td>
<td>1.06</td>
<td>0.72</td>
<td>-0.32, 2.46</td>
<td></td>
</tr>
<tr>
<td>Stanley et al. (2009)</td>
<td>129</td>
<td>0.90</td>
<td>0.20</td>
<td>0.52, 1.28</td>
<td></td>
</tr>
<tr>
<td>Stanley et al. (2014)</td>
<td>160</td>
<td>0.33</td>
<td>0.15</td>
<td>0.03, 0.64</td>
<td></td>
</tr>
<tr>
<td>Wetherell et al. (2009)</td>
<td>20</td>
<td>0.35</td>
<td>0.36</td>
<td>-0.36, 1.07</td>
<td></td>
</tr>
<tr>
<td><strong>Subtotal (NNTB = 3)</strong></td>
<td>444</td>
<td>0.87</td>
<td>0.16</td>
<td>0.36, 0.98</td>
<td></td>
</tr>
<tr>
<td><strong>c. CBT vs. Active Treatment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breines et al. (2015)</td>
<td>129</td>
<td>0.97</td>
<td>0.19</td>
<td>0.60, 1.33</td>
<td></td>
</tr>
<tr>
<td>Stanley et al. (1996)</td>
<td>30</td>
<td>0.06</td>
<td>0.36</td>
<td>-0.63, 0.76</td>
<td></td>
</tr>
<tr>
<td>Wetherell et al. (2003)</td>
<td>30</td>
<td>0.17</td>
<td>0.33</td>
<td>-0.47, 0.91</td>
<td></td>
</tr>
<tr>
<td>Wetherell et al. (2011)</td>
<td>12</td>
<td>-0.29</td>
<td>0.54</td>
<td>-1.33, 0.77</td>
<td></td>
</tr>
<tr>
<td>Wetherell et al. (2013)</td>
<td>28</td>
<td>0.75</td>
<td>0.36</td>
<td>0.05, 1.45</td>
<td></td>
</tr>
<tr>
<td><strong>Subtotal (NNTB = 4)</strong></td>
<td>242</td>
<td>0.42</td>
<td>0.24</td>
<td>-0.05, 0.89</td>
<td></td>
</tr>
<tr>
<td><strong>d. CBT vs. Any Control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (NNTB = 3)</td>
<td>772</td>
<td>0.66</td>
<td>0.12</td>
<td>0.42, 0.90</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>g</th>
<th>S.E.</th>
<th>95% C.I.</th>
<th>Pooled Hedges' g effect size (Random), 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>e. CBT vs. Passive Control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breines et al. (2012)</td>
<td>55</td>
<td>0.79</td>
<td>0.30</td>
<td>0.20, 1.35</td>
<td></td>
</tr>
<tr>
<td>Stanley et al. (2006)</td>
<td>115</td>
<td>0.85</td>
<td>0.22</td>
<td>0.43, 1.27</td>
<td></td>
</tr>
<tr>
<td><strong>Subtotal (NNTB = 2)</strong></td>
<td>170</td>
<td>0.83</td>
<td>0.16</td>
<td>0.62, 1.14</td>
<td></td>
</tr>
<tr>
<td><strong>f. CBT vs. Active Control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stanley et al. (1996)</td>
<td>21</td>
<td>-0.06</td>
<td>0.30</td>
<td>-0.75, 0.64</td>
<td></td>
</tr>
<tr>
<td>Wetherell et al. (2003)</td>
<td>35</td>
<td>0.13</td>
<td>0.34</td>
<td>-0.53, 0.60</td>
<td></td>
</tr>
<tr>
<td>Wetherell et al. (2011)</td>
<td>12</td>
<td>0.16</td>
<td>0.59</td>
<td>-0.91, 1.21</td>
<td></td>
</tr>
<tr>
<td><strong>Subtotal (NNTB = 30)</strong></td>
<td>68</td>
<td>0.06</td>
<td>0.22</td>
<td>-0.37, 0.48</td>
<td></td>
</tr>
<tr>
<td><strong>g. CBT vs. Any Control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (NNTB = 4)</td>
<td>238</td>
<td>0.46</td>
<td>0.20</td>
<td>0.07, 0.85</td>
<td></td>
</tr>
</tbody>
</table>

### Efficacy of CBT for GAD in Older Adults: A Review

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>g</th>
<th>S.E.</th>
<th>95% C.I.</th>
<th>Pooled Hedges' g effect size (Random), 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Follow-up analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 3. Funnel plot of end of treatment PSWQ/PSWQ-A Hedges’ g effect sizes from all primary studies included in the meta-analysis (k =14)