Evidence-based assessment of treatment outcomes for late-life Generalised Anxiety Disorder using the Penn State Worry Questionnaire (PSWQ) and Penn State Worry Questionnaire – Abbreviated (PSWQ-A).

Carly Johnco, PhD,[a,b] Viviana M. Wuthrich, PhD,[a,b] Gretchen A. Brenes, PhD,[c] Julie Loebach Wetherell, PhD,[d] Jan Mohlman, PhD,[e]

[a]Centre for Emotional Health, Department of Psychology, Macquarie University, Sydney, Australia

[b]Centre for Ageing, Cognition and Wellbeing, Macquarie University, Sydney, Australia

[c] Department of Internal Medicine, Section on Gerontology and Geriatric Medicine, Wake Forest School of Medicine

[d]Mental Health Impact Unit 3, VA San Diego Healthcare System and Department of Psychiatry, University of California, San Diego

[e]Department of Psychology, William Paterson University, Wayne, NJ USA

Corresponding author: Carly Johnco, Department of Psychology, Macquarie University, Sydney, NSW Australia, 2109. E-mail: carly.johnco@mq.edu.au, Ph: +61 9850 8053.

Funding

This work was supported by the National Health and Medical Research Council of Australia, National Institute of Mental Health and National Alliance for Research on Schizophrenia and Depression (NARSAD) Young Investigator Grant.

Declarations of interest

None.

Highlights

- Developed guidelines to asses treatment outcome using Penn State Worry Questionnaire (PSWQ)
- Remission is defined by scores ≤51 on the PSWQ, and ≤24 on the abbreviated scale.
- Treatment response is defined by a 9% reduction or ≥4 point reduction on the PSWQ.
- The PSWQ-Abbreviated was poor at identifying treatment response status
- New guidelines will improve consistency when assessing treatment for late-life GAD.

Abstract

Objective: The Penn State Worry Questionnaire (PSWQ) is a commonly used measure of treatment outcome for late-life Generalised Anxiety Disorder (GAD). However, there is considerable variability in the definitions used to define treatment response and remission. This study aimed to provide empirically derived guidelines for assessing treatment response and remission among older adults with GAD using the PSWQ and the abbreviated PSWQ (PSWQ-A).

Design: Longitudinal assessment of GAD symptoms pre- and post-treatment.

Participants: Participants were 259 older adults aged 60 – 86 with a diagnosis of GAD who were assessed before and after treatment.

Intervention: Participants were randomly assigned to Cognitive Behavioural Therapy or Control (waitlist, discussion group or supportive therapy) conditions.

Measurements: Signal detection analyses using receiver operating characteristic methods were used to determine optimal agreement between structured diagnostic interviews and scores on the PSWQ and PSWQ-A.

Results: Results suggest that a score of ≤51 was optimal for defining diagnostic remission status on the PSWQ, and a score of ≤24 was optimal on the PSWQ-A. A 9% reduction or ≥4 point reduction was optimal for assessing treatment response on the PSWQ. The PSWQ-A was poor at identifying treatment response status.

Conclusions: Findings suggest that most of the previously used definitions have underestimated the treatment effects for late-life GAD. However overall, the PSWQ and PSWQ-A are suboptimal for assessing treatment outcome for late-life GAD. The standardisation of response and remission criteria has implications for comparison between treatment trials, and for the benchmarking of outcomes in clinical practice.

Keywords: Anxiety; Generalized Anxiety Disorder; GAD; Penn State Worry Questionnaire; PSWQ; Geriatric; Older Adult; Elderly.

Evidence-based assessment of treatment outcomes for late-life Generalised Anxiety Disorder using the Penn State Worry Questionnaire and Penn State Worry Questionnaire - Abbreviated.

Generalised Anxiety Disorder (GAD) is one of the most common anxiety disorders among older adults (Wolitzky-Taylor et al., 2010, Beekman et al., 1998). Although the majority of psychological treatment studies for late-life anxiety have focused on treatment of GAD (e.g., Thorp et al., 2009, Hall et al., 2016, Gonçalves and Byrne, 2012, Hendriks et al., 2008, Gould et al., 2012, Pinquart and Duberstein, 2007), there is considerable variability in the assessment of treatment outcomes. The Penn State Worry Questionnaire (PSWQ; Meyer et al., 1990) is one of the most widely-used measures of worry severity among clinical trials for late-life GAD. In comparison to diagnostic clinical interviews, self-report measures are less costly and time-consuming to administer, making them desirable for use in research and clinical practice to inform treatment progression and discontinuation. Change on the PSWQ is the most commonly used metric of treatment efficacy in randomised control trials for late-life GAD. Despite this, there is considerable variability in the way this measure has been used to define treatment outcome among studies. Empirically derived guidelines for defining treatment response and remission on the PSWQ are necessary to help standardise research outcomes. These guidelines are also necessary for use in clinical practice to benchmark client progress and inform treatment decision-making.

Among older adults, the PSWQ has demonstrated good internal consistency, convergent validity, and modest test-retest reliability (Stanley et al., 2001, Wuthrich et al., 2014). However, discriminant validity from measures of depression is variable, with some studies reporting moderate correlations (r = .3-.7; Wuthrich et al., 2014, Hopko et al., 2003) and others reporting small correlations (r = .2; Stanley et al., 2001). At the recommended threshold of scores >50, the PSWQ is able to differentiate GAD from non-clinical samples

with 62-90% sensitivity and 70-94% specificity (Wuthrich et al., 2014, Stanley et al., 2003b, Webb et al., 2008), with moderate to good agreement with presence of a GAD diagnosis on structured diagnostic interviews (kappa = .5-.7; Wuthrich et al., 2014, Webb et al., 2008).

Originally developed with younger adults, the PSWQ demonstrates problems with factor structure in older samples. Despite being designed to assess a unitary construct, the PSWQ commonly shows a two-factor structure among older adults, with reverse-coded items loading on a second factor (Beck et al., 1995), or a poor model fit demonstrated for both a one and two factor structure (Hopko et al., 2003, Wuthrich et al., 2014). The problematic factor structure has raised concern as to the suitability of this measure for older samples, with suggestions that geriatric-specific measure may be more appropriate, such as the Geriatric Anxiety Inventory (Pachana et al., 2007) or Geriatric Anxiety Scale (Segal et al., 2010). To overcome this limitation, an abbreviated 8-item version (PSWQ-A; Hopko et al., 2003) was developed, which has shown similar psychometric properties to the full PSWQ, but with a single factor. At the recommended cut-off of >22, the PSWQ-A has similar predictive validity to the full PSWQ to discriminate between those with a GAD diagnosis and those without, with 66-80% sensitivity and 63-93% specificity and moderate to good agreement with diagnostic interviews (kappa = .4-.7; Wuthrich et al., 2014, Webb et al., 2008, Stanley et al., 2003b). The PSWQ more accurately predicts GAD status than the PSWQ-A (Wuthrich et al., 2014, Webb et al., 2008).

