Predictors and moderators of treatment outcome in late-life anxiety: A systematic review

Authors

Theofanis Kyriacou[1], Jade Hodges[1], Rebecca L. Gould[1]

[1] Division of Psychiatry, University College London, London, UK

***Corresponding author.** Theofanis Kyriacou, Tel.: Division of Psychiatry, University College London, Maple House, 149 Tottenham Court Road, London W1T 7NF

Abstract

Background: The aim of this review was to identify and critically appraise predictors and moderators of outcomes of psychological and pharmacological treatments for late-life anxiety disorders. Their identification may guide the development of personalised treatments for older people with anxiety disorders.

Methods: Web of Science, PsychINFO, CINAHL, Embase, and Pubmed were searched for studies published up to 12 May 2022. Randomised controlled trials and observational studies reporting treatment predictors and moderators were included. Participants with a diagnosis of any anxiety disorder who were aged over 60 years were included. Treatment outcomes included response, remission, and change in anxiety score.

Results: Thirteen studies met the inclusion criteria. Twenty-three out of 49 predictors or moderators assessed at post-treatment, and 14 out of 33 predictors or moderators assessed at follow-up were statistically significant. Only one predictor, baseline worry severity at post-treatment, was reported in at least three studies. Most studies were rated as having a low risk of bias in at least three areas and satisfied important quality criteria for predictor and moderator analyses.

Limitations: Samples were predominantly white, female and highly educated, and most studies were secondary analyses.

Conclusions: There is evidence that baseline worry severity appears to predict treatment outcome in late-life anxiety disorders. However, this was only explored in psychological intervention studies and therefore its predictive ability in pharmacotherapy remains unknown. Future research should explore predictors and moderators in a range of anxiety disorders and design methodologically-strong and adequately-powered studies with the primary aim of assessing predictors of treatment outcomes.

Keywords: systematic review, predictor, moderator, anxiety disorders, late life.

Introduction

Anxiety disorders are one of the most prevalent psychiatric disorders in late life (Porensky et al., 2009). Their 12-month prevalence is estimated to be about 10% and their lifetime prevalence is believed to be up to 15.3% (Bower et al., 2015). If late-life anxiety disorders remain untreated, the consequences are large and include an increased risk of late-life major depression, lower quality of life, and higher healthcare consumption (Hendriks et al., 2010). Current recommendations for the treatment of anxiety disorders in older adults include pharmacological interventions such as selective serotonin reuptake inhibitors and serotonin norepinephrine reuptake inhibitors, and psychological interventions such as cognitive behavioural therapy (APA, 2009; NICE, 2011; 2013).

Identifying sub-populations of older adults with late-life anxiety who are likely to benefit from a given pharmacological or psychological treatment is important for several reasons. First, it may inform treatment decisions about how to appropriately tailor treatments to specific individuals, leading to more personalised interventions. Second, it may facilitate the adjustment of such personalised interventions in a more timely fashion. Third, it may improve outcomes of existing treatments in older adults with anxiety disorders (Hult et al., 2019).

Tailoring pharmacological or psychological treatments for older adults with anxiety disorders can be accomplished by considering biopsychosocial factors, such as demographic and clinical characteristics, that act as predictors and/or moderators of clinical improvement following late-life anxiety treatment. Predictors are variables that influence the likelihood of a specific treatment outcome, whether positive or negative, irrespective of the treatment condition (Nierenberg, 2003). They offer valuable prognostic information regarding the potential effectiveness of treatment for an individual (Papakostas & Fava, 2008). On the other hand, differential predictors, also known as moderators, are variables that interact significantly with a specific treatment condition (Papakostas & Fava, 2008). For instance, in a study on collaborative care for anxiety in adults, gender moderated treatment outcomes, with women responding more favourably to the collaborative care intervention than to usual care, whilst no such association was observed among men (Grubbs et al., 2015). This example demonstrates the potential utility of moderators in identifying subgroups of older adults with

late-life anxiety who may derive greater benefits from a specific treatment (Kraemer et al., 2002). From a clinical perspective, moderators are helpful in precisely identifying those who might be most suitable for a particular intervention, as well as subgroups that may require additional therapeutic support. Therefore, moderators can serve as the basis of personalised medicine by indicating the optimal approach for adapting treatments to the specific needs of individual patients.

A previous meta-analysis identified seven predictors of treatment outcome in late-life depression that were consistently reported in at least three studies: baseline depression severity, executive dysfunction, early improvement, physical illness, current episode duration, baseline anxiety symptoms, and age (Tunvirachaisakul et al., 2018). However, to the authors' knowledge, no previous systematic reviews have examined predictors and/or moderators of treatment outcome in late-life anxiety disorders. Consequently, the aim of this systematic review was to identify and critically appraise predictors and moderators of treatment outcome in studies of pharmacological and/or psychological interventions for late-life anxiety disorders. Outcomes of interest included treatment response, remission, or change in score on standardised anxiety questionnaires.

Method

Study protocol

The PRISMA guidelines for reporting systematic reviews and meta-analyses were followed (Moher et al., 2009) and the PRISMA checklist is listed in Supplementary Material 1. The review protocol was pre-registered with the PROSPERO database of systematic review protocols (registration number CRD42021242286).

Eligibility criteria

The inclusion criteria were:

1. Randomised controlled trials (RCTs), non-randomised controlled trials, cohort studies, cross-sectional studies, and case-control studies.

2. A sample size greater than 10 participants in observational studies and greater than 5 in each arm in RCTs.
3. Studies where the mean, median or modal age of participants was 60 years or more, but the minimum age of participants was 50 years or more, in accordance with definitions of 'older people' used by the United Nations (2007) and the Department of Health (2001). Specifically, the United Nations defines older people as 60 years and older, whilst the Department of Health defines 'old age' as a socially constructed definition that can include people as young as 50 or from the official retirement age of 60 for women and 65 for men. Studies that involved both older and younger-age people were included if age-specific analyses were reported.
4. Participants with a principal diagnosis of an anxiety disorder based on DSM or ICD criteria.
5. Any pharmacological intervention (e.g., benzodiazepines, selective serotonin reuptake inhibitors, etc.), psychological or psychosocial intervention (e.g., cognitive behavioural therapy, psychoeducation, problem-solving or stress-management skills, etc.) or care management intervention (e.g., collaborative care intervention, etc.).
6. Reported statistical data on predictors and/or moderators of treatment outcome in late-life anxiety. Treatment outcome was categorised as treatment response (i.e., a pre-defined change in score on a measure of anxiety), remission (i.e., absence of anxiety disorder defined as scoring below a pre-defined threshold on an anxiety scale), and/or change in the score on standardised anxiety questionnaires (i.e., an unspecified change in score on a measure of anxiety).
7. Published in peer-reviewed journal studies and written in English.

The exclusion criteria were:

1. Reviews, non-human studies, case reports, qualitative studies, meeting or conference abstracts, and study protocols.
2. Studies focusing solely on people with post-traumatic stress disorder, as this reflects the change in DSM 5 criteria whereby PTSD is no longer classified as an anxiety disorder (APA, 2013).

Search strategy

Search terms listed below were combined using the Boolean AND operator:

i) anxiety OR anxious OR worry OR worried OR panic OR GAD OR phobi* OR agoraphobia;

ii) 'older adult' OR 'old age' OR elder* OR geriatr* OR 'late onset' OR late-onset OR 'late life' OR late-life;

iii) predict* OR moderat* OR regress* OR modulat*.

The following online bibliographic databases were searched on 03/03/2021 and included all years: Web of Science, PsychINFO, CINAHL, Embase, Pubmed. The search was repeated on 12/05/2022 to identify any further relevant papers. References of published reviews and studies which met the eligibility criteria were also manually searched. Following de-duplication, abstracts and titles were initially screened to determine which studies potentially met the inclusion criteria. Full-text screening of potentially relevant studies followed whereby eligibility against the inclusion criteria was assessed. Data from eligible papers were extracted using a structured, standardised form. The screening of titles and abstracts, full-text screening, and data extraction were independently and blindly conducted by two researchers (TK and JH). Any discrepancies were reviewed by a third researcher (RG) and resolved through discussion.

Assessment of study quality

The Cochrane Risk of Bias tool was used to assess the study quality of RCTs (Eldridge et al. 2016). It examined the risk of bias in five domains: selection bias, reporting bias, performance bias, detection bias, and attrition bias. Each domain consisted of different subdomains that were rated with high, low, or unclear risk of bias. The Newcastle-Ottawa Scale (NOS) for case-control studies (Wells et al., 2000) was used to assess the quality of case-control studies. The NOS for cohort studies (Wells et al., 2000) and cross-sectional studies (Alshabanat et al., 2015) were planned to be used to assess the study quality of other types of observational studies but no such studies were identified. The Checklist for the Appraisal of Moderators and Predictors (CHAMP) was used to critically appraise moderator and predictor analyses (Van Hoorn et al., 2017). Quality assessments were carried out independently and blindly by two authors (TK and JH). Any discrepancies were resolved through discussion with a third rater (RG).

Data Extraction

Data extraction was performed using a standardised form for evidence synthesis. Study characteristics were extracted including study, year, country, study design, outcome, setting, diagnostic criteria for anxiety disorder, type of anxiety disorder, treatment condition, comparator/control condition(s), old age sample size, length of intervention, type of analysis performed, and source RCT. The following demographic and clinical characteristics were extracted: mean age, mean years of education, % women, % white ethnicity, screening tool and mean cognitive score, anxiety rating scale name, anxiety rating scale criteria for inclusion, mean anxiety severity at baseline, mean age of onset, and duration of illness at baseline (in years). If studies of secondary data analyses provided insufficient information, studies reporting the primary data analyses were also retrieved for data extraction. Data extraction was carried out independently and blindly by two raters (TK and JH). Any discrepancies were resolved through discussion with a third rater (RG).

Data Synthesis

Data analysis was informed through a narrative synthesis approach, which permits: i) the development of a theory of how the intervention works, why and for whom; ii) the development of the preliminary synthesis of findings of included studies; iii) the exploration of relationships in the data; and iv) the assessment of the robustness of synthesis (Popay et al., 2006). The synthesis was structured around study characteristics, demographic and clinical characteristics, study quality, and predictors and moderators of treatment outcome identified within included studies. In line with the methodology employed in Tunvirachaisakul et al.'s (2018) meta-analysis on predictors of treatment outcome in late-life depression, separate meta-analyses were planned for predictor and/or moderator variables that showed statistical significance (at an alpha level of $p<0.05$) across three or more independent studies. Baseline worry severity was the only characteristic that met this criterion However, a meta-analysis could not be performed due to heterogeneity in the analysis methods employed by the relevant studies, which resulted in effect sizes that could not be meaningfully combined to obtain a reliable pooled estimate (see the results for more discussion of this).

Results

Study Selection

As shown in the PRISMA flow diagram (Figure 1), searches of the online databases generated 8990 records. Following de-duplication, the abstracts and titles of 3911 records were screened which resulted in full-text studies being retrieved for 127 records. Following full-text screening, 114 records were excluded. The most common reasons for exclusion included no analysis of predictors and moderators, not being focused on anxiety disorders, and review studies. This left thirteen studies that were eligible for inclusion in the review. In two different instances, two studies reported findings from the same dataset: Gulpers et al. (2020) and Hendriks et al. (2011), and Caudle et al. (2007) and Wetherell et al. (2005).

Study characteristics

Study characteristics were similar across eligible studies (Table 1). Of the thirteen studies included, nine were conducted in the US and three were conducted in the Netherlands. The studies were published from 2005 to 2020 and participants were recruited either from a practice setting (5/13 studies), community setting (5/13) or both (3/13). Twelve studies adopted an RCT design, and one study adopted a case-control design. All studies used DSM diagnostic criteria to diagnose the anxiety disorder. Most studies focused on generalised anxiety disorder (8/13), two studies focused on panic disorder with agoraphobia, and three studies focused on more than one type of anxiety disorder. The following treatment conditions were employed: eleven studies examined CBT, one study examined escitalopram, and one study examined buspirone. Of the twelve studies reporting RCTs, five studies included a non-active control (e.g., usual care, waiting list), one study included an active control (e.g., drug placebo), two studies included a treatment comparator (e.g., pharmacotherapy vs. psychological therapy), and four studies reported more than one comparator and/or control condition. The sample size was less than 50 participants in six studies, between 50-100 participants in four studies, and between 101-155 participants in three studies. Most studies (9/13) had an intervention lasting between

12-15 weeks, with two studies offering an intervention of 24 weeks. Nine studies reported secondary data analyses of previously published RCTs.

Demographic and clinical characteristics

Demographic and clinical characteristics varied across eligible studies (Table 2). Twelve out of 13 studies included participants with a mean age between 63 to 72 years old, with one study not reporting the mean age (Majercsik et al., 2004). Seven studies reported the mean years of education, which ranged from 13.5 to 16.5 years. All studies had more female participants than male participants, except for Hundt et al. (2014) who had an equal male-to-female ratio. In eight of the nine studies reporting data on ethnicity, the percentage of white ethnicity was above 70%. Twelve studies screened for dementia. Ten studies used more than one anxiety scale, with the Hamilton Anxiety Rating Scale (HAM-A; Thompson et al., 2015) and the Penn State Worry Questionnaire (PSWQ; Meyer et al., 1990) being the most frequently employed scales (in 6 and 8 studies, respectively). In most studies, participants had either mild to moderate anxiety severity at baseline as measured by HAM-A (6/13) or moderate worry severity as measured by PSWQ (3/13). Only one study reported the mean age of onset (Wetherell et al., 2005), and only four studies reported the duration of illness, which ranged between 6 to 32 years.