Although several studies have examined the psychometric properties of the PSWQ and PSWQ-A as a screening measure for identifying GAD, there are no studies assessing its validity for assessing treatment outcome. The two important measures of treatment outcome are treatment response (a clinically meaningful improvement in symptoms as a result of treatment), and remission (disappearance or absence of a diagnosis). Although these terms are often used interchangeably, they reflect qualitatively different things and have different

implications. Remission is the ultimate goal of treatment, reflecting that the individual no longer meets criteria for a diagnosis. However, assessing treatment response is important to determine whether a treatment is effective in reducing symptoms, and to inform whether a treatment should be continued, augmented or changed. The severity of a client's disorder also affects these metrics. Those with higher baseline severity may show a meaningful reduction in symptoms (i.e., treatment response), but may still exceed a diagnostic threshold (i.e., not be in remission). This information is important to inform treatment planning, and indications of response but not remission status would suggest a need to continue treatment. In contrast, those with mild disorder severity may achieve remission with a smaller reduction in disorder severity. Thus, both response and remission metrics are important to assess.

Recommended cut-off scores for detecting GAD diagnosis on the PSWQ cannot be applied to determine treatment response, as these absolute values do not provide a metric of change. Further, despite the similar goal of differentiating those with a GAD diagnosis from those without, it cannot be assumed that cut-off scores derived for screening purposes will be optimal for the assessment of diagnostic remission. Cut-off scores recommended for screening purposes prioritise sensitivity – the ability of a test to accurately identify those with the disorder. Although specificity (the ability for a test to correctly identify those without the disease) is also important, the risk of a false negative test may result in a disorder going undetected and untreated. While cut-off scores aim to optimise all screening metrics, sensitivity is typically prioritised over specificity when determining thresholds for screening measures in order to avoid false negatives. In contrast, the primary concern when assessing remission status is whether people may be incorrectly classified as in remission when they are not, resulting in premature treatment discontinuation. As such, the positive predictive value (the proportion who are classified as being remitted by the test who are actually in remission)

and specificity are prioritised when assessing remission. As such, there is a need for empirically defined guidelines for assessing treatment outcome on the PSWQ.

Despite being one of the most widely-used measures of treatment outcome in clinical trials for late-life GAD, there is significant variability among trials in how this measure is used to assess treatment outcome. For example, definitions of treatment response (sometimes referred to minimal clinically significant change) have included an 8.5point reduction on the PSWQ (Stanley et al., 2009, Bradford et al., 2011, Wetherell et al., 2013), or a 20% reduction on 66-75% of outcome measures, including PSWQ (Wetherell et al., 2003, Stanley et al., 1996, Stanley et al., 2003c, Stanley et al., 2003a). One study used a reliable change index (Wetherell et al., 2005). Others have used a more conservative approach that combines traditional definitions of remission and response, defining treatment responders as those who have both an absence of GAD diagnosis and 20% reduction in 75-80% of outcome measures (Mohlman, 2008, Mohlman and Gorman, 2005, Mohlman et al., 2003). Definitions of remission status or high endstate functioning have included achieving a score within one standard deviation of the mean for normal older adults samples across 3-4 outcomes measures, although the exact threshold level on the PSWQ has again differed among studies, including a score of ≤37 (Wetherell et al., 2003, Stanley et al., 1996), ≤36 (Stanley et al., 2003a), and ≤43 (Mohlman, 2008). To date, only one study has attempted to examine the validity of different definitions; however, this study only assessed agreement between different definitions, but not against gold-standard diagnostic interviews (Roseman et al., 2011). This study found moderate agreement (kappa=.46) between an 8.5 point reduction and a 20% reduction on 75% of outcome measures that included the PSWQ. Despite this moderate agreement between these two indices, it is important to assess convergence with gold-standard diagnostic interviews.

Given its more recent development and primary use as a screening measure, few studies have used the PSWQ-A to assess treatment outcome. Brenes et al. (2015) defined treatment response as a 5.5 point decrease in PSWQ-A score, based on the mean reduction in symptoms from Barrera et al. (2015), and Stanley et al. (2014) defined response based on a 20% reduction in 75% of outcome measures. The notable variability in thresholds, and the inconsistent assessment of both treatment response and remission in studies, makes it difficult to compare results across trials and for clinicians to benchmark client progress in practice.

This study aimed to define optimal guidelines for assessing treatment response and remission among older adults with GAD using the PSWQ and PSWQ-A. Remission status was defined as loss of GAD diagnosis using a standardised diagnostic interview. Treatment response was defined using a reliable change index from change in GAD diagnosis severity on a standardised diagnostic interview. Raw score threshold levels are most relevant to the assessment of remission status (i.e., if an individual reaches the threshold level, they are likely to be in remission from a GAD diagnosis), thus guidelines were developed for defining treatment remission based on a threshold score. Percent reduction in symptoms, raw score reductions and threshold values have all been used in previous studies defining treatment responder status. Only percent reduction and raw score reduction metrics were considered when developing guidelines for predicting treatment responder status given that threshold values do not provide an index of change.

Method

Participants

The sample utilised secondary data analysis from seven studies of cognitive behavioural therapy (CBT) for late-life anxiety disorders from four sites: Macquarie University, Australia (Wuthrich and Rapee, 2013, Johnco et al., 2014), University of California San Diego, USA

(UCSD; Wetherell et al., 2003), William Paterson University, USA (Mohlman, 2008, Mohlman and Gorman, 2005, Mohlman, 2020) and Wake Forest School of Medicine, USA (Brenes et al., 2015). The total sample consisted of 259 older participants (age range 60–86 years; M=67.22, SD=5.75; 74.13% female). Participants had a diagnosis of Generalised Anxiety Disorder (GAD) at entry into the respective treatment trial based on a structured diagnostic interview - either the Anxiety Disorders Interview Schedule for DSM-IV (ADIS-IV; Di Nardo et al., 1994) or Structured Clinical Interview for the DSM (SCID; First et al., 2002). PSWQ data was available used from three sites: Macquarie University, UCSD (2003) and William Paterson University. PSWQ-A data was extracted from full-scale responses from Macquarie University, UCSD and William Paterson University, with additional PSWQ-A data (N=118) included from Wake Forest University. Thus, analyses on the PSWQ utilised a subsample of 141 participants (Age Range 60-86 years, M=87.83, SD=5.79; 66.67% female) and analyses for the PSWQ-A utilised the total sample. Table 1 summarises demographic and clinical characteristics.

2.2. Measures

2.2.1. Diagnostic Interviews. Structured clinical interviews were conducted pre- and post-treatment in all studies. The ADIS-IV (Di Nardo et al., 1994) was used at Macquarie University and UCSD, and the SCID-IV (SCID; First et al., 2002) was used at Wake Forest and William Paterson University. The ADIS-IV and SCID-IV are semi-structured interviews for diagnosing anxiety and related disorders according to DSM-IV criteria. GAD diagnosis severity was rated using the 0-8 Clinician Severity Ratings from the ADIS at MQ, UCSD and Wake Forest, with scores of ≥4 indicating severity above diagnostic thresholds. The 0-3 disorder severity ratings from the SCID was used for William Paterson (0=*no diagnosis/none*, 1=*mild*, 2=*moderate*, 3=*severe),* with scores ≥1 indicating severity above diagnostic

thresholds. Diagnostic interviews were administered by graduate-level Clinical Psychology students, Psychologists and Clinical Psychologists blind to treatment allocation and supervised by licenced Clinical Psychologists. Based on data from the original studies from which this data was drawn, interrater reliability for the diagnosis of GAD was excellent at Macquarie University (kappa=1.00), strong at William Paterson (kappa=.90) and moderate at UCSD (kappa=.87).

2.2.2. Penn State Worry Questionnaire (PSWQ; Meyer et al., 1990) and Penn State Worry Questionnaire – Abbreviated (PSWQ-A; Hopko et al., 2003). The PSWQ is a 16-item measure of worry severity, and the PSWQ-A is the 8-item abbreviated version. Both measures ask individuals to rate their level of trait worry on a 5-point likert scale. Total scores range from 16-80 for the PSWQ and 8-40 for the PSWQ-A, with higher scores reflected greater anxiety. Internal consistency was excellent for the PSWQ and PSWQ-A at pre-treatment (α=.87 and .86) and post-treatment (α=.91 and .91) in this sample.

2.3. Treatment Conditions

Participants from Macquarie University were randomised to a manualised group Cognitive Behavioural Therapy (CBT) condition or a waitlist control condition (Wuthrich and Rapee, 2013). CBT utilised the Ageing Wisely treatment program (Wuthrich, 2009) which consisted of 12x2hours group therapy sessions for anxiety and depression in older adults, and included behavioural activation, cognitive restructuring, problem solving, exposure therapy, assertiveness, and managing bereavement.

Participants from UCSD were randomly assigned in small groups to one of three conditions: a 12x1.5 hour CBT group, discussion group, or a 12-week waitlist condition. CBT consisted of relaxation training, cognitive restructuring, and worry exposure based on the Craske et al. (1992) manual, with minor modifications for older adults (omission of time

management and benzodiazepine withdrawal and examples made relevant to older people). The discussion group consistent of group discussions focused on topics known to be worry-provoking for older adults, including memory problems, health concerns, and bereavement. Group leaders provided validation and supportive listening, encouraged peer support and information sharing, but did not teach skills or differentially reinforce coping strategies.

Participants from William Paterson University were randomly assigned to either CBT or a waitlist condition. CBT consisted of 13 individual 50-minute sessions that included relaxation, cognitive restructuring, worry exposure, activity planning, time management, breaking tasks down into small steps, communication skills, and mindfulness. .

Participants from Wake Forest University were randomized to either CBT or supportive therapy. CBT consisted of 9-11 weekly sessions via telephone and included relaxation, cognitive restructuring, coping statements, problem solving, worry control, behavioural activation, and exposure therapy. Supportive therapy consisted of 10 weekly non-directive sessions that emphasised supportive and reflective communication, but avoided advice or coping methods.

2.4. Procedure

Participants were randomly allocated to a manualised CBT or a control condition within the original randomised clinical trials from which this data was drawn. Data from diagnostic assessments and PSWQ/PSWQ-A administered at pre- and post-treatment were used in this study.

2.5. Data analysis

There was a minimal amount of missing data (.65%), and the pattern of missingness was considered missing completely at random (Little's MCAR $\chi(337)^2=314.73$, p=.803). As such, missing values were imputed using the expectation maximisation algorithm.

ROC analyses were used to assess optimal thresholds on the PSWQ and PSWQ-A for defining treatment remission and response status based on diagnostic interviews. Remission was defined as loss of GAD diagnosis on diagnostic interviews. Treatment responder status was defined based on a reliable change index (RCI) utilising GAD diagnosis severity ratings from diagnostic interviews, with a RCI (Jacobson and Truax, 1991) calculated separately for severity ratings on the ADIS (range 0-8) and SCID (range 0-3) to determine responder status.

The Area Under the Curve (AUC) statistic was used to assess the accuracy of the PSWQ and PSWQ-A to differentiate those who were remitted/responders and those who were not, with values of 1 representing perfect prediction and values of .5 being no better than chance. AUC of .5-.6 is considered poor, .7-.8 is acceptable, .8-.9 is excellent and >.9 is outstanding (Mandrekar, 2010). Consistent with previous signal detection studies (e.g., Johnco et al., 2015b, Johnco et al., 2015a, Caporino et al., 2013, Storch et al., 2011, Storch et al., 2010, Houghton et al., 2015, Skarphedinsson et al., 2017), six metrics were used to evaluate various cut-off scores. Sensitivity (i.e., true positive rate) is the proportion of participants who meet remission status based on diagnostic interviews who would be correctly classified as in remission based on the cut-off score (true positive/[true positive + false positive]). Specificity (i.e., true negative) is the proportion of participants who do not meet remission status, who would be correctly classified as not remitted based on the test cut-off score (true negative/[false positive + true negative]). Positive Predictive Value (PPV) is the proportion of participants who are identified as being remitted by the test, who are also classified as remitted based on clinical interviews (True Positive/[True Positive + False Positive]). Negative Predictive Value (NPV) is the proportion of participants who are

identified as not remitted based on the test cut-off who are also not considered remitted based on diagnostic interviews (True Negative/[True Negative + False Negative]. Percent agreement (Agreement) between remission status as classified by clinical interviews and test cut-off scores was calculated. Finally, agreement between clinical interviews and test cut-off scores was also evaluated using Cohen's kappa, given this metric accounts for chance agreement and is a preferred metric to simple percent agreement. A kappa of .2-.4 indicates fair agreement, .4-.6 indicates moderate agreement, .6-.8 indicates substantial agreement, and ≥.8 indicates almost perfect agreement (Landis and Koch, 1977).

When deciding on optimal cut-offs, the purpose of test use was considered, and a conservative approach was adopted. Classifying an individual as 'remitted' would indicate treatment success and likely prompt treatment discontinuation. Given the risk of prematurely discounting treatment, kappa, PPV and specificity metrics were prioritised when evaluating cut-off scores. Given that classifying an individual as having 'responded' to treatment would indicate the need for continued treatment but not necessarily discontinuation, metrics were weighted more equally when evaluating optimal cut-off scores.