Quality assessment

The Cochrane Risk of Bias tool was utilised to examine the risk of bias within the twelve RCTs identified (Table 3). Five studies attained a low risk of bias in random sequence generation. Ten studies attained an unclear risk of bias for allocation concealment prior to assignment. All studies had either a high (8/12) or unclear risk of bias (4/12) for the blinding of participants and personnel. Five studies adequately blinded outcome assessors to allocated interventions. Ten studies handled incomplete outcome data adequately and eight studies reported all outcomes assessed.

The NOS for Case-Control studies was used to assess the quality of Mohlman et al. (2013; Table 4). The study provided an adequate description of the cases and of the controls, ensured

representativeness of the cases and the same method used for the ascertainment of cases and controls, and described the non-respondents.

The Checklist for the Appraisal of Moderators and Predictors tool assessed the quality of predictor and moderator analyses examined (Table 5). Eleven studies measured all predictors and/or moderators before the allocation or the start of the intervention and ten studies used reliable and validated assessment tools to assess predictors and/or moderators. Twelve studies presented results for all candidate predictors and/or moderators and all studies specified the moderators and the predictors *a priori*. No study reported power analyses or sample size calculations to ensure adequate sample sizes for the analyses.

Predictors and moderators of treatment outcome

Summaries of the predictors and moderators of treatment outcomes at post-treatment and at follow-up (with respect to treatment response, remission, and/or change in anxiety scores) are reported in Table 6. A total of 49 predictors and/or moderators were explored at post-treatment, 23 of which were statistically significant in at least one study. Thirty-two predictors and/or moderators were reported by one study, 12 of which were statistically significant. Seven predictors and/or moderators were reported by two studies, of which three were statistically significant in at least one study. Ten predictors and/or moderators were reported by three or more studies, seven of which were statistically significant in at least one study.

At follow-up, a total of thirty-three predictors and/or moderators were explored, with 14 being statistically significant in at least one study. Twenty-four predictors and/or moderators were reported by one study, seven of which were statistically significant. Seven predictors and/or moderators were reported by two studies, of which three were statistically significant in at least one study. Two predictors were reported by three or more studies, and both predictors were statistically significant in at least one study. The main results of all the included studies are shown in Supplementary Material 2.

Statistically significant predictors of treatment outcome assessed at post-treatment and at follow-up.

Age. Out of five studies assessing age as a predictor of treatment outcome, only one study showed that age significantly predicted changes in scores on worry measures, whereby younger age predicted higher scores on PSWQ-A (Conti et al., 2017).

Race/Ethnicity. Conti et al. (2017) revealed that African American race predicted changes in worry outcome measures, i.e., higher scores on PSWQ-A and GAI-SF when treated with CBT.

Baseline anxiety. (i) Baseline worry severity, GAD, and state-anxiety severity: Four studies showed that baseline worry severity measured by PSWQ was a significant predictor of treatment outcome (Bradford et al., 2011; Hundt et al., 2014; Mohlman et al., 2013; Mohlman et al., 2020). All four studies recruited participants with GAD (with Ayers et al. also recruiting two participants with an anxiety disorder not otherwise specified [ADNOS]), had CBT as a treatment condition, and usual care, waiting list, or enhanced usual care as control conditions. All four studies showed that less severe baseline worry predicted better response to CBT at post-treatment. The same association was also present at follow-up in one of these four studies (Mohlman et al., 2020). (ii) Baseline anxiety severity: Higher baseline anxiety severity predicted poorer CBT outcome at post-treatment and at follow-up (Ayers et al. 2010). (iii) Baseline GAD severity: Caudle et al. (2007) revealed that higher baseline GAD severity predicted better response to CBT at post-treatment, which was also observed at follow-up (Caudle et al., 2007; Wetherell et al., 2005). (iv) State anxiety: Higher baseline state anxiety was a predictor of poorer CBT outcomes at post-treatment (Hundt et al., 2014).

PSWQ 1-month score change. Reduction in PSWQ scores after the first month predicted lower worry severity at post-treatment and follow-up following CBT (Bradford et al. 2011).

Psychiatric comorbidity. In participants receiving CBT, it was found that the higher the number of comorbid psychiatric diagnoses, the lower the worry severity at post-treatment (Mohlman et al. 2020) and at follow-up (Caudle et al., 2007; Mohlman et al. 2020; Wetherell et al., 2005).

Disease score. Majercsik et al. (2004) demonstrated poor health status (i.e., high number of diseases experienced) predicted worse treatment outcome measured by HAM-A following buspirone treatment.

Therapy. (i) Number of sessions completed: Hundt et al. (2004) showed that a greater number of CBT sessions completed predicted greater improvement in PSWQ-A. (ii) Homework completion: Wetherell et al. (2005) and Caudle et al. (2007) showed that greater homework completion predicted better treatment response to CBT at post-treatment and at follow-up.

Cognitive factors. (i) Executive skills: Mohlman et al. (2013) showed that people with improved executive skills during CBT had a greater reduction in PSWQ scores. (ii) Information processing bias: Biased processing of affective material (i.e., faster response to positive words at 0, 2, 4, 8, and 12 weeks) predicted greater reduction in GAD symptoms post-treatment following escitalopram treatment (Steiner et al., 2013). (iii) VPAR scores: Higher VPAR scores were associated with lower endpoint PSWQ scores at follow-up following CBT treatment (Mohlman et al. 2020). (iv) Orientation: Caudle et al. (2007) found that the absence of errors on the orientation domain of MMSE at baseline was associated with greater improvement in the mean reliable change index (RCI) consisting of three anxiety outcomes (PSWQ, HAM-A, and GAD Severity) at follow-up compared to the presence of at least one error in this domain following CBT.

Psychotropic medication. Psychotropic medication use was associated with higher anxiety severity at follow-up following CBT (Ayers et al., 2010).

Social support. Majercsik et al. (2004) showed that a higher number of social contacts led to a significantly larger improvement in HAM-A scores at post-treatment following buspirone treatment. In the placebo group, the improvement did not depend on the number of social contacts.

Negative life events. Participants who reported experiencing more negative life events at post-treatment predicted higher anxiety severity at follow-up following CBT (Ayers et al., 2010).

Recruitment site. Recruitment site predicted treatment outcome in one study. That is, participants recruited from the Baylor community clinics who received CBT had lower PSWQ and STAI-T scores post-treatment compared to participants recruited from the Veteran Affairs clinic (Hundt et al., 2014).

Source study. Caudle et al. (2007) pooled the results from three different studies and showed that enrolment in the study sample of Stanley et al. (1996) predicted improvement in the mean RCI at

post-treatment and at follow-up following CBT. This association was also observed at follow-up in Wetherell et al. (2005).

Avoidant coping style. Participants engaging in an avoidant coping style, which was reported at post-treatment, was associated with higher anxiety levels at follow-up following CBT (Ayers et al., 2010).

3.7 Statistically significant moderators of treatment outcome at post-treatment and at follow-up.

Baseline depression severity. Schuurmans et al. (2009) found that higher comorbid depression measured by CES-D predicted worse treatment outcome at post-treatment following CBT treatment. No such association was found following sertraline treatment.

Duration of illness and late-onset panic disorder. In older adults with late-life panic disorder with agoraphobia, duration of illness and late-onset panic disorder measured at baseline were found to moderate post-treatment outcome, such that those with higher age at onset and shorter duration of illness benefited more from CBT than paroxetine with respect to both agoraphobic cognitions and avoidance behaviour (Hendriks et al., 2012).

Personality pathology. With respect to treatment of avoidance behaviours in panic disorder with agoraphobia, a higher number of cluster A features and a higher number of features of overall personality pathology at baseline were significantly associated with a worse treatment outcome in those treated with paroxetine, and a significantly better treatment outcome in those treated with CBT (Gulpers et al., 2020).

Neuroticism and perceived health. Neuroticism and lower perceived health measured at baseline were associated with higher anxiety severity at post-treatment and at follow-up following CBT treatment. No such association was identified following sertraline treatment (Schuurmans et al., 2009).

Meta-analysis

Baseline worry severity was the sole characteristic to demonstrate statistical significance across three or more independent studies. However, only Bradford et al. (2011) reported effect sizes (i.e., odds ratios) controlling for treatment arm i.e., the active and control group. Hundt et al. (2014) exclusively

focused their analyses on those randomly assigned to receive CBT in the RCT, whilst Mohlman et al. (2013) focused their analyses on participants who underwent CBT in their case-control study. It is important to recognise that any odds ratios derived from these latter analyses offer specific information applicable to the respective active or case groups investigated but are not directly comparable to odds ratios resulting from a comparison between active and control groups. Consequently, a meta-analysis was not performed.

Discussion

Main findings

In the current review, predictors, and moderators of treatment outcomes for older adults receiving pharmacological and/or psychological interventions for late-life anxiety were examined. Out of the 49 predictors and/or moderators assessed at post-treatment, 23 were statistically significant, and 14 out of 33 predictors and/or moderators assessed at follow-up were also significant. However, caution should be exercised in interpreting most of these predictors and/or moderators, as they were consistently reported as statistically significant by only one or two studies. Baseline worry severity assessed at post-treatment emerged as the only predictor consistently reported as statistically significant in at least three studies.

Baseline anxiety symptoms

Several predictors related to baseline anxiety symptoms were found to be significant, including worry severity, state-anxiety severity, anxiety severity, and GAD severity. Lower baseline levels of worry severity, state-anxiety severity, and anxiety severity in older adults with GAD consistently predicted better CBT outcomes at both post-treatment (Bradford et al., 2010; Hundt et al.,2014; Mohlman et al., 2013; Mohlman et al., 2020) and at follow-up (Ayers et al., 2010; Mohlman et al., 2020). The negative relationship between baseline anxiety symptoms and treatment outcome was also reported in studies with working-age adults with social anxiety disorder (Mululo et al., 2012) and panic disorder (Porter & Chambless, 2015). A possible explanation for this association is that older adults with high levels of worry and anxiety may engage in maladaptive thinking patterns,

which can decrease cognitive flexibility, the ability to adapt and learn new ways of handling situations (Johnco et al., 2015). Conversely, lower anxiety symptom severity at baseline, coupled with higher cognitive flexibility in older adults, may facilitate better engagement in CBT and result in improved outcomes.

Moreover, a previous meta-analysis on late-life depression found that higher baseline anxiety symptoms predicted poorer treatment outcomes in pharmacological treatments and repetitive transcranial magnetic stimulation (rTMS) and electroconvulsive therapy (ECT) trials (Tunvirachaisakul et al., 2018). The authors speculated that comorbid anxiety symptoms may contribute to more severe brain pathology or genetic vulnerabilities that reduce the effectiveness of these treatments (Goodkind et al., 2016; Pimontel et al., 2012). However, such conclusions cannot be drawn from the present review, as it focused solely on psychological treatments.

However, it is important to note that two studies exploring baseline GAD severity reported opposite findings, where higher baseline GAD severity predicted better CBT outcome (Caudle et al., 2007; Wetherell et al., 2005). As the results of these studies were pooled from the same three randomised RCTs, the agreement between them was expected. One possible explanation is that higher anxiety severity at baseline may motivate older adults to seek and engage in treatment, resulting in better outcomes (Newman et al., 2010).

The variation in findings among these studies could also be explained by the methodology employed to measure anxiety. Specifically, Caudle et al. and Wetherell et al. calculated a mean reliable change index (RCI) across three anxiety outcome measures (PSWQ, HAM-A, and GAD Severity rating) to measure improvement. This approach considered multiple dimensions of anxiety and thus provided a comprehensive assessment of anxiety severity (Newman & Fisher, 2010). In contrast, the studies reporting that lower baseline severity was associated with better CBT outcomes utilised a single measure, such as the PSWQ, HAM-A, or STAI-T. These scales are designed to evaluate specific facets of anxiety, such as excessive worry or general anxiety symptoms (Julian, 2011). Whilst this approach allows for a focused assessment of targeted anxiety constructs, it may provide a less comprehensive assessment of anxiety severity. Therefore, the findings from studies

employing a single outcome measure may fail to account for the broader or more diverse spectrum of anxiety symptoms experienced by individuals. Further investigation of the relationship between different dimensions of anxiety and treatment outcome will contribute to a more comprehensive understanding of late-life anxiety and its varied symptomatology.

The inclusion of nine different anxiety outcome measures in the reviewed studies also highlights the heterogeneous nature of anxiety disorders and the diverse cognitive and behavioural symptoms associated with them. Each measure focuses on specific constructs related to anxiety. For instance, STAI-T assesses an individual's overall or dispositional level of anxiety, whilst PSWQ specifically targets excessive worry, a key feature of GAD. The overlap observed among some predictors suggests shared characteristics among certain dimensions of anxiety. However, it is also evident that anxiety can manifest idiosyncratically, leading to variations in treatment outcomes across different measures, as discussed earlier. Whilst individuals may experience improvements in certain aspects of their anxiety symptoms, their progress may not be consistent across all dimensions. Furthermore, variation in the mode of assessment employed in anxiety outcome measures introduces another layer of complexity to the findings. Some measures are administered by clinicians in a structured clinical setting (e.g., HAM-A), whilst others rely on self-ratings (e.g., PSWQ). This difference in assessment method may have influenced treatment outcomes as clinician-rated outcome measures may result in larger effect sizes in comparison to self-rated measures (e.g., Cuijpers et al., 2010; Gould et al., 2012; Pinquart et al., 2006). This emphasises the need to evaluate the consistency of findings across different outcome measures.