Results

Preliminary analyses

Demographic and clinical differences between samples are noted in Table 1. There were no significant group differences in age or employment status. There was some sample differences in sex distribution, with a greater proportion of males in the MQ sample and the lowest proportion in the Wake Forest sample. There were also some sample differences in comorbidity, with a greater proportion of comorbid mood disorders and a lower proportion with no comorbid disorder in the Macquarie and Wake Forest samples. There was no evidence that these sample differences in sex distribution influenced core variables of

interest, with no significant sex differences in PSWQ or PSWQ-A scores at pre- or post-treatment, or in remission or responder status (all *p*'s > .05). Similarly, there were no differences in remission or responder status based on having a comorbid mood disorder or no comorbidity, and no differences in pre- or post-PSWQ scores based on having no comorbidity (all *p*'s > .05). However those with a comorbid mood disorder reported significant lower pre- and post-treatment PSWQ (t(139)=2.60, p=.010, and t(139)=3.56, p=.001), and lower post-treatment PSWQ-A scores (t(257)=2.80, p=.006), but not pre-treatment PSWQ-A scores (t(139)=.54, p=.590). Given that differential treatment efficacy based on comorbidity status was not a primary outcome of interest in this study, samples were combined for all analyses.

To examine change in symptoms from pre- to post-treatment based on diagnostic interviews; 56.78% of the sample were remitted from their Generalised Anxiety Disorder diagnosis at the end of treatment, with a significantly greater proportion of these from the CBT conditions (n=94, 69.63%) compared with the control conditions (n=53, 42.74%; $\chi(1)^2$=19.04, p<.001). Similarly, there was a significantly higher proportion of participants in the CBT conditions (n=79, 58.52%) compared with the control conditions (n=47, 37.90%) that were classified as treatment responders, $\chi(1)^2$=11.00, p=.001. Symptom change on PSWQ and PSWQ-A showed a similar pattern, with a significant decrease in PSWQ (t=4.25, p<.001) and PSWQ-A scores (t=9.71, p<.001) from pre- to post-treatment. Repeated measures ANOVA showed a significant time by condition interaction for PSWQ scores (F(1, 139)=16.47, p <.001), with those in the CBT conditions showing a significant reduction in PSWQ scores over treatment (M_{diff}=-7.71, SE=1.28, F(1,139)=36.28, p<.001, $\eta_p^{2=}$.21), but no significant change in PSWQ scores among those in the control conditions (M_{diff}=-.19, SE=1.47, F(1,139)=.02, p=.898, $\eta_p^{2=}$.02). Based on PSWQ-A scores, there was a significant time by condition interaction (F(1, 257)=15.77, p <.001) showing a greater reduction in

PSWQ-A scores among those in the CBT conditions (M_{diff}=6.50, SE=.65, F(1,257)=99.21, p <.001, $\eta_p^{2=}$.28) compared with the control condition (M_{diff}=-2.75, SE=.68, F(1,257)=16.37, p<.001, $\eta_p^{2=}$.06).

Prediction of Remission Status using the PSWQ

An ROC analysis assessed the predictive validity of scores on the PSWQ and PSWQ-A at post-treatment to identify remission status. The AUC was fair for both measures, .70 (SE=.05, p <.001) for the PSWQ and .74 (SE=.03, p<.001) for the PSWQ-A (see Figures 1a and 1b).

Table 2 summarises the metrics for predicting GAD remission status using raw score cut-off scores on the PSWQ. Results suggest that a cut-off score of ≤51 was the optimal cut-off score for defining remission, with a minimal level of agreement between the cut-off score and diagnostic criteria (kappa=.34). At this cut-off score, 60% of those who were remitted would correctly be identified as such (sensitivity=.60), 74% of those who were not remitted would be correctly classified as such (specificity=.74), 68% of those who would be classified as remitted based on PSWQ scores would also be classified as such based on clinical interviews (PPV=.68), and 67% of those who would be classified as not remitted based on PSWQ score would also be classified as such based on clinical interviews (NPV=.67). Although a cut-off score of ≤57 was also considered on the basis of higher kappa (.41), NPV (.76) and sensitivity (.78), the more conservative criteria of ≤51 was chosen on the basis of the higher specificity value and similar PPV (.66). Specificity and NPV were both chosen a priori as the prioritised statistics; however, the two cut-off scores differed in their superiority on each metric. Given the risk of premature treatment discontinuation in a client who is not in remission, optimising specificity was considered the priority. In comparison, the implication

of a lower NPV would be additional treatment in a client who was actually in remission. On balance, premature discontinuation was prioritised, along with the specificity metric.

Prediction of Remission Status using the PSWQ-A

The lower section of Table 2 summarises results for predicting GAD remission status using PSWQ-A cut-off scores. Results suggest an optimal cut-off score of ≤24, with this cut-off resulting in 69% agreement with diagnostic interviews, k=.38, sensitivity=.65, specificity=.73, PPV=.76 and NPV=.62. Cut-off scores of ≤27 and ≤28 were also considered due to their slightly higher kappa, sensitivity and NPV statistics; however, ≤24 was chosen as the optimal cut-off given higher specificity and PPV values for the same reasons noted above for the full PSWQ.

Prediction of Treatment Response

Four ROC analyses examined the predictive validity of changes in scores on the PSWQ and PSWQ-A to identify treatment response status, using percent reduction in scores from pre-treatment to post-treatment, as well as raw score reductions. The AUC was fair for percent reduction on the PWSQ, .70 (SE=.04, p <.001), but poor for percent reduction on the PSWQ-A (AUC=.65, SE=.05, p =.001; see Figures 1c and d). Similarly, the AUC was acceptable for predicting treatment response based on raw score reductions on the PSWQ (AUC=.70, SE=.04, p<.001) but poor for predicting response based on PSWQ-A raw sore reduction (AUC=.62, SE=.04, p=.007; see figures 1e and f).

The upper section of Table 3 summarises statistics for the percent reduction in PSWQ scores to predict treatment responder status. A 9% reduction in PSWQ score optimally predicted treatment response, with fair agreement with diagnostic assessments (k=.32), sensitivity=.59, specificity=.72, PPV=.69, NPV=.63 and agreement=.66. At the previously

used threshold of 20%, sensitivity=.42, specificity=.83, PPV=.67, NPV=.64, agreement=.64 and k=.25.

The ability for the percent reduction in PSWQ-A scores to predict treatment response is summarised in the lower section of Table 3. A 13% or 14% reduction were the best cut-off scores, however both showed suboptimal metrics, with 51-52% of treatment responders identified (sensitivity=.52 and .51, respectively), 74% of non-responders correctly identified (specificity=.74 for both cut-offs), PPV=.62 and .61, NPV=.66 and .65, agreement=.64 for both, and k=.27 and .26.