Duration of illness is another factor that could have influenced the association between baseline anxiety and treatment outcomes. Although only four of the reviewed studies reported on this factor, and only one of these found it to be a significant predictor in panic disorder with agoraphobia (Hendriks et al., 2012), it still highlights an important consideration. Hendriks et al. suggested that a shorter duration of panic disorder was associated with better outcomes in CBT. Similar findings have been observed in studies involving younger adults with panic disorder (Nakano et al., 2008) and late-life depression (Tunvirachaisakul et al., 2018). Whilst there may be multiple factors at play,

prolonged illness duration may have cumulative effects, resulting in increased cognitive, emotional, or behavioural impairment. These cumulative effects may contribute to higher baseline symptom severity and treatment resistance (Satre et al., 2006). Other factors such as deep-seated views of self, prior negative experiences with treatments, and established unhelpful coping strategies may additionally contribute to treatment resistance and impact on treatment outcomes (Lawrence et al., 2019). However, further research is necessary to investigate the relationship between illness duration, baseline anxiety, and subsequent treatment outcomes to understand the interplay between these factors.

Overall, the findings of the review highlight the significance of baseline anxiety severity as a predictor of treatment outcome in late-life anxiety disorders. It may also have a potential transdiagnostic predictive ability given that baseline anxiety severity also predicted treatment outcomes in pharmacological, rTMs and ECT studies in late-life depression (Tunvirachaisakul et al., 2018). However, its predictive ability following pharmacotherapy remains unknown, as none of the studies used comparator treatments. The heterogeneous nature of anxiety disorders, the varying cognitive and behavioural symptoms, and the use of different assessment measures also underscore the complexity of anxiety as a multidimensional construct. Additionally, factors such as illness duration should be considered to better understand their impact on treatment response. Future studies should further explore the relationship between baseline anxiety and worry severity and outcomes in pharmacotherapies and psychotherapies for anxiety disorders.

Research Implications

This review included 13 studies and identified only one statistically significant predictor that was reported in at least three studies. In contrast, Tunvirachaisakul et al.'s (2018) systematic review of predictors of treatment outcome in late-life depression included 67 studies and identified seven statistically significant predictors that were reported in three or more studies. This highlights the paucity of research in late-life anxiety disorders and hence the pressing need to conduct more research in this area.

Moreover, there was variability in the reported data, making it hard to synthesise findings. To address this issue, authors should prioritise the reporting of effect sizes and confidence intervals which can enable interpretations based on the clinical significance of the findings. In this manner, findings of non-significant p-values with small to moderate effect sizes with the potential to be clinically relevant will be captured.

Clinical implications

Identifying predictors and moderators of treatment outcome in late-life anxiety may help clinicians in several ways: i) it may lead to more informed or faster treatment decisions, e.g., providing the necessary treatment modifications or augmentations to achieve more significant symptom reduction or improved quality of life; and ii) it may provide patients with more information regarding their potential prognosis. Baseline worry/anxiety severity was the only statistically significant predictor reported in at least three studies, whereby lower worry/anxiety severity at baseline predicted better CBT outcomes at post-treatment and follow-up in older people with GAD. Whilst this may suggest that CBT should be offered to those with milder anxiety symptoms at baseline, more research is needed in order to ascertain what type of psychotherapy and/or pharmacotherapy should be offered to those with more severe symptoms at baseline.

Strengths and limitations

The main strength of the current review is that the CHAMP tool showed that most studies (11/13) used reliable and validated assessment tools to assess predictors or moderators. Therefore, observed moderator or predictor effects were less likely to be underestimated or overestimated (Van Hoorn et al., 2017). Moreover, all studies specified the moderators and the predictors *a priori*, which minimises the likelihood of false-positive findings (Van Hoorn et al., 2017).

However, there were a number of limitations of this review. First, the predominance of white, female, and highly educated individuals within the samples limits the generalisability of the findings to broader populations. Additionally, the homogeneity of the samples hinders the exploration of demographic factors such as ethnicity, gender, education level, or socioeconomic status as potential

predictors or moderators of treatment outcome in the existing literature. A study by Conti et al. (2017) made efforts to recruit older adults from ethnic minorities and found that race/ethnicity was a predictor of treatment outcome. Specifically, African American participants showed less improvement in anxiety symptoms compared to their White and Hispanic counterparts. This highlights the importance of studies recruiting larger and more diverse samples in order to ensure sufficient power to detect potential differences and better understand the influence of demographic factors on treatment outcomes. Second, most studies (8/13) reported predictors and/or moderators of outcome of psychological therapies. Thus, the generalisability of these findings to pharmacotherapies is limited. Third, only six out of 13 studies had an active comparator condition, which limits the assessment of predictors and/or moderators of improvement across different types of interventions. Fourth, the use of nine different anxiety measurement scales across studies contributed to heterogeneity in the assessment of anxiety symptoms, as discussed earlier. Fifth, the generalisability of findings to other anxiety disorders is limited given that most studies examined GAD only (8/13), with only two studies examining panic disorder with agoraphobia. Sixth, the current review followed Tunvirachaisakul et al.'s (2018) methodology for identifying consistent predictors and/or moderators in order to permit a comparison of results between older people with late-life depression and older people with late-life anxiety. However, the reliance on statistical significance rather than effect sizes to define consistent predictors and/or moderators may have resulted in some predictors and/or moderators not being identified as consistent across three or more studies. Finally, no study reported conducting a power analysis to ensure that sample sizes were adequate for predictor and moderator analyses based on the CHAMP tool. Indeed, ten out of 13 studies were secondary analyses, which can lack statistical power and inflate the chances of a type 1 error following multiple testing (Andrade, 2019). Of note, approximately half of the studies (6/13) tested numerous moderators and predictors, with the majority of these (5/13) reporting sample sizes smaller than 50 (a previously recommended sample size ‘rule of thumb’ for data analyses of this nature; Pincus et al., 2011; VanVoorhis et al., 2007). Consequently, limited power may be a potential explanation for some of the non-significant results reported here.

Conclusions

In summary, to the authors' knowledge, this is the first systematic review to investigate predictors and moderators of treatment outcome in late-life anxiety. Although 23 statistically significant predictors and/or moderators at post-treatment and 14 at follow-up were identified study, baseline worry severity assessed at post-treatment was the only predictor identified by three or more studies. Consequently, the strength of conclusions that can be drawn and the clinical and policy implications of these findings are limited. Furthermore, these conclusions are based on relatively homogenous samples and from secondary analyses of RCT data, which questions the generalisability and validity of these results. Future researchers should aim to replicate these findings in larger and more diverse samples and design methodologically-strong and adequately-powered studies with the main objective of assessing predictors or moderators. Identifying predictors and moderators of treatment outcome in late-life anxiety disorders may lead to improved understanding of this condition and more informed treatment decisions.

References

Alshabanat, A., Zafari, Z., Albanyan, O., Dairi, M., FitzGerald, J.M., 2015. Asthma and COPD overlap syndrome (ACOS): A systematic review and meta analysis. PLoS One 10. https://doi.org/10.1371/JOURNAL.PONE.0136065.

American Psychiatric Association Division of Research., 2013. Highlights of Changes from DSM-IV to DSM-5: Posttraumatic Stress Disorder. FOCUS. 11, 358–361. https://doi.org/10.1176/APPI.FOCUS.11.3.358.

American Psychiatric Association., 2009. Practice guidelines for the treatment of patients with panic disorder, second ed. Washington, DC.

Andrade, C., 2019. Multiple testing and protection against a type 1 (false positive) error using the Bonferroni and Hochberg corrections. Indian journal of psychological medicine. 41, 99–100. https://doi.org/10.4103/IJPSYM.IJPSYM_499_18.

Ayers, C.R., Petkus, A., Liu, L., Patterson, T.L., Wetherell, J.L., Wetherell, J., 2010. Negative life events and avoidant coping are associated with poorer long-term outcome in older adults treated for generalized anxiety disorder. Journal of Experimental Psychopathology. 1, 146–154. https://doi.org/10.5127/jep.003110.

Bower, E. S., Wetherell, J. L., Mon, T., & Lenze, E. J., 2015. Treating anxiety disorders in older adults: current treatments and future directions. Harvard review of psychiatry. 23, 329-342. https://doi.org/10.1097/HRP.0000000000000064.

Bradford, A., Cully, J., Rhoades, H., Kuenik, M., Kraus-Schuman, C., Wilson, N., & Stanley, M., 2011. Early response to psychotherapy and long-term change in worry symptoms in older adults with generalized anxiety disorder. The American Journal of Geriatric Psychiatry. 19, 347-356. https://doi.org/10.1097/JGP.0b013e3181f18061.

Caudle, D. D., Senior, A. C., Wetherell, J. L., Rhoades, H. M., Beck, J. G., Kunik, M. E., Snow, A. L., Wilson, N. L., & Stanley, M. A., 2007. Cognitive errors, symptom severity, and response to cognitive behavior therapy in older adults with generalized anxiety disorder. The American Journal of Geriatric Psychiatry. 15, 680-689. https://doi.org/10.1097/JGP.0b013e31803c550d.

Conti, E.C., Barrera, T.L., Amspoker, A.B., Wilson, N., Shrestha, S., Wagener, P., Armento, M., Evans-Hudnall, G., Cummings, J.P., Kunik, M.E., Stanley, M.A., 2017. Predictors of outcomes for older adults participating in Calmer Life, a culturally tailored intervention for anxiety. Clinical Gerontologist. 40, 172–180. https://doi.org/10.1080/07317115.2017.1291547.
Cuijpers, P., Li, J., Hofmann, S. G., & Andersson, G., 2010. Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: a meta-analysis. Clinical psychology review. 30, 768-778. https://doi.org/10.1016/j.cpr.2010.06.001.
Department of Health and Social Care, 2001. National Service Framework for Older People. https://www.gov.uk/government/publications/quality-standards-for-care-services-for-older-people (accessed 29 January 2023).
Eldridge, S., Campbell, M., Campbell, M., Drahota-Towns, A., Giraudeau, B., Higgins, J., Reeves, B., & Siegfried, N.., 2016. Revised Cochrane risk of bias tool for randomized trials (RoB 2.0): additional considerations for cluster-randomized trials. https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool.
Goodkind, M.S., Gallagher-Thompson, D., Thompson, L.W., Kesler, S.R., Anker, L., Flournoy, J., Berman, M.P., Holland, J.M., & O'hara, R.M., 2016. The impact of executive function on response to cognitive behavioral therapy in late-life depression. International journal of geriatric psychiatry. 31, 334–339. https://doi.org/10.1002/gps.4325.
Gould, R.L., Coulson, M.C. & Howard, R.J., 2012. Cognitive Behavioral Therapy for depression in older people: A meta-analysis and meta-regression of randomized controlled trials. Journal of American Geriatrics Society. 60, 1817-1830. https://doi.org/10.1111/j.1532-5415.2012.04166.x.
Grubbs, K. M., Cheney, A. M., Fortney, J. C., Edlund, C., Han, X., Dubbert, P., Sherbourne, C. D., Craske, M. G., Stein, M. B., Roy-Byrne, P. P., & Sullivan, J. G., 2015. The role of gender in moderating treatment outcome in collaborative care for anxiety. Psychiatric Services. 66, 265-271. https://doi.org/10.1176/appi.ps.201400049.
Gulpers, B., Voshaar, R. O., Kampman, M., Verhey, F., Van Alphen, S., & Hendriks., 2020. The impact of personality pathology on treatment outcome in late-life panic disorder. Journal of Psychiatric Practice. 26, 164-174. https://doi.org/10.1097/PRA.0000000000000472.
Hendriks, G. J., Keijsers, G. P. J., Kampman, M., Oude Voshaar, R. C., Verbraak, M. J. P. M., Broekman, T. G., & Hoogduin, C. A. L., 2010. A randomized controlled study of paroxetine and cognitive-behavioural therapy for late-life panic disorder. Acta Psychiatrica Scandinavica. 122, 11–19. https://doi.org/10.1111/j.1600-0447.2009.01517.x.
Hendriks, G.-J., Keijsers, G.P.J., Kampman, M., Hoogduin, C.A.L., & Oude Voshaar, R.C., 2011. Predictors of outcome of pharmacological and psychological treatment of late-life panic disorder with agoraphobia. International journal of geriatric psychiatry. 27, 146–150. https://doi.org/10.1002/gps.270.
Hult, K.J., 2014. Measuring the potential health impact of personalized medicine: evidence from multiple sclerosis treatments, in: Berndt, E.S., Goldman, D.P., Rowe, J.W. (Eds), Economic Dimensions of Personalized and Precision Medicine. University of Chicago Press, pp. 185 – 216.
Hundt, N. E., Amspoker, A. B., Kraus-Schuman, C., Cully, J. A., Rhoades, H., Kunik, M. E., & Stanley, M. A., 2014. Predictors of CBT outcome in older adults with GAD. Journal of anxiety disorders. 28, 845-850. https://doi.org/10.1016/j.janxdis.2014.09.012.
Julian, L. J., 2011. Measures of anxiety. Arthritis care & research, 63, 467-472. https://doi.org/10.1002%2Facr.20561.
Johnco, C., Wuthrich, V. M., & Rapee, R. M., 2015. The impact of late-life anxiety and depression on cognitive flexibility and cognitive restructuring skill acquisition. Depression and anxiety. 32, 754–762. https://doi.org/10.1002/da.22375.
Johnco, C., Wuthrich, V. M., & Rapee, R. M.., 2013. The role of cognitive flexibility in cognitive restructuring skill acquisition among older adults. Journal of Anxiety disorders, 27, 576-584. https://doi.org/10.1016/j.janxdis.2012.10.004.
Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S., 2002. Mediators and moderators of treatment effects in randomized clinical trials. Archives of general psychiatry. 59, 877–883. https://doi.org/10.1001/archpsyc.59.10.877.
Lawrence, V., Kimona, K., Howard, R. J., Serfaty, M. A., Wetherell, J. L., Livingston, G., Wilkinson, P., Walters, K., Jones, R., Wuthrich, V. M., & Gould, R. L., 2019. Optimising the acceptability and

feasibility of acceptance and commitment therapy for treatment-resistant generalised anxiety disorder in older adults. Age and ageing. 48, 735-744. https://doi.org/10.1093/ageing/afz082.
Majercsik, E., & Haller, J., 2004. Interactions between anxiety, social support, health status and buspirone efficacy in elderly patients. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 28, 1161-1169. https://doi.org/10.1016/j.pnpbp.2004.06.007.
Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D,. 1990. Development and validation of the penn state worry questionnaire. Behaviour research and therapy. 28, 487-495. https://doi.org/10.1016/0005-7967(90)90135-6.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., Altman, D., Antes, G., Atkins, D., Barbour, V., Barrowman, N., Berlin, J.A., Clark, J., Clarke, M., Cook, D., D'Amico, R., Deeks, J.J., Devereaux, P.J., Dickersin, K., Egger, M., Ernst, E., Gøtzsche, P.C., Grimshaw, J., Guyatt, G., Higgins, J., Ioannidis, J.P.A., Kleijnen, J., Lang, T., Magrini, N., McNamee, D., Moja, L., Mulrow, C., Napoli, M., Oxman, A., Pham, B., Rennie, D., Sampson, M., Schulz, K.F., Shekelle, P.G., Tovey, D., Tugwell, P., 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Ann Intern Med 151, 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135.
Mohlman, J., 2013. Executive skills in older adults with GAD: Relations with clinical variables and CBT outcome. Journal of Anxiety Disorders. 27, 131-139. https://doi.org/10.1016/j.janxdis.2012.12.001.
Mohlman., 2020. Neurocognitive predictors of long-term outcome in CBT for late life generalized anxiety disorder. Journal of Anxiety Disorders. 74, 102246. https://doi.org/10.1016/j.janxdis.2020.102246.
Mululo, S. C. C., Menezes, G. B. D., Vigne, P., & Fontenelle, L. F. 2012. A review on predictors of treatment outcome in social anxiety disorder. Brazilian Journal of Psychiatry. 34, 92-100. https://doi.org/10.1590/S1516-44462012000100016.
Nakano, Y., Lee, K., Noda, Y., Ogawa, S., Kinoshita, Y., Funayama, T., Watanabe, N., Chen, J., Noguchi, Y., Toshiaki, A., & Furukawa, T. A. 2008. Cognitive-behavior therapy for Japanese patients with panic disorder: Acute phase and one-year follow-up results. Psychiatry and clinical neurosciences, 62, 313-321. https://doi.org/10.1111/j.1440-1819.2008.01799.x.
National Institute for Health and Care Excellence. (2011). Generalised anxiety disorder and panic disorder in adults: management. [Clinical guideline CG113]. https://www.nice.org.uk/guidance/ng43.
National Institute for Health and Care Excellence. (2013). Social anxiety disorder: recognition, assessment and treatment. [Clinical guideline CG159]. https://www.nice.org.uk/guidance/CG159.
Newman, M. G., & Fisher, A. J., 2010. Expectancy/credibility change as a mediator of cognitive behavioral therapy for generalized anxiety disorder: Mechanism of action or proxy for symptom change?. International journal of cognitive therapy. 3, 245-261. https://doi.org/10.1521/ijct.2010.3.3.245.
Newman, M. G., Przeworski, A., Fisher, A. J., & Borkovec, T. D., 2010. Diagnostic comorbidity in adults with generalized anxiety disorder: Impact of comorbidity on psychotherapy outcome and impact of psychotherapy on comorbid diagnoses. Behavior Therapy. 41, 59-72. https://doi.org/10.1016/j.beth.2008.12.005.
Nierenberg, A. A., 2003. Predictors of response to antidepressants: General principles and clinical implications. Psychiatric Clinics. 26, 345-352. https://doi.org/10.1016/S0193-953X(02)00105-3.
Papakostas, G.I., Fava, M., 2008. Predictors, moderators, and mediators (correlates) of treatment outcome in major depressive disorder. Dialogues Clin Neurosci. 10, 439–451. https://doi.org/10.31887/DCNS.2008.10.4/GIPAPAKOSTAS.
Pimontel, M.A., Culang-Reinlieb, M.E., Morimoto, S.S., Sneed, J.R., 2011. Executive dysfunction and treatment response in late-life depression. International journal of geriatric psychiatry. 27, 893–899. https://doi.org/10.1002/gps.2808.
Pincus, T., Miles, C., Froud, R., Underwood, M., Carnes, D., Taylor, S.J., 2011. Methodological criteria for the assessment of moderators in systematic reviews of randomised controlled trials: A consensus study. BMC Med Res Methodol 11, 1-14. https://doi.org/10.1186/1471-2288-11-14.
Pinquart, M., Duberstein, P. R., & Lyness, J. M., 2006. Treatments for later-life depressive conditions: a meta-analytic comparison of pharmacotherapy and psychotherapy. American Journal of Psychiatry. 163, 1493-1501. https://doi.org.10.1176/ajp.2006.163.9.1493.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., Duffy, S., 2006. Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme Version, 1, b92.
Porensky, E. K., Dew, M. A., Karp, J. F., Skidmore, E., Rollman, B. L., Shear, M. K., & Lenze, E. J., 2009. The burden of late-life generalized anxiety disorder: effects on disability, health-related quality of life, and healthcare utilization. The American Journal of Geriatric Psychiatry. 17, 473-482. https://doi.org/10.1097/JGP.0b013e31819b87b2.
Porter, E., & Chambless, D. L., 2015. A systematic review of predictors and moderators of improvement in cognitive-behavioral therapy for panic disorder and agoraphobia. Clinical Psychology Review. 42, 179-192. https://doi.org/10.1016/j.cpr.2015.09.004.
Satre, D. D., Knight, B. G., & David, S., 2006. Cognitive-behavioral interventions with older adults: Integrating clinical and gerontological research. Professional Psychology: research and practice. 37, 489. https://psycnet.apa.org/doi/10.1037/0735-7028.37.5.489.
Schuurmans, J., Comijs, H., Emmelkamp, P. M., Gundy, C. M., Weijnen, I., Van Den Hout, M., & Van Dyck, R,. 2006. A randomized, controlled trial of the effectiveness of cognitive–behavioral therapy and sertraline versus a waitlist control group for anxiety disorders in older adults. The American journal of geriatric psychiatry, 14, 255-263. https://doi.org/10.1097/01.JGP.0000196629.19634.00.
Schuurmans, J., Comijs, H., Emmelkamp, P.M.G., Weijnen, I.J.C., van den Hout, M., van Dyck, R., 2099. Long-term effectiveness and prediction of treatment outcome in cognitive behavioral therapy and sertraline for late-life anxiety disorders. International Psychogeriatrics. 21, 1148-1159. https://doi.org/10.1017/S1041610209990536.
Stanley, M. A., Beck, J. G., & Glassco, J. D., 1996. Treatment of generalized anxiety in older adults: A preliminary comparison of cognitive-behavioral and supportive approaches. Behavior therapy. 27, 565-581. https://doi.org/10.1016/S0005-7894(97)80035-4.
Stanley, M. A., Wilson, N. L., Novy, D. M., Rhoades, H. M., Wagener, P. D., Greisinger, A. J., Cully, J.A., & Kunik, M. E., 2009. Cognitive behavior therapy for generalized anxiety disorder among older adults in primary care: a randomized clinical trial. Jama. 301, 1460-1467. https://doi.org/10.1001/jama.2009.458.
Stanley, M. A., Wilson, N., Shrestha, S., Amspoker, A. B., Armento, M., Cummings, J. P., Evans-Hudnall, G., Wagener, P. D., & Kunik, M. E., 2016. Calmer life: a culturally tailored intervention for anxiety in underserved older adults. The American Journal of Geriatric Psychiatry. 24, 648-658. https://doi.org/10.1016/j.jagp.2016.03.008.
Stanley, M.A., Beck, J.G., Novy, D.M., Averill, P.M., Swann, A.C., Diefenbach, G.J., Hopko, D.R., 2003. Cognitive-behavioral treatment of late-life generalized anxiety disorder. J Consult Clin Psychol. 71, 309–319. https://doi.org/10.1037/0022-006X.71.2.309.
Stanley, M.A., Wilson, N.L.W., Amspoker, A.B., Kraus-Schuman, C., Wagener, P.D., Calleo, J.S., Cully, J.A., Teng, E., Rhoades, H.M., Williams, S., Masozera, N., Horsfield, M., Kunik, M.E., 2014. Lay providers can deliver effective cognitive behavior therapy for older adults with generalized anxiety disorder: a randomized trial. Depression and Anxiety. 31, 391–401. https://doi.org/10.1002/da.22239.
Steiner, A. R., Petkus, A. J., Nguyen, H., & Wetherell, J. L., 2013. Information processing bias and pharmacotherapy outcome in older adults with generalized anxiety disorder. Journal of anxiety disorders. 27, 592-597. https://doi.org/10.1016/j.janxdis.2012.10.007.
Thompson, E., 2015. Hamilton rating scale for anxiety (HAM-A). Occupational Medicine. 65, 601. https://doi.org/10.1093/occmed/kqv054.
Tunvirachaisakul, C., Gould, R. L., Coulson, M. C., Ward, E. V., Reynolds, G., Gathercole, R. L., Grocott, H., Supasitthumrong, T., Tunvirachaisakul, A., Kimona, K., & Howard, R. J., 2018. Predictors of treatment outcome in depression in later life: A systematic review and meta-analysis. Elsevier. https://doi.org/10.1016/j.jad.2017.10.008.
United Nations. Department of Economic and Social Affairs., 2007. World Economic and Social Survey 2007: Development in an ageing world. https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/2007wess.pdf.

van Hoorn, R., Tummers, M., Booth, A., Gerhardus, A., Rehfuess, E., Hind, D., Bossuyt, P.M., Welch, V., Debray, T.P.A., Underwood, M., Cuijpers, P., Kraemer, H., van der Wilt, G.J., Kievit, W., 2017. The development of CHAMP: A checklist for the appraisal of moderators and predictors. BMC Med Res Methodol. 17, 1-9. https://doi.org/10.1186/S12874-017-0451-0.
VanVoorhis, C. W., & Morgan, B. L., 2007. Understanding power and rules of thumb for determining sample sizes. Tutorials in quantitative methods for psychology. 3, 43-50. https://doi.org/ https://doi.org/10.20982/tqmp.03.2.p043.
Wells, G., Shea, B., O'Connell, D., Peterson, J., Welch, V., 2000. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Microsoft PowerPoint - Newcastle Ottowa Scale Pope Bruce.ppt (evidencebasedpublichealth.de).
Wetherell, J. L., Ayers, C. R., Sorrell, J. T., Thorp, S. R., Nuevo, R., Belding, W., Gray, M., Stanley, M. A., Areán, P. A., Donohue, M., Unützer, J., Ramsdell, J., Xu, R., & Patterson, T. L., 2009. Modular psychotherapy for anxiety in older primary care patients. The American Journal of Geriatric Psychiatry. 17, 483-492. https://doi.org/10.1097/JGP.0b013e3181a31fb5.
Wetherell, J. L., Gatz, M., & Craske, M. G., 2003. Treatment of generalized anxiety disorder in older adults. Journal of consulting and clinical psychology. 71, 31. https://doi.org/10.1037/0022-006X.71.1.31.
Wetherell, J. L., Hopko, D. R., Diefenbach, G. J., Averill, P. M., Beck, J. G., Craske, M. G., Gatz, M., Novy, D.M., & Stanley, M. A., 2005. Cognitive-behavioral therapy for late-life generalized anxiety disorder: Who gets better?. Behavior Therapy. 36, 147-156. https://doi.org/10.1016/S0005-7894(05)80063-2.
Wetherell, J.L., Petkus, A.J., White, K.S., Nguyen, H., Kornblith, S., Andreescu, C., Zisook, S., Lenze, E.J., 2013. Antidepressant medication augmented with cognitive-behavioral therapy for generalized anxiety disorder in older adults. American Journal of Psychiatry 170, 782–789. https://doi.org/10.1176/APPI.AJP.2013.12081104.