The upper section of Table 4 summarises prediction of treatment response based on a raw score reduction on the full PSWQ. The optimal cut-off was ≥4, which had sensitivity of .65, specificity of .64, PPV of .65, NPV of .64, agreement of .64 and k=.29. At the previously used threshold of ≥8.5, sensitivity=.51, specificity=.78, PPV=.71, NPV=.61, agreement=.64 and k=.29.

The lower section of Table 4 summarises the prediction of treatment response based on a raw score reduction on the PSWQ-A. Consistent with the poor AUC, there were no cut-off scores that were considered appropriate. The highest kappa value was .21 and even the highest sensitivity value (≥5: kappa=.46) indicated that fewer than 50% of those who had responded to treatment would be identified using this measure.

Discussion

Given the considerable variability in the definitions and use of the PSWQ to define treatment outcome in previous studies, this study aimed to develop empirically derived guidelines for defining treatment response and remission status among older adults with GAD using the PSWQ and PSWQ-A. Results suggest that a score of ≤51 was optimal for defining remission status on the PSWQ, and a score of ≤24 was optimal on the PSWQ-A. A 9% reduction or ≥4-

point reduction was optimal for assessing treatment response on the PSWQ. The PSWQ-A was poor at identifying treatment response status, and no guidelines are recommended.

The recommended thresholds for assessing remission status are similar to those recommended for screening purposes. A score of 51 or above on the PSWQ is recommended when screening for likely GAD (Stanley et al., 2003b). The recommendation of a threshold of ≤51 for assessing remission status is similar. The guideline of ≤24 for assessing remission status on the PSWQ-A is also similar to the threshold cut-off score of ≥23 and above for screening purposes (Stanley et al., 2003b). However this is notably different from thresholds used in previous studies assessing remission or high end state functioning that have used scores of ≤37 (Wetherell et al., 2003, Stanley et al., 1996), ≤36 (Stanley et al., 2003a), and ≤43 (Mohlman, 2008) on the PSWQ. Our results suggest that these conservative thresholds are likely to have underestimated the effects of treatment in older adults, with results from this study suggesting that these conservative thresholds would have detected less than 40% of remitted cases, but would have greater than 80% specificity. The use of diagnostic interviews to assess symptom change in this study is superior to the exclusive reliance on self-report measures in these previous studies, which can be affected by self-report biases.

The recommended thresholds for defining treatment response on the PSWQ (≥9% reduction or ≥4 point reduction) are much lower than thresholds used in previous studies, which have typically utilised a 8.5 point reduction (Stanley et al., 2009, Bradford et al., 2011, Wetherell et al., 2013) or 20% reduction (Wetherell et al., 2003, Stanley et al., 1996, Stanley et al., 2003c, Stanley et al., 2003a). Given that a 20% reduction in symptoms only detected 42% of those who were classified as treatment responders, and an 8.5 point reduction detected 51%, these previous thresholds are likely to have notably underestimated treatment effects when categorical outcome measures have been reported.

Despite adequately detecting remission status, the PSWQ-A was poor at identifying treatment response. Although this measure demonstrates superior factor structure to the full PSWQ and good ability to detect GAD status (Wuthrich et al., 2014, Hopko et al., 2003, Webb et al., 2008), results suggest that the PSWQ-A would be no better than chance at determining treatment response and should only be used to determine remission status when considering treatment discontinuation. Despite this, it is important to note that this is only relevant when treatment response is a categorical outcome (i.e., treatment responder or non-responder). The majority of research studies examine average group change in symptoms, and these results do not necessarily suggest problems when using these statistical methods. However, linear change metrics are only relevant when reporting results of research trials, given that beta-weights and p-values are of no use for evaluating treatment of an individual patient. In treatment contexts, dichotomous metrics for determining treatment effects are ideal to provide an objective measure of progress.

This study has several limitations to note. Even at optimal cut-offs, the PSWQ showed only fair to moderate ability to predict remission or response (kappa = .3-.4; AUC = .7). Although these are acceptable ranges (Landis and Koch, 1977), they are slightly lower than other signal detection studies (e.g., Johnco et al., 2015b, Johnco et al., 2015a, Caporino et al., 2013, Storch et al., 2011, Storch et al., 2010, Houghton et al., 2015, Skarphedinsson et al., 2017), and are relatively low in absolute terms. Overall, these results suggest that the PSWQ is suboptimal for evaluating treatment outcome in geriatric GAD. These findings highlight the need to consider whether geriatric-specific measures of anxiety (e.g., Geriatric Anxiety Scale/Geriatric Anxiety Inventory) may be superior for assessing treatment outcomes. However, using these empirically derived guidelines in conjunction with other measures may improve predictive capacity of the PSWQ. It is notable that most of the previous studies that have used the PSWQ in defining outcome status have used this as one of

three or four measures that are used in combination to assess outcome. While this may improve accuracy, the requirement to administer and score up to four self-report measures of the same symptom cluster risks increasing client and clinician burden, and may deter use of objective outcome measures in clinical practice. Combining data from multiple sites and countries improved power and generalisability of these findings; however, the inconsistent use of diagnostic interviews (e.g., ADIS vs. SCID) and the discrepant use of a 0-8 or 0-3 scale to rate disorder severity between studies was a limitation. Although the RCI used to define treatment responder status on diagnostic interviews was calculated separately for each of these, it would be advantageous for consistent use of severity ratings and diagnostic interviews. Similarly, the site differences in comorbidity profiles were reflective of different trial inclusion criteria (i.e., the requirement of comorbid depression at Macquarie University; Wuthrich and Rapee, 2013). While differential treatment outcomes by site or based on comorbidity status was not a primary outcome of interest in this study, results warrant replication in a large and more clinically homogenous sample with consistent measurement approaches.

The PSWQ is one of the most widely used self-report measures of worry severity used in treatment trials for late-life GAD. Despite this, the definitions and thresholds used to assess treatment outcome has varied between studies. Using a large sample of older adults who received treatment for late-life GAD, this study provides empirically-derived guidelines for defining remission and treatment response status. A score of ≤51 on the PSWQ and ≤24 on the PSWQ-A are recommended for defining remission. A 9% reduction or ≥4 point reduction on the PSWQ is recommended for defining treatment response, whereas the PSWQ-A was not suitable for assessing treatment response status. Findings suggest that most of the previously used definitions have underestimated the treatment effects for late-life GAD. The

standardisation of response and remission criteria is important for comparison between treatment trials, and for the benchmarking of outcomes in clinical practice.

Conflict of Interest Declaration: None.