Author Disclosure

Author contribution

TK and RLG participated in study design and data analysis. TK and JH reviewed, extracted data, and assessed quality of the included studies. All co-authors substantially contributed to manuscript preparation and approved it for submission.

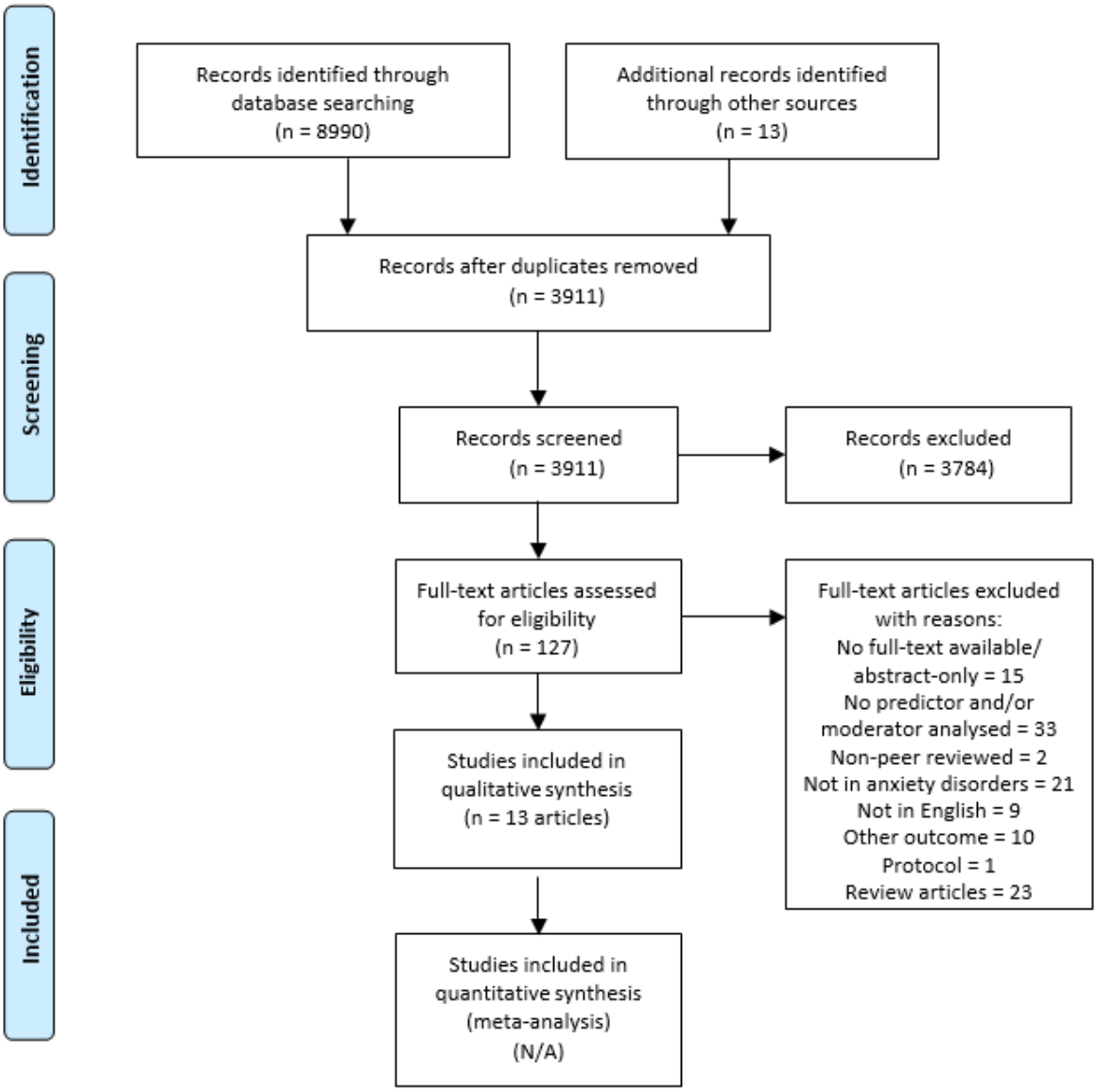
Role of funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest

The authors confirm that there are no conflicts of interest associated with this publication and there is no financial support for this publication that can influence the outcome.

Figure 1: PRISMA Flow Diagram



Adapted from: Moher et al. (2009)

Table 1: Study characteristics

Study year country	Study Design	Treatment Outcome	Setting	Diagnostic criteria used for anxiety disorder	Type of anxiety disorder	Treatment condition (s)	Comparator / control condition(s)	Sample size	Length of intervention (in weeks)	Source RCT
Ayers et al. (2010) US	RCT	SCORE	Practice	DSM-IV	GAD & ADNOS	CBT	Enhanced community treatment	27	12	Wetherell et al. (2009)
Bradford et al. (2011) Netherlands	RCT	RES	Practice	DSM-IV	GAD	CBT	Enhanced usual care	76	12	Stanley et al. (2009)
Caudle et al. (2007) US	RCT *	SCORE	Community	DSM-III-R or DSM-IV	GAD	CBT	Supportive psychotherapy, or discussion group & 12 week waiting group, or minimal contact control	65	12-15	Stanley et al. (1996), Stanley et al. (2003), Wetherell et al. (2003)
Conti et al. (2017) US	RCT **	SCORE	Community	DSM-IV	GAD & ADNOS	CBT	Enhanced community care	54	12	Stanley et al. (2016)
Gulpers et al. (2020) Netherlands	RCT	SCORE	Practice	DSM-IV	Panic disorder with agoraphobia	CBT	Paroxetine	34	14	Hendriks et al. (2010)
Hendriks et al. (2012) Netherlands	RCT	SCORE	Practice	DSM-IV	Panic disorder with agoraphobia	CBT	Paroxetine, Waiting-list group	49	14	Hendriks et al. (2010)
Hundt et al. (2014) US	RCT ***	SCORE	Practice / Community	DSM-IV	GAD	CBT	Usual care	150	24	Stanley et al. (2014)
Majercsik et al. (2004) Hungary	RCT	SCORE	Practice	DSM-IV	GAD	buspirone	Placebo	155	6	N/A
Mohlman et al. (2013) US	Case-control	REM/SCORE	Community	DSM-IV	GAD	CBT	N/A	GAD = 69, Control = 52^	8	N/A
Mohlman et al. (2020) US	RCT	SCORE	Community	DSM-IV	GAD	CBT	Waiting list	44	24	N/A

Study year country	Study Design	Treatment Outcome	Setting	Diagnostic criteria used for anxiety disorder	Type of anxiety disorder	Treatment condition (s)	Comparator / control condition(s)	Sample size	Length of intervention (in weeks)	Source RCT
Schuurmans et al. (2009) US	RCT	SCORE	Practice/ Community	DSM-IV-TR	GAD, panic disorder, agoraphobia, social phobia	CBT	Sertraline, Waiting list	47	15	Schuurmans et al. (2006)
Steiner et al. (2013) US	RCT †	SCORE	Practice/ Community	DSM-IV-TR	GAD	Escitalopram	Escitalopram (followed by placebo or CBT)	25	12	Wetherell et al. (2013)
Wetherell et al. (2005) US	RCT *	SCORE	Community	DSM-III-R or DSM-IV	GAD	CBT	supportive psychotherapy, or discussion group & 12 week waiting group, or minimal contact control	65	12-15	Stanley et al. (1996), Stanley et al. (2003), Wetherell et al. (2003)

Note: N/A = Not applicable, * = Caudle et al. (2007) and Wetherell et al. (2005) pooled findings from the same three datasets, ** = Conti et al. (2017) reported data for participants who received the intervention from either an RCT or an open trial, *** = Hundt et al. (2014) examined only data from patients randomised to CBT (data taken from a published RCT of CBT), † = In Steiner et al (2013) participants were awaiting randomisation into a study evaluating the effects of cognitive-behavioural therapy as an augmentation to pharmacotherapy, ^ = Sample size only reported per condition

Treatment Outcome RES = response, REM = remission, SCORE = score on anxiety questionnaire

Study design RCT = randomised controlled trial

Diagnostic criteria DSM = Diagnostic and Statistical Manual of Mental Disorders

Type of anxiety disorder GAD = Generalised anxiety disorder, ADNOS = anxiety disorder not otherwise specified

Treatment conditions CBT = Cognitive behavioural therapy

Table 2: Demographic and clinical characteristics.

Study year	Mean age (SD)	Mean years of education (SD)	% Women	% White ethnicity	Cognitive screening tool and mean cognitive score	Anxiety Rating scale(s)	Anxiety Rating scale criteria for inclusion	Mean anxiety severity at baseline (SD)	Mean age of onset (SD)	Duration of illness at baseline (in years)
Ayers et al. (2010)	72 (6.9)	16 (2.7)	85.2	96.3	MMSE (N/A)	HAM-A	N/A	HAM-A = 18.52 (5.42)	N/A	N/A
Bradford et al. (2011)	CBT = 66.7 (6.6) EUC = 68.2 (5.5)	CBT = 16.07 (2.7) EUC = 16.00 (3.0)	CBT = 84.4 EUC = 74.2	CBT = 73.3 EUC = 87.1	MMSE (N/A)	PSWQ	N/A	CBT PSWQ = 53.8 (10.1), EUC PSWQ = 58.4 (12.0)	N/A	N/A
Caudle et al. (2007) *	66.7 (6.6)	14.7 (2.4)	78.5	83.1	MMSE (28.7)	HAM-A PSWQ	N/A	HAM-A = 18.6 (6.4) PSWQ = 60.9 (10.3)	N/A	N/A
Conti et al. (2017)	63.4 (7.8)	13.5 (2.1)	96.3	20.4	6-item screen	PSWQ-A GAI-SF	PSWQ-A >22	PSWQ-A = 28.5 (6.8) GAI-SF = 3.4 (1.6)	N/A	N/A
Gulpers et al. (2020)	Paroxetine = 69.0 (4.4) CBT = 69.1 (5.0)	N/A	Paroxetine = 63 CBT = 56	N/A	MMSE (N/A)	Dutch versions of MI-A and ACQ	N/A	Paroxetine MI-A = 2.4 (0.7), ACQ = 1.7 (0.4), CBT MI-A = 2.4 (1.0), ACQ = 1.6 (0.5)	N/A	Paroxetine = 6.0 (17.5) CBT = 6.5 (21.4) **
Hendriks et al. (2012)	69.2 (4.7)	N/A	57	N/A	MMSE (N/A)	Dutch versions of the MI-A and ACQ	N/A	MI-A = 2.4 (0.9) ACQ = 1.7 (0.5)	N/A	13.8 (15.8)
Hundt et al. (2014)	66.79 (6.38)	15.46 (3.11)	50	80	6-item screener (N/A)	PSWQ-A STAI-T	N/A	PSWQ-A = 24.47 (7.95) STAI-T = 46.89 (10.20)	N/A	N/A
Majercsik et al. (2004)	N/A	N/A	71.0	N/A	N/A (N/A)	HAM-A	HAM-A >15	HAM-A = 20.9 (2.3)	N/A	N/A
Mohlman et al. (2013)	GAD = 70.0 (6.7) Control = 70.1 (6.9)	N/A	GAD = 72 Control = 73	GAD = 88 Control = 82	MMSE	PSWQ GAD-Q-IV	N/A	GAD PSWQ = 59.2 (13.8) GAD-Q-IV = 8.5	N/A	N/A

Study year	Mean age (SD)	Mean years of education (SD)	% Women	% White ethnicity	Cognitive screening tool and mean cognitive score	Anxiety Rating scale(s)	Anxiety Rating scale criteria for inclusion	Mean anxiety severity at baseline (SD)	Mean age of onset (SD)	Duration of illness at baseline (in years)
								(2.5), Control PSWQ = 30.5 (7.5) GAD-Q-IV = 1.9 (1.5)		
Mohlman et al. (2020)	70.93 (5.42)	N/A	Waitlist = 59 CBT = 65	Waitlist = 71 CBT = 70	MMSE (121) ***	PSWQ-A STAI-T	N/A	N/A	N/A	N/A
Schuurmans et al. (2009)	Sertraline = 69.35 (5.88) CBT = 70.60 (6.52)	N/A	70	N/A	MMSE (N/A)	HAM-A BAI	N/A	CBT BAI = 19.09 (12.09) HAM-A = 14.48 (8.00), Sertraline BAI = 22.90 (13.19) HAM-A = 17.86 (8.22)	N/A	CBT = 27.85 (25.20) Sertraline = 30.47 (23.52)
Steiner et al. (2013)	68.8 (7.6)	16.4 (2.5)	60	80	MMSE (N/A)	HAM-A PSWQ GADSS	HAM-A >16	HAM-A = 21.63 (3.61) PSWQ = 55.23 (10.95) GADSS = 14.80 (3.75)	N/A	N/A
Wetherell et al. (2005) *	67.8 (6.5)	14.7 (2.4)	69	83	MMSE (28.7)	ADIS-IV PSWQ HAM-A	N/A	ADIS = 5.01 (1.0) HAM-A = 18.6 (6.4) PSWQ = 60.9 (10.3)	36.1 (26.3)	31.7 (26.5)

Note: N/A = Not applicable or not available, EUC = enhanced usual care, * = Caudle et al. (2007) and Wetherell et al. (2005) pooled findings from the same three datasets; there were certain discrepancies in their reported demographic and clinical characteristics, ** = Median reported, *** = MMSE has a maximum score of 30. The mean cognitive score of 121 does not reflect this score range.

Cognitive screening tools MMSE = Mini-Mental State Examination

Anxiety rating scale name ACQ = Agoraphobic Cognitions Questionnaire, ADIS-IV = Anxiety Disorders Interview Schedule for DSM-IV, BAI = Beck Anxiety Inventory, GADSS = Generalized Anxiety Disorder Severity Scale, GAD-Q-IV = Generalized Anxiety Disorder-Questionnaire-IV, GAI-SF = Geriatric Anxiety Inventory Short Form, HAM-A = Hamilton Anxiety Rating Scale, MI-A = Mobility Inventory Avoidance scale, PSWQ = Penn State Worry Questionnaire, PSWQ-A = Penn State Worry Questionnaire – short version, STAI-T = Spielberger State-Trait Anxiety Inventory - trait subscale

Table 3: Risk of bias - Cochrane Risk of Bias Tool

Study year	Random Sequence Generation	Allocation Concealment	Blinding of participants and personnel	Blinding of Outcome Assessments	Incomplete Outcomes	Selective Reporting	Additional details
Ayers et al., 2010	?	?	?	✓	✓	✓	No information reported about random sequence generation and allocation concealment. The experimental and control groups were merged for the analyses; information may had been lost.
Bradford et al. (2011)	✓	✓	X	?	✓	✓	No blinding of participants and personnel to the allocated interventions.
Caudle et al. (2007) *	?	?	X	?	✓	?	No information provided on random sequence generation and allocation concealment. Missing information on blinding participants, personnel and the outcome assessors.
Conti et al. (2017) **	?	?	?	?	?	?	No sufficient information was reported across the different domains.
Gulpers et al. (2020)	✓	?	X	✓	✓	✓	No sufficient information provided for allocation concealment and no blinding of participants and personnel.
Hendriks et al. (2012)	✓	?	X	✓	✓	X	No sufficient information provided for allocation concealment and no blinding of participants and personnel. No findings presented about waiting-list group suggesting selective reporting.
Hundt et al. (2014) ***	✓	?	X	✓	✓	✓	No blinding of participants and personnel.
	?				✓	✓	

Study year	Random Sequence Generation	Allocation Concealment	Blinding of participants and personnel	Blinding of Outcome Assessments	Incomplete Outcomes	Selective Reporting	Additional details
Majercsik et al. (2004)		?	?	?			No information was reported about random sequence generation and allocation concealment. Authors report that 'no special care was taken to ensure complete blindness of experimenters to treatment'.
Mohlman et al. (2020)	?	?	X	✓	✓	✓	Insufficient information regarding random sequence generation and allocation concealment. Personnel and participants were not blinded to the interventions.
Schuurmans et al. (2009)	✓	✓	X	X	?	✓	Study participants, personnel and assessors were not blinded.
Steiner et al. (2013) †	?	?	?	?	✓	✓	Insufficient information regarding random sequence generation, allocation concealment, blinding of participants and personnel, and blinding of outcome assessments.
Wetherell et al. (2005) *	?	?	X	?	✓	?	No information provided on random sequence generation and allocation concealment. Information on blinding participants, personnel and the outcome assessors were missing.

Note: ✓ = low risk of bias, X = high risk of bias, ? = uncertain risk of bias, * = Caudle et al. (2007) and Wetherell et al. (2005) pooled findings from the same three datasets, ** = Conti et al. (2017) reports data for participants who received the intervention from either an RCT or an open trial, *** = In Steiner et al (2013) participants were awaiting randomisation into a study evaluating the effects of cognitive-behavioural therapy as an augmentation to pharmacotherapy, † = Hundt et al. (2014) examine only data from patients randomised to CBT (data taken from a published RCT of CBT)

Table 4: Risk of bias - Newcastle-Ottawa Scale for Case-Control Studies.

Author year	Domains								**Additional details**
	Selection				**Comparability**	**Exposure**			
	Is the case definition adequate?	Representativeness of the cases	Selection of controls	Definition of control	Comparability of cases and controls on the basis of the design or analysis	Ascertainment of exposure	Same method of ascertainment for cases and controls	Non-response rate	
Mohlman et al. (2013)	Yes	Consecutive or obviously representative	Community controls	No history of the disease	Study controls for GAD	Interview not blinded to case/control status	Yes	Non-respondents described	Controls were age- and sex-matched control free of lifetime psychiatric disorders based on the SCID.

Additional details SCID = Structured Clinical Interview for DSM Disorders

Table 5: Checklist for the Appraisal of Moderators and Predictors.

Author year	Domain												**Overall judgment**	Clarification of the main arguments	Body of evidence
	Design				**Analysis**			**Results**			**Transferability**				
	Was there sufficient empirical or theoretical support for the moderator or predictor that was examined?	Was the moderator or predictor specified a priori?	Was the moderator or predictor variable measured before the allocation or start of the intervention?	Was measurement of the moderator or predictor reliable and valid in the target population?	In case of a moderator, was an interaction test used?	Was a limited number of moderators and predictors tested?	Was sample size adequate for the moderator or predictor analysis?	Were results presented for all candidate moderators or predictors that were examined?	Did statistical tests or confidence intervals indicate that observed moderator or predictor effects were unlikely to be merelv due to chance variation?	Was the moderator or predictor effect consistent with related moderators or predictors, or across related outcomes measured within the study?	Were the setting and study population comparable to the setting and population in which the information would be used?	Is the moderator or predictor effect clinically important?	Would the claims regarding moderation or prediction of treatment outcomes be sufficiently substantiated and sufficiently relevant to take into account when making recommendations for treatment decisions?		
Ayers et al. (2010)	Yes	Yes	?	?	No	No	No	Yes	Yes	Yes	No	?	No	No sample size calculation No validation of the life events measure Life events and avoidant coping style measured post-baseline	Preliminary findings for life events and avoidant coping style as a predictor
Bradford et al. (2011)	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	N/A	Yes	?	No	No sample size calculation No confidence intervals for odds ratios Change in PSWQ was not clinically significant	Preliminary findings for 1-month change in PSWQ score as a predictor
Caudle et al. (2007)	Yes	Yes	Yes	Yes	N/A	Yes	DK	Yes	Yes	No	Yes	No	No	No sample size calculation	Evidence on the predictive ability of cognitive functioning deficits remains scarce

Author year	Domain												**Overall judgment**	Clarification of the main arguments	Body of evidence
	Design				**Analysis**			**Results**			**Transferability**				
	Was there sufficient empirical or theoretical support for the moderator or predictor that was examined?	Was the moderator or predictor specified a priori?	Was the moderator or predictor variable measured before the allocation or start of the intervention?	Was measurement of the moderator or predictor reliable and valid in the target population?	In case of a moderator, was an interaction test used?	Was a limited number of moderators and predictors tested?	Was sample size adequate for the moderator or predictor analysis?	Were results presented for all candidate moderators or predictors that were examined?	Did statistical tests or confidence intervals indicate that observed moderator or predictor effects were unlikely	Was the moderator or predictor effect consistent with related moderators or predictors, or across related outcomes measured within the study?	Were the setting and study population comparable to the setting and population in which the information would be used?	Is the moderator or predictor effect clinically important?	Would the claims regarding moderation or prediction of treatment outcomes be sufficiently substantiated and sufficiently relevant to take into account when making recommendations for treatment decisions?		
Conti et al. (2017)	Yes	Yes	Yes	Yes	N/A	Yes	DK	Yes	Yes	?	Yes	No	No	No sample size calculation Six-item screener for cognitive impairment requires further validation	Baseline severity score of anxiety previously identified as a predictor
Gulpers et al. (2020)	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	Yes	No	No	No sample size calculation Unknown whether the Dutch versions of the Personality Diagnostic Questionnaire were validated in older adults Small to medium effect sizes observed	First study on personality pathology in panic disorder with agoraphobia
Hendriks et al. (2012)	No	Yes	N/A	Yes	Yes	No	DK	Yes	Yes	Yes	No	No	No	Small sample size No sample size calculation	Preliminary findings
Hundt et al. (2014)	No	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	No	No	No sample size calculation Sample size appears inadequate for the predictor analysis	Baseline anxiety severity and no of CBT sessions replicated

Author year	Domain														
	Design				**Analysis**			**Results**			**Transferability**		**Overall judgment**	Clarification of the main arguments	Body of evidence
	Was there sufficient empirical or theoretical support for the moderator or predictor that was examined?	Was the moderator or predictor specified a priori?	Was the moderator or predictor variable measured before the allocation or start of the intervention?	Was measurement of the moderator or predictor reliable and valid in the target population?	In case of a moderator, was an interaction test used?	Was a limited number of moderators and predictors tested?	Was sample size adequate for the moderator or predictor analysis?	Were results presented for all candidate moderators or predictors that were examined?	Did statistical tests or confidence intervals indicate that observed moderator or predictor effects were unlikely	Was the moderator or predictor effect consistent with related moderators or predictors, or across related outcomes measured within the study?	Were the setting and study population comparable to the setting and population in which the information would be used?	Is the moderator or predictor effect clinically important?	Would the claims regarding moderation or prediction of treatment outcomes be sufficiently substantiated and sufficiently relevant to take into account when making recommendations for treatment decisions?		
Majercsik et al. (2004)	Yes	Yes	Yes	?	No	Yes	DK	Yes	DK	DK	Yes	No	No	Sufficient theoretical background No sample size calculation present	Findings requiring further replication
Mohlman et al. (2013)	Yes	Yes	Yes	Yes	N/A	DK	DK	Yes	Yes	No	Yes	No	No	No sample size calculation Unknown whether the sample size was adequate for predictor analysis	Preliminary findings on the predictive ability of executive functioning
Mohlman et al. (2020)	Yes	Yes	Yes	Yes	No	No	DK	Yes	Yes	DK	Yes	?	Yes	Small sample size Lack of power to identify predictors	Preliminary findings

Author year	Domain												**Overall judgment**	Clarification of the main arguments	Body of evidence
	Design				**Analysis**			**Results**			**Transferability**				
	Was there sufficient empirical or theoretical support for the moderator or predictor that was examined?	Was the moderator or predictor specified a priori?	Was the moderator or predictor variable measured before the allocation or start of the intervention?	Was measurement of the moderator or predictor reliable and valid in the target population?	In case of a moderator, was an interaction test used?	Was a limited number of moderators and predictors tested?	Was sample size adequate for the moderator or predictor analysis?	Were results presented for all candidate moderators or predictors that were examined?	Did statistical tests or confidence intervals indicate that observed moderator or predictor effects were unlikely	Was the moderator or predictor effect consistent with related moderators or predictors, or across related outcomes measured within the study?	Were the setting and study population comparable to the setting and population in which the information would be used?	Is the moderator or predictor effect clinically important?	Would the claims regarding moderation or prediction of treatment outcomes be sufficiently substantiated and sufficiently relevant to take into account when making recommendations for treatment decisions?		
Schuurmans et al. (2009)	?	Yes	Yes	Yes	No	No	No	Yes	DK	N/A	Yes	No	No	Insufficient theoretical background Recruited fewer participants than planned Large no of predictors tested given the sample size	Preliminary findings
Steiner et al. (2013)	Yes	Yes	Yes	Yes	N/A	Yes	DK	Yes	DK	?	Yes	DK	No	No sample size calculation Findings may be an artifact of the small sample size	Preliminary findings
Wetherell et al. (2005)	Yes	Yes	Yes	Yes	DK	Yes	DK	Yes	Yes	DK	Yes	No	No	Relatively high number of predictors were assessed considering the sample size	Preliminary findings requiring independent replication

Note: ? = partially, N/A = not applicable, DK = do not know

Table 6. Combined results of predictors and moderators assessed at post-treatment and follow-up.