Description of Authors' Roles: Carly Johnco designed the study, collected and analysed data and wrote the manuscript. Viviana Wuthrich collected data, assisted with data interpretation, and provided feedback on the manuscript. Gretchen Brenes, Julie Wetherell and Jan Mohlman collected data and provided feedback on the manuscript.

Acknowledgements: None.

References

Barrera, T. L., Cully, J. A., Amspoker, A. B., Wilson, N. L., Kraus-Schuman, C., Wagener, P. D., et al. 2015. Cognitive-behavioral therapy for late-life anxiety: Similarities and differences between Veteran and community participants. *Journal of anxiety disorders,* 33**,** 72-80.

Beck, J. G., Stanley, M. A. & Zebb, B. J. 1995. Psychometric properties of the Penn State Worry Questionnaire in older adults. *Journal of Clinical Geropsychology,* 1**,** 33-42.

Beekman, A. T., Bremmer, M. A., Deeg, D. J., van Balkom, A. J., Smit, J. H., de Beurs, E., et al. 1998. Anxiety disorders in later life: a report from the Longitudinal Aging Study Amsterdam. *International journal of geriatric psychiatry,* 13**,** 717-726.

Bradford, A., Cully, J., Rhoades, H., Kunik, M., Kraus-Schuman, C., Wilson, N., et al. 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 : official journal of the American Association for Geriatric Psychiatry,* 19**,** 347-356.

Brenes, G. A., Danhauer, S. C., Lyles, M. F., Hogan, P. E. & Miller, M. E. 2015. Telephone-Delivered Cognitive Behavioral Therapy and Telephone-Delivered Nondirective Supportive Therapy for Rural Older Adults With Generalized Anxiety Disorder: A Randomized Clinical Trial. *JAMA Psychiatry,* 72**,** 1012-20.

Caporino, N. E., Brodman, D. M., Kendall, P. C., Albano, A. M., Sherrill, J., Piacentini, J., et al. 2013. Defining treatment response and remission in child anxiety: signal detection analysis using the pediatric anxiety rating scale. *Journal of the American Academy of Child and Adolescent Psychiatry,* 52**,** 57-67.

Craske, M. G., Barlow, D. H. & O'Leary, T. A. 1992. *Mastery of your anxiety and worry,* Albany, NY, Graywind Publications.

Di Nardo, P. A., Brown, T. A. & Barlow, D. H. 1994. *Anxiety Disorders Interview Schedule for DSM-IV,* Boston, Center for Stress and Anxiety Related Disorders,Boston University.

First, M. B., Spitzer, R. L., Miriam, G. & Williams, J. B. W. 2002. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders,* New York, Biometrics Research, New York State Psychiatric Institute,.

Gonçalves, D. C. & Byrne, G. J. 2012. Interventions for generalized anxiety disorder in older adults: Systematic review and meta-analysis. *Journal of Anxiety Disorders,* 26**,** 1-11.

Gould, R. L., Coulson, M. C. & Howard, R. J. 2012. Efficacy of Cognitive Behavioral Therapy for Anxiety Disorders in Older People: A Meta-Analysis and Meta-Regression of Randomized Controlled Trials. *Journal of the American Geriatrics Society,* 60**,** 218-229.

Hall, J., Kellett, S., Berrios, R., Bains, M. K. & Scott, S. 2016. Efficacy of Cognitive Behavioral Therapy for Generalized Anxiety Disorder in Older Adults: Systematic Review, Meta-Analysis, and Meta-Regression. *Am J Geriatr Psychiatry,* 24**,** 1063-1073.

Hendriks, G. J., Oude Voshaar, R. C., Keijsers, G. P. J., Hoogduin, C. A. L. & Van Balkom, A. J. L. M. 2008. Cognitive-behavioural therapy for late-life anxiety disorders: a systematic review and meta-analysis. *Acta Psychiatrica Scandinavica,* 117**,** 403-411.

Hopko, D. R., Stanley, M. A., Reas, D. L., Wetherell, J. L., Beck, J. G., Novy, D. M., et al. 2003. Assessing worry in older adults: confirmatory factor analysis of the Penn State Worry Questionnaire and psychometric properties of an abbreviated model. *Psychol Assess,* 15**,** 173-83.

Houghton, D. C., Capriotti, M. R., De Nadai, A. S., Compton, S. N., Twohig, M. P., Neal-Barnett, A. M., et al. 2015. Defining treatment response in trichotillomania: a signal detection analysis. *J Anxiety Disord,* 36**,** 44-51.

Jacobson, N. S. & Truax, P. 1991. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology,* 59**,** 12-19.

Johnco, C. J., De Nadai, A. S., Lewin, A. B., Ehrenreich-May, J., Wood, J. J. & Storch, E. A. 2015a. Defining treatment response and symptom remission for anxiety disorders in pediatric autism spectrum disorders using the Pediatric Anxiety Rating Scale. *Journal of autism and developmental disorders,* 45**,** 3232-3242.

Johnco, C. J., Salloum, A., Lewin, A. B. & Storch, E. A. 2015b. Refining Clinical Judgment of Treatment Response and Symptom Remission Identification in Childhood Anxiety Using a Signal Detection Analysis on the Pediatric Anxiety Rating Scale. *Journal of child and adolescent psychopharmacology,* 25**,** 674-683.

Johnco, C. J., Wuthrich, V. M. & Rapee, R. M. 2014. The influence of cognitive flexibility on treatment outcome and cognitive restructuring skill acquisition during cognitive behavioural treatment for anxiety and depression in older adults: Results of a pilot study. *Behav Res Ther,* 57**,** 55-64.

Landis, J. R. & Koch, G. G. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics,* 33**,** 159-174.

Mandrekar, J. N. 2010. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology,* 5**,** 1315-1316.

Meyer, T. J., Miller, M. L., Metzger, R. L. & Borkovec, T. D. 1990. Development and validation of the Penn State Worry Questionnaire. *Behav Res Ther,* 28**,** 487-95.

Mohlman, J. 2008. More Power to the Executive? A Preliminary Test of CBT Plus Executive Skills Training for Treatment of Late-Life GAD. *Cognitive and Behavioral Practice,* 15**,** 306-316.

Mohlman, J. 2020. Neurocognitive Predictors of Long-Term Outcome in CBT for Late Life Generalized Anxiety Disorder. *Journal of Anxiety Disorders,* 74**,** 102246.

Mohlman, J., Gorenstein, E. E., Kleber, M., de Jesus, M., Gorman, J. M. & Papp, L. A. 2003. Standard and enhanced cognitive-behavior therapy for late-life generalized anxiety disorder: two pilot investigations. *Am J Geriatr Psychiatry,* 11**,** 24-32.