Type of factor	Predictors* / Moderators†	Post-treatment: Response / Remission / Score – Regression model of outcome	Follow-up: Response / Remission / Score – Regression model of outcome
Demographic	Age*	**Conti et al. (2017)$_{b}$ – PSWQ-A** Conti et al. (2017)$_{b}$ – GAI-SF Hendriks et al. (2012)$_{ab}$ – MI-A Hendriks et al. (2012)$_{ab}$ – ACQ Hundt et al. (2014)$_{b}$ – PSWQ Hundt et al. (2014)$_{b}$ -STAI-T Schuurmans et al. (2009)$_{ab}$ – CBT Schuurmans et al. (2009)$_{ab}$ – Sertraline Wetherell et al. (2005)b – Mean RCI	Schuurmans et al. (2009)$_{ab}$ – CBT Wetherell et al. (2005)b – Mean RCI
	Race/ethnicity*	**Conti et al. (2017)$_{b}$ – PSWQ-A** **Conti et al. (2017)$_{b}$ – GAI-SF** Hundt et al. (2014)$_{b}$ -PSWQ Hundt et al. (2014)$_{b}$ -STAI-T Wetherell et al. (2005)b – Mean RCI	Wetherell et al. (2005)b – Mean RCI
	Educational level*	Mohlman et al. (2013)b - PSWQ Conti et al. (2017)$_{b}$ – PSWQ-A Conti et al. (2017)$_{b}$ – GAI-SF Hundt et al. (2014)$_{b}$ -PSWQ Hundt et al. (2014)$_{b}$ -STAI-T Schuurmans et al. (2009)$_{ab}$ – CBT Schuurmans et al. (2009)$_{ab}$ – Sertraline Wetherell et al. (2005)b – Mean RCI	Schuurmans et al. (2009)$_{ab}$ – CBT Wetherell et al. (2005)b – Mean RCI
	Gender*	Majercsik et al. (2004)a - HAM-A Wetherell et al. (2005)b – Mean RCI	Wetherell et al. (2005)b – Mean RCI
	Income*	Conti et al. (2017)$_{b}$ – PSWQ-A Conti et al. (2017)$_{b}$ – GAI-SF	
	Marital Status*	Conti et al. (2017)$_{b}$ – PSWQ-A Conti et al. (2017)$_{b}$ – GAI-SF Wetherell et al. (2005)b – Mean RCI	Wetherell et al. (2005)b – Mean RCI
	Employment status*	Wetherell et al. (2005)b – Mean RCI	Wetherell et al. (2005) – MEAN RCI
Anxiety	Baseline worry severity*	**Bradford et al. (2011)b - PSWQ** **Hundt et al. (2014)$_{b}$ – PSWQ** **Mohlman et al. (2013)b - PSWQ** Mohlman et al. (2020)$_{b}$ -VPAR scores **Mohlman et al. (2020)$_{b}$ -Hippocampal volumes** Wetherell et al. (2005)b – Mean RCI	Bradford et al. (2011)$_{b}$ **Mohlman et al. (2020)$_{b}$ -VPAR scores** Mohlman et al. (2020)$_{b}$ -Hippocampal volumes Wetherell et al. (2005)b – Mean RCI
	Baseline anxiety severity*	**Ayers et al. (2010)b - HAM-A**	**Ayers et al. (2010)b - HAM-A**
	Baseline state-anxiety severity*	**Hundt et al. (2014) -STAI-T**	
	Baseline GAD Severity*	**Caudle et al. (2007)b - Mean RCI** Wetherell et al. (2005)b – Mean RCI	**Caudle et al. (2007)b - Mean RCI** **Wetherell et al. (2005)b – Mean RCI**
	PSWQ 1-month score change*	**Bradford et al. (2011)b - PSWQ**	**Bradford et al. (2011)$_{b}$ - PSWQ**
	Duration of illness†*	**Hendriks et al. (2012)$_{ab}$ – MI-A*** **Hendriks et al. (2012)$_{ab}$ – ACQ†** Schuurmans et al. (2009)$_{ab}$ – CBT† Schuurmans et al. (2009)$_{ab}$ – Sertraline† Wetherell et al. (2005)b – Mean RCI	Schuurmans et al. (2009)$_{ab}$ – CBT† Wetherell et al. (2005)b – Mean RCI

Type of factor	Predictors* / Moderators†	Post-treatment: Response / Remission / Score – Regression model of outcome	Follow-up: Response / Remission / Score – Regression model of outcome
	Late-onset panic disorder†	**Hendriks et al. (2012)$_{ab}$ – MI-A†** **Hendriks et al. (2012)$_{ab}$ – ACQ†**	
Other mental health symptoms	Baseline depression severity*†	Mohlman et al. (2013)b - PSWQ Conti et al. (2017)$_{b}$ – PSWQ-A Conti et al. (2017)$_{b}$ – GAI-SF Hundt et al. (2014)$_{b}$ -PSWQ Mohlman et al. (2013)b - PSWQ **Schuurmans et al. (2009)$_{ab}$ – CBT†** Schuurmans et al. (2009)$_{ab}$ – Sertraline† Wetherell et al. (2005)b – Mean RCI	Schuurmans et al. (2009)$_{ab}$ – CBT Schuurmans et al. (2009)$_{ab}$ – Sertraline Wetherell et al. (2005)b – Mean RCI
	Psychiatric comorbidity*	Caudle et al. (2007)b - Mean Hundt et al. (2014)$_{b}$ -PSWQ Hundt et al. (2014)$_{b}$ -STAI-T **Mohlman et al. (2020)$_{b}$ -VPAR scores** **Mohlman et al. (2020)$_{b}$ -Hippocampal volumes** Wetherell et al. (2005)b – Mean RCI	**Caudle et al. (2007)b - Mean RCI** Mohlman et al. (2020)$_{b}$ -VPAR scores **Mohlman et al. (2020)$_{b}$ -Hippocampal volumes** **Wetherell et al. (2005)b – Mean RCI**
	Agoraphobic symptoms*	Schuurmans et al. (2009)$_{ab}$ – CBT Schuurmans et al. (2009)$_{ab}$ – Sertraline	Schuurmans et al. (2009)$_{ab}$ – CBT Schuurmans et al. (2009)$_{ab}$ – Sertraline
Physical	Hypertension*	Mohlman et al. (2013)b - PSWQ Mohlman et al. (2020)$_{b}$ -VPAR scores Mohlman et al. (2020)$_{b}$ -Hippocampal volumes	 Mohlman et al. (2020)$_{b}$ -VPAR scores Mohlman et al. (2020)$_{b}$ -Hippocampal volumes
	Perceived health†	**Schuurmans et al. (2009)$_{ab}$ – CBT** Schuurmans et al. (2009)$_{ab}$ – Sertraline	**Schuurmans et al. (2009)$_{ab}$ – CBT** Schuurmans et al. (2009)$_{ab}$ – Sertraline
	Disease score*	**Majercsik et al. (2004)a - HAM-A**	
Personality	Total number of personality features†	**Gulpers et al. (2020)$_{ab}$ – MI-A** Gulpers et al. (2020)$_{ab}$ - ACQ	
	Cluster A features†	**Gulpers et al. (2020)$_{ab}$ – MI-A** Gulpers et al. (2020)$_{ab}$ - ACQ	
	Cluster C features*	Gulpers et al. (2020)$_{ab}$ – MI-A Gulpers et al. (2020)$_{ab}$ - ACQ	
	Neuroticism†	**Schuurmans et al. (2009)$_{ab}$ – CBT** Schuurmans et al. (2009)$_{ab}$ – Sertraline	**Schuurmans et al. (2009)$_{ab}$ – CBT** Schuurmans et al. (2009)$_{ab}$ – Sertraline
Therapy	Number of completed sessions*	Conti et al. (2017)$_{b}$ – PSWQ-A Conti et al. (2017)$_{b}$ – GAI-SF Hundt et al. (2014)$_{b}$ -PSWQ **Hundt et al. (2014)$_{b}$ -STAI-T** Wetherell et al. (2005)b – Mean RCI	 Wetherell et al. (2005)b – Mean RCI
	Expectancies of therapy*	Conti et al. (2017)$_{b}$ – PSWQ-A Conti et al. (2017)$_{b}$ – GAI-SF Hundt et al. (2014)$_{b}$ -PSWQ Hundt et al. (2014)$_{b}$ -STAI-T Wetherell et al. (2005)b – Mean RCI	 Wetherell et al. (2005)b – Mean RCI
	Credibility of therapy*	Conti et al. (2017)$_{b}$ – PSWQ-A Conti et al. (2017)$_{b}$ – GAI-SF Hundt et al. (2014)$_{b}$ -PSWQ Hundt et al. (2014)$_{b}$ -STAI-T	

Type of factor	Predictors* / Moderators†	Post-treatment: Response / Remission / Score – Regression model of outcome	Follow-up: Response / Remission / Score – Regression model of outcome
	Therapist adherence*	Hundt et al. (2014)$_b$ -PSWQ Hundt et al. (2014)$_b$ -STAI-T	
	Therapist competence*	Hundt et al. (2014)$_b$ -PSWQ Hundt et al. (2014)$_b$ -STAI-T	
	Homework completion*	**Caudle et al. (2007)b - Mean RCI** Conti et al. (2017)$_b$ – PSWQ-A Conti et al. (2017)$_b$ – GAI-SF Hundt et al. (2014)$_b$ -PSWQ Hundt et al. (2014)$_b$ -STAI-T **Wetherell et al. (2005)b – Mean RCI**	**Caudle et al. (2007)b - Mean RCI** **Wetherell et al. (2005)b – Mean RCI**
Cognitive functioning	Executive skills group based on level, i.e., low, intact, improved*	**Mohlman et al. (2013)b - PSWQ**	
	Non-verbal executive skills*	Mohlman et al. (2013)b - PSWQ	
	Basic cognitive skills group*	Mohlman et al. (2013)b - PSWQ	
	Cognitive impairment (measured by MMSE)*	Wetherell et al. (2005)b – Mean RCI	Wetherell et al. (2005)b – Mean RCI
	Information processing bias*	**Steiner et al. (2013)$_a$ –GADSS** Steiner et al. (2013)$_a$– PSWQ Steiner et al. (2013)$_a$ – HAM-A	
	VPAR scores*	Mohlman et al. (2020)$_b$ -VPAR scores	**Mohlman et al. (2020)$_b$ -VPAR scores**
	Orientation*	Caudle et al. (2007)b - Mean RCI	**Caudle et al. (2007)b - Mean RCI**
	Working Memory*	Caudle et al. (2007)b - Mean RCI	Caudle et al. (2007)b - Mean RCI
	Delayed Recall*	Caudle et al. (2007)b - Mean RCI	Caudle et al. (2007)b - Mean RCI
	Language/Praxis*	Caudle et al. (2007)b - Mean RCI	Caudle et al. (2007)b - Mean RCI
Neuroanatomical Medication	Hippocampal volume*	Mohlman et al. (2020)$_b$ -Hippocampal volumes	Mohlman et al. (2020)$_b$ -Hippocampal volumes
	Psychotropic medication use*	Hundt et al. (2014)$_b$ -PSWQ Hundt et al. (2014)$_b$ -STAI-T	**Ayers et al. (2010)b - HAM-A**
Support provision	Use of other professionally delivered mental health services*	Ayers et al. (2010)b - HAM-A	Ayers et al. (2010)b - HAM-A
	Social support*†	Hundt et al. (2014)$_b$ -PSWQ Hundt et al. (2014)$_b$ -STAI-T **Majercsik et al. (2004)$_a$† - HAM-A**	
Life events	Number of positive life events*		Ayers et al. (2010)b - HAM-A

Type of factor	Predictors* / Moderators†	Post-treatment: Response / Remission / Score – Regression model of outcome	Follow-up: Response / Remission / Score – Regression model of outcome
	Number of negative life events*		**Ayers et al. (2010)b - HAM-A**
Other factors	Recruitment site*	**Hundt et al. (2014)$_b$ -PSWQ** **Hundt et al. (2014)$_b$ -STAI-T**	
	Problem-solving confidence*	Hundt et al. (2014)$_b$ -PSWQ Hundt et al. (2014)$_b$ -STAI-T	
	Source study*	**Caudle et al. (2007)b - Mean** Wetherell et al. (2005)b – Mean RCI	**Caudle et al. (2007)b - Mean RCI** **Wetherell et al. (2005)b – Mean RCI**
	Provided type (expert non-expert)*	Conti et al. (2017)$_b$ – PSWQ-A Conti et al. (2017)$_b$ – GAI-SF	
	Escape/avoidance coping style*		**Ayers et al. (2010)b - HAM-A**

Notes: a= pharmacological study, b=psychological therapy study, ab= combined pharmacotherapy and psychological therapy. * = predictor, †= moderator. Factors assessed as predictors and moderators have both * and † next to them. In such cases, the † symbol has been added next to the studies in the post-treatment and/or follow-up columns indicating which study assessed the factor as a moderator. Bold text = statistically significant factor. VPAR = Verbal Paired Associates Retention, GAD = Generalised Anxiety Disorder, MMSE = Mini-Mental State Examination

Post-treatment: Response / Remission / Score – Regression models of outcome & *Follow-up: Response / Remission / Score – Regression model of outcome* ACQ = Agoraphobic Cognitions Questionnaire, CBT = Cognitive Behavioural Therapy, GAI-SF = Geriatric Anxiety Inventory Short Form, HAM-A = Hamilton Anxiety Rating Scale, MI-A = Mobility Inventory Avoidance scale, PSWQ = Penn State Worry Questionnaire, PSWQ-A = Penn State Worry Questionnaire – short version, RCI = Reliable Change Index, STAI-T = Spielberger State-Trait Anxiety Inventory - trait subscale.
The regression model of outcome indicates that the separate regression models conducted for each outcome explored within each study.

Supplementary Material 1: PRISMA checklist.

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	1
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	3-4
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	4
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	4-5
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	5-6
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	5
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	5-6
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	5-7
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	6-7
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	6-7
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	6
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	7
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	7
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	7
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	7
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the	7

Section and Topic	Item #	Checklist item	Location where item is reported
		model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	N/A
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	N/A
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	N/A
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	N/A
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	7-8, 22
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	N/A
Study characteristics	17	Cite each included study and present its characteristics.	8-9, 22-26
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	9, 26-29, 30-33
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	10-13, 34-37
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	6-8
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	N/A
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	N/A
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	N/A
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	N/A
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	13-15
	23b	Discuss any limitations of the evidence included in the review.	16
	23c	Discuss any limitations of the review processes used.	16
	23d	Discuss implications of the results for practice, policy, and future research.	15
OTHER INFORMATION			

Section and Topic	Item #	Checklist item	Location where item is reported
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	4
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	4
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	4
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	21
Competing interests	26	Declare any competing interests of review authors.	21
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	N/A

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. https://dio.org/10.1136/bmj.n71.

Supplementary Material 2: Main results of included studies.

Condition	Author/ year	Statistical analysis	Predictor(s) / Moderator(s)	Outcomes	Results for late-life anxiety in relation to predictor analysis and moderator analysis
GAD	Ayers et al. (2010)	Linear random effect regressions	Baseline anxiety, psychotropic medication use, use of other professionally-delivered mental health services, escape/avoidance coping style, number of positive life events, number of negative life events	HAM-A score at post-treatment	• Use of other professionally-delivered mental health services (P = 0.76) and number of positive life events (P = 0.39) did not predict anxiety symptoms. • Baseline anxiety (β = 0.75, SE = 0.12, P < 0.01), psychotropic medication use (β = 4.02, SE = 1.32, P < 0.01), escape/avoidance coping style (β = 4.76, SE = 1.75, P < 0.02) and number of negative life events (β = 0.19, SE = 0.08, P < 0.02) predicted anxiety symptoms.
GAD	Bradford et al. (2011)	Binary logistic regression	Baseline PSWQ score, PSWQ 1-month change score	PSWQ at post-treatment and at follow-up	• Baseline PSWQ score predicted anxiety symptoms at 3 months - $\chi 2$ (N = 76, df=1) = 4.93, P = 0.026, OR = 1.08 - but not at 15 months, P = 0.82. • PSWQ 1-month change score predicted anxiety symptoms at 3 months - $\chi 2$ (N = 76, df = 1) = 4.65, p = 0.031, OR= 0.92 - and at 15 months - $\chi 2$ (N = 76, df = 1) = 6.15, p = 0.013, OR = 0.92.
GAD	Caudle et al. (2007)	Hierarchical regression	Baseline GAD severity, number of comorbid psychiatric disorders, homework completion, study, presence/absence of errors in the following cognitive domains: orientation, working memory, language/praxis, delayed recall.	Reliable change index (RCI) based on the anxiety symptoms measured across the anxiety outcome variables used at post-treatment & at 6-month follow-up	• When controlling for the effects of study, GAD severity at baseline, homework completion, and psychiatric comorbidity, the model accounted for a significant amount of variance at both posttreatment (R^2 = 25%, F(5,59) = 3.87, p = 0.005) and at 6-month follow-up (R^2 = 40%, F(5,58) = 7.77, p <0.0001). • When controlling for the above variables and the presence of errors on each MMSE domain, the model remained significant at post-treatment (R^2 = 29%, F(9,55) = 2.53, p = 0.02). No significant cognitive predictors of RCT were identified. The model also remained significant at 6-month follow-up (R^2 = 48%, F(9,54) = 5.50, p <0.0001). The Orientation domain of the MMSE was the only individually significant predictor of RCI at follow-up (β = 0.24, t(54) = 2.30, p = 0.03), over and above the already established effects of study, GAD severity, comorbidity, and homework completion.

Condition	Author/ year	Statistical analysis	Predictor(s) / Moderator(s)	Outcomes	Results for late-life anxiety in relation to predictor analysis and moderator analysis
GAD & ADNOS	Conti et al. (2017)	Separate linear regression equation for each individual predictor variable	Group, provider type, age, race, education, income, marital status, baseline PHQ-8, expectancies of therapy, credibility of therapy, homework completion, number of sessions	3-month PSWQ-A and GAI-SF	• Younger age (p= .031) and African American race (p= .028) were both significantly associated with higher post-treatment PSWQ-A scores, controlling for pre-treatment PSWQ-A score. Baseline PSWQ-A score was a significant predictor of post-treatment PSWQ-A in all models (all p-values < 0.01). The rest variables were non-significant predictors (P > 0.05). • African American race (p= .033) was significantly associated with higher post-treatment GAI-SF scores, controlling for pre-treatment GAI-SF scores. Baseline GAI-SF score was a significant predictor in all models (all p-values < .01). The rest variables were non-significant predictors (P > 0.05).
Panic disorder with agoraphobia	Gulpers et al. (2020)	Multivariate regression analyses	Total number of personality features, cluster A features, cluster B features, cluster C features	MI-A and ACQ at post-treatment	• Cluster B features and cluster C features did not predict avoidance behaviour and agoraphobic conditions in either the paroxetine or the CBT groups ($P > 0.05$). • Total number of personality features predicted avoidance behaviours ($P = 0.028$), although, it did not predict agoraphobic cognitions ($P = 0.178$). • Cluster A features predicted avoidance behaviour in both paroxetine ($\beta = 0.49$, SE = 0.03, $P = 0.019$) and CBT groups ($\beta = -0.44$, SE = 0.03, $P = 0.017$). Cluster A features did not predict agoraphobic cognitions in the CBT nor paroxetine group ($P > 0.05$).
Panic disorder with agoraphobia	Hendriks et al. (2012)	Multiple regression analyses	Chronological age, age of onset and duration of illness	MI-A and ACQ at post-treatment	• Chronological age did not predict agoraphobic conditions nor avoidance behaviour in any of the treatment groups (P > 0.05). • Late-onset panic disorder did not predict agoraphobic cognitions (P = 0.11) nor avoidance behaviour (P = 0.54) in the paroxetine group. Late-onset panic disorder predicted agoraphobic cognitions (R2 = 0.61, β = 0.40, P = 0.03) and avoidance behaviour (R2 = 0.60, β = 0.51, P = 0.01) in the CBT group. • Duration of illness predicted avoidance behaviours (R2 = 0.52, β = 0.26, P = 0.05) in both groups. Duration of illness predicted agoraphobic cognitions in the CBT group (R2 = 0.68, β = 0.48, P = 0.01) but not in the paroxetine group (P = 0.78).

Condition	Author/ year	Statistical analysis	Predictor(s) / Moderator(s)	Outcomes	Results for late-life anxiety in relation to predictor analysis and moderator analysis
GAD	Hundt et al. (2014)	Intent-to-treat multivariate analyses	Baseline worry severity, baseline anxiety severity, homework completion, problem-solving confidence, no of CBT sessions completed, expectancies for therapy, credibility for therapy, age, race/ethnicity, education, social support, PHQ-9, no of mental health diagnoses, mental health medications, therapist adherence, therapist competence	6-month PSWQ-A & 6-month STAI-T	• Baseline PSWQ-A predicted 6-month PSWQ-A (β = 0.553, SE = 0.102, P < 0.0001). Baseline STAI-T predicted 6-month STAI-T (β = 0.57, SE = 0.09, *P* < 0.001). • The site predicted 6-month PSWQ-A (β = 2.788, SE = 1.233, P = 0.028) and 6-month STAI-T (β = 3.50, SE = 1.56, *P* = 0.03). • Sessions completed predicted 6-month STAI-T (*β* = 6.07, *SE* = 2.48, *P* = 0.02) but not 6-month PSWQ-A (*P* = 0.109). • The rest predictors did not predict treatment outcomes (*P* > 0.05).
GAD	Majercsik et al. (2004)	Multiple regression analysis	Gender, number of social contacts, health status	HAM-A score at post-treatment	• Pharmacological therapy (R=0.76; F(4,150)=52.38; p<0.0001), number of social contacts (β=0.13±0.05; p<0.016), and the health status (β=0.11±0.05; p<0.05) predicted anxiety severity at post-treatment, i.e. improvement. Gender had a minimal effect on improvement (β=−0.04±0.05; p<0.4). • Improvement was significantly larger in buspirone-treated patients that had many social contacts (F(3,101)=3.47; p<0.018) • A similar effect was found for the disease score, but differences did not reach statistical significance (F(3,101)=1.38; p<0.3)
GAD	Mohlman et al. (2013)	Hierarchical regression & binomial logistic regression	Education, hypertension status, baseline PSWQ scores, baseline BDI scores, basic skills scores, group category based on executive skills i.e. low, intact, improved.	Pre-posttreatment change scores on the PSWQ and GAD-Q-IV, the proportion of patients who met a threshold criterion for high end state functioning	• A PSWQ model comprising of pre-treatment scores on the measures of anxiety on the first step, education, hypertension status and BDI scores on the second step, and basic scores and group based on executive skills on the third step, was significant, F(7,51) = 3.944, p < 0.01, adj. R2 = 0.38. Significant individual predictors were baseline PSWQ (t(51) = −3.240, p < .005) and ES Group (t(51) = 2.559, p < .05). The full GAD-Q-IV model was not significant (p > .05). • The model with the following variables: education, hypertension status, BDI, basic, Verbal ES, Nonverbal ES scores, predicting high end-state functioning was not significant (p > .05).

Condition	Author/ year	Statistical analysis	Predictor(s) / Moderator(s)	Outcomes	Results for late-life anxiety in relation to predictor analysis and moderator analysis
GAD	Mohlman et al. (2020)	Multiple regression analyses	Number of comorbid disorders, baseline PSWQ scores, hypertension, VPAR subtest of the Wechsler Memory Scales-III, hippocampal volume	PSWQ score at post-treatment	• The full multiple regression model including VPAR as an index of WM was significant at post-treatment, F(4,40)=4.057, p< .01, adj. r2=.26. On the first step, the number of comorbid disorders was negatively associated with PSWQ immediately following treatment. The full posttreatment model including hippocampal volume as an index of WM was significant, F(3,35)=8.155, p< .001, adj. r2=.37. On the first step, baseline PSWQ scores were positively associated, and the number of comorbid disorders at baseline was negatively associated to posttreatment PSWQ. • The full model of VPAR was significant at follow-up, F(5,22)=9.062, p< .001, adj. r2=.60. On the first step, baseline PSWQ scores were positively associated, and on the third step, VPAR scores were negatively associated to follow-up PSWQ scores. The full model of hippocampal volume was significant at follow-up, F(3,16)=4.964, p<05, adj. r2=.53,. On the first step, baseline PSWQ scores were positively associated to endpoint PSWQ.
GAD, panic disorder, agoraphobia, social phobia	Schuurmans et al. (2009)	Multiple forward regression analyses	Age, educational level, duration of symptoms, depressive symptoms, agoraphobic symptoms, neuroticism, perceived health	HAM-A score at post-treatment	• At post-treatment, lower perceived health predicted anxiety severity following CBT (β=−0.63, $P < 0.01$). No other predictors were identified for CBT or sertraline outcome at post-treatment. • At one-year follow-up, neuroticism predicted anxiety severity following CBT (β=−0.63, $P < 0.01$). No other predictors were identified for CBT or sertraline outcome at one-year follow-up.
GAD	Steiner et al. (2013)	Repeated measures mixed-effect regression models	Information processing bias	HAM-A score, PSWQ score, GADSS score	• The information response bias was associated with GAD symptoms, $F(1,24) = 4.28$; $P = 0.009$. • The information response bias was not associated with clinical improvement in anxiety symptoms, $P = 0.49$, nor with worry symptoms, $P = 0.06$.

Condition	Author/ year	Statistical analysis	Predictor(s) / Moderator(s)	Outcomes	Results for late-life anxiety in relation to predictor analysis and moderator analysis
GAD	Wetherell et al. (2005)	Separate multiple regression analysis based variables group	Demographic variables: Study, gender, age, ethnicity, education, marital status, work status. Clinical variables: psychiatric comorbidity, GAD duration, cognitive impairment measured by the MMSE. Initial levels of anxiety and depressive symptoms measured by baseline GAD severity, PSWQ, HAM-A, BDI, and HAMD. Treatment variables: treatment expectancy, number of sessions attended, homework adherence.	Reliable change index based on the anxiety symptoms measured across the anxiety outcome variables used at post-treatment & at 6-month follow-up	• Demographic variables did not predict mean RCI at posttreatment, or 6-month follow-up ($p > 0.05$). • Clinical variables did not predict mean RCI at posttreatment, ($p > 0.05$). However, clinical variables predicted RCI at 6-month follow-up, ($F(5, 58) = 2.55$, $p = .04$). Enrollment in Stanley's 1996 study, (beta = .342, $p = .02$), and the presence of a comorbid psychiatric diagnosis, (beta -- .310, $p = .01$) were significant predictors. • Baseline severity of psychopathology did not predict mean RCI at posttreatment ($p > 0.05$), although it predicted mean RCI at 6-month follow-up ($F(7, 55) = 2.58$, $p = .02$). Baseline GAD severity was the only significant variable (beta = .285, $p = .04$); higher levels of GAD severity were associated with better treatment response. • Treatment variables were significant both at posttreatment, $F(5, 57) = 2.57$, $p = .04$, and at follow-up, $F(5, 56) = 2.78$, $p = .03$. Homework adherence was the only variable that achieved significance at post-treatment (beta = .291, $p = .04$) and at 6-month follow-up (beta = .344, $p = .02$), with greater homework completion associated with better response.

Results for late-life anxiety in relation to predictor analysis or moderator analysis PSWQ = Penn-state worry questionnaire, CBT = cognitive behavioural therapy, WM = working memory, VPAR = Verbal paired associates retention, GAD = generalised anxiety disorder