Mohlman, J. & Gorman, J. M. 2005. The role of executive functioning in CBT: a pilot study with anxious older adults. *Behav Res Ther,* 43**,** 447-65.

Pachana, N. A., Byrne, G. J., Siddle, H., Koloski, N., Harley, E. & Arnold, E. 2007. Development and validation of the Geriatric Anxiety Inventory. *International Psychogeriatrics,* 19**,** 103-114.

Pinquart, M. & Duberstein, P. R. 2007. Treatment of Anxiety Disorders in Older Adults: A Meta-analytic Comparison of Behavioral and Pharmacological Interventions. *The American Journal of Geriatric Psychiatry,* 15**,** 639-651.

Roseman, A. S., Cully, J. A., Kunik, M. E., Novy, D. M., Rhoades, H. M., Wilson, N. L., et al. 2011. Treatment response for late-life generalized anxiety disorder: moving beyond symptom-based measures. *J Nerv Ment Dis,* 199**,** 811-4.

Segal, D. L., June, A., Payne, M., Coolidge, F. L. & Yochim, B. 2010. Development and initial validation of a self-report assessment tool for anxiety among older adults: The Geriatric Anxiety Scale. *Journal of Anxiety Disorders,* 24**,** 709-714.

Skarphedinsson, G., De Nadai, A. S., Storch, E. A., Lewin, A. B. & Ivarsson, T. 2017. Defining cognitive-behavior therapy response and remission in pediatric OCD: a

signal detection analysis of the Children's Yale-Brown Obsessive Compulsive Scale. *European child & adolescent psychiatry,* 26**,** 47-55.

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.

Stanley, M. A., Beck, J. G., Novy, D. M., Averill, P. M., Swann, A. C., Diefenbach, G. J., et al. 2003a. Cognitive-behavioral treatment of late-life generalized anxiety disorder. *J Consult Clin Psychol,* 71**,** 309-19.

Stanley, M. A., Diefenbach, G. J., Hopko, D. R., Novy, D. M., Kunik, M. E., Wilson, N. L., et al. 2003b. The Nature of Generalized Anxiety in Older Primary Care Patients: Preliminary Findings. *Journal of Psychopathology and Behavioral Assessment,* 25**,** 273-280.

Stanley, M. A., Hopko, D. R., Diefenbach, G. J., Bourland, S. L., Rodriguez, H. & Wagener, P. 2003c. Cognitive-behavior therapy for late-life generalized anxiety disorder in primary care: preliminary findings. *Am J Geriatr Psychiatry,* 11**,** 92-6.

Stanley, M. A., Novy, D. M., Bourland, S. L., Beck, J. G. & Averill, P. M. 2001. Assessing older adults with generalized anxiety: a replication and extension. *Behaviour Research and Therapy,* 39**,** 221-235.

Stanley, M. A., Wilson, N. L., Amspoker, A. B., Kraus-Schuman, C., Wagener, P. D., Calleo, J. S., et al. 2014. Lay providers can deliver effective cognitive behavior therapy for older adults with generalized anxiety disorder: a randomized trial. *Depress Anxiety,* 31**,** 391-401.

Stanley, M. A., Wilson, N. L., Novy, D. M., Rhoades, H. M., Wagener, P. D., Greisinger, A. J., et al. 2009. Cognitive behavior therapy for generalized anxiety disorder among older adults in primary care: a randomized clinical trial. *JAMA,* 301**,** 1460-1467.

Storch, E. A., De Nadai, A. S., Lewin, A. B., McGuire, J. F., Jones, A. M., Mutch, P. J., et al. 2011. Defining treatment response in pediatric tic disorders: a signal detection analysis of the Yale Global Tic Severity Scale. *J Child Adolesc Psychopharmacol,* 21**,** 621-7.

Storch, E. A., Lewin, A. B., De Nadai, A. S. & Murphy, T. K. 2010. Defining treatment response and remission in obsessive-compulsive disorder: a signal detection analysis of the Children's Yale-Brown Obsessive Compulsive Scale. *J Am Acad Child Adolesc Psychiatry,* 49**,** 708-17.

Thorp, S. R., Ayers, C. R., Nuevo, R., Stoddard, J. A., Sorrell, J. T. & Wetherell, J. L. 2009. Meta-Analysis Comparing Different Behavioral Treatments for Late-Life Anxiety. *The American Journal of Geriatric Psychiatry,* 17**,** 105-115.

Webb, S. A., Diefenbach, G., Wagener, P., Novy, D. M., Kunik, M., Rhoades, H. M., et al. 2008. Comparison of Self-report Measures for Identifying Late-life Generalized Anxiety in Primary Care. *Journal of Geriatric Psychiatry and Neurology,* 21**,** 223-231.

Wetherell, J. L., Gatz, M. & Craske, M. G. 2003. Treatment of generalized anxiety disorder in older adults. *J Consult Clin Psychol,* 71**,** 31-40.

Wetherell, J. L., Hopko, D. R., Diefenbach, G. J., Averill, P. M., Beck, J. G., Craske, M. G., et al. 2005. Cognitive-behavioral therapy for late-life generalized anxiety disorder: Who gets better? *Behavior Therapy,* 36**,** 147-156.

Wetherell, J. L., Petkus, A. J., White, K. S., Nguyen, H., Kornblith, S., Andreescu, C., et al. 2013. Antidepressant medication augmented with cognitive-behavioral therapy for generalized anxiety disorder in older adults. *The American journal of psychiatry,* 170**,** 782-789.

Wolitzky-Taylor, K. B., Castriotta, N., Lenze, E. J., Stanley, M. A. & Craske, M. G. 2010. Anxiety disorders in older adults: a comprehensive review. *Depression and Anxiety,* 27**,** 190-211.

Wuthrich, V. M. 2009. *Ageing Wisely: Group Program for the treatment of worry and low mood in older adults,* Sydney, Australia., Centre for Emotional Health, Department of Psychology, Macquarie University.

Wuthrich, V. M., Johnco, C. & Knight, A. 2014. Comparison of the Penn State Worry Questionnaire (PSWQ) and abbreviated version (PSWQ-A) in a clinical and non-clinical population of older adults. *J Anxiety Disord,* 28**,** 657-63.

Wuthrich, V. M. & Rapee, R. M. 2013. Randomised controlled trial of group cognitive behavioural therapy for comorbid anxiety and depression in older adults. *Behaviour Research and Therapy,* 51**,** 779-786.

Table 1. Sample demographic and clinical characteristics.

	Total sample (N=259)	**Macquarie University (n=57)**	**UCSD (n=48)**	**William Paterson (n=36)**	**Wake Forest (n=118)**	
	M (SD)	***M (SD)***	***M (SD)***	***M (SD)***	***M (SD)***	***F***
Age	67.22 (5.75)	67.39 (5.34)	69.21 (6.68)	66.69 (4.91)	66.49 (5.65)	3.09
Pre-treatment PSWQ	58.17 (10.65)	54.98 (10.83)	63.27 (10.31)	56.44 (9.24)	-	9.57***
Post-treatment PSWQ	53.88 (12.64)	49.23 (11.94)	60.80 (11.67)	52.03 (11.13)	-	13.48***
Pre-treatment PSWQ-A	29.71 (6.40)	26.40 (7.14)	31.74 (7.00)	28.14 (4.90)	30.97 (5.43)	9.88***
Post-treatment PSWQ-A	25.01 (7.78)	23.10 (7.05)	31.15 (6.36)	24.97 (7.34)	23.44 (7.59)	14.75***
	n (%)	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**χ^2**
Male	67 (25.87%)	24 (42.11%)	10 (20.83%)	13 (36.11%)	20 (16.95%)	15.34**
Employment Status						7.44
Working (full-time/part-time)	79 (30.38%)	19 (32.76%)	10 (20.83%)	17 (47.22%)	33 (27.97%)	
Not working/retired	180 (69.50%)	38 (66.67%)	38 (79.17%)	19 (52.78%)	85 (72.03%)	
Comorbidity						
None	47 (18.08%)	0 (0.00%)	13 (27.08%)	17 (47.22%)	17 (14.41%)	36.82***
Other anxiety disorder	108 (41.54%)	30 (52.63%)	13 (27.08%)	12 (33.33%)	53 (44.92%)	8.56*
Mood disorder	161 (61.82%)	53 (92.98%)	8 (16.67%)	12 (33.33%)	88 (74.58%)	85.71***
PTSD	19 (7.31%)	2 (3.51%)	3 (6.25%)	2 (5.56%)	12 (10.17%)	2.87
OCD	11 (4.23%)	2 (3.51%)	4 (8.33%)	0 (0.00%)	5 (4.24%)	3.64
Substance Use disorder	14 (5.38%)	4 (7.02%)	0 (0.00%)	1 (2.78%)	9 (7.93%)	4.66
Treatment						26.67***
CBT	135 (51.92%)	45 (78.95%)	16 (33.33%)	19 (52.78%)	55 (46.61%)	
Control (Waitlist, non-directive supportive therapy, discussion group).	124 (47.69%)	12 (21.05%)	32 (66.66%)	17 (47.22%)	63 (53.39%)	
Remission[a]	147 (56.76%)	27 (47.37%)	24 (50.00%)	16 (44.44%)	80 (67.80%)	11.02*
Responder[b]	126 (48.65%)	21 (36.84%)	13 (27.08%)	15 (41.67%)	77 (65.25%)	25.84***

PSWQ = Penn State Worry Questionnaire; PSWQ-A = Penn State Worry Questionnaire – Abbreviated. [a]Remission was defined as loss of GAD diagnosis on diagnostic interviews. [b]Treatment responder status was defined based on a reliable change index of GAD diagnosis severity ratings from diagnostic interviews.

Table 2. Prediction of clinical remission based on loss of Generalised Anxiety Disorder diagnosis using the PSWQ.

PSWQ Cut-off Score	Sensitivity	Specificity	PPV	NPV	Agreement	*k*
≤45	0.39	0.85	0.70	0.61	0.63	0.25
≤46	0.39	0.82	0.67	0.60	0.62	0.22
≤47	0.43	0.77	0.63	0.60	0.61	0.21
≤48	0.46	0.76	0.63	0.6	0.62	0.22
≤49	0.51	0.75	0.65	0.63	0.63	0.25
≤50	0.55	0.74	0.66	0.65	0.65	0.30
≤51	0.60	0.74	0.68	0.67	0.67	0.34
≤52	0.60	0.73	0.67	0.67	0.67	0.33
≤53	0.61	0.69	0.64	0.66	0.65	0.30
≤54	0.63	0.65	0.62	0.66	0.64	0.28
≤55	0.64	0.65	0.62	0.67	0.65	0.29
≤56	0.70	0.64	0.64	0.70	0.67	0.34
≤57	0.78	0.64	0.66	0.76	0.70	0.41
≤58	0.79	0.57	0.62	0.75	0.67	0.35
≤59	0.84	0.50	0.60	0.77	0.66	0.33
≤60	0.87	0.46	0.59	0.79	0.65	0.32
PSWQ-A Cut-off Score	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Agreement**	***K***
≤20	0.45	0.86	0.80	0.54	0.63	0.29
≤21	0.46	0.86	0.81	0.55	0.63	0.30
≤22	0.52	0.82	0.79	0.57	0.65	0.33
≤23	0.56	0.76	0.75	0.57	0.65	0.31
≤24	0.65	0.73	0.76	0.62	0.69	0.38
≤25	0.67	0.68	0.73	0.61	0.68	0.35
≤26	0.71	0.63	0.72	0.63	0.68	0.35
≤27	0.77	0.61	0.72	0.67	0.70	0.38
≤28	0.82	0.56	0.71	0.71	0.71	0.40
≤29	0.85	0.51	0.69	0.72	0.70	0.37
≤30	0.88	0.46	0.68	0.75	0.70	0.37
≤31	0.91	0.39	0.66	0.77	0.69	0.32
≤32	0.92	0.32	0.64	0.75	0.66	0.26
≤33	0.93	0.29	0.63	0.74	0.65	0.23
≤34	0.96	0.24	0.62	0.82	0.65	0.22
≤35	0.97	0.21	0.61	0.82	0.64	0.19

K = kappa, NPV = Negative Predictive Value, PPV = Positive Predictive Value, PSWQ = Penn State Worry Questionnaire; PSWQ-A = Penn State Worry Questionnaire – Abbreviated

Table 3. Prediction of treatment response based on reliable change in GAD diagnostic severity from percent reduction in PSWQ score from pre to post treatment.

% reduction in PSWQ Score	Sensitivity	Specificity	PPV	NPV	Agreement	*K*
≥5%	0.66	0.58	0.62	0.63	0.62	0.24
≥6%	0.65	0.62	0.64	0.63	0.64	0.27
≥7%	0.63	0.65	0.65	0.63	0.64	0.29
≥8%	0.61	0.68	0.66	0.63	0.64	0.29
≥9%	0.59	0.72	0.69	0.63	0.66	0.32
≥10%	0.56	0.74	0.69	0.62	0.65	0.30
≥11%	0.52	0.78	0.71	0.61	0.65	0.30
≥12%	0.49	0.78	0.70	0.60	0.64	0.27
≥13%	0.48	0.80	0.71	0.60	0.64	0.28
≥14%	0.45	0.80	0.70	0.59	0.62	0.25
≥15%	0.45	0.81	0.71	0.59	0.63	0.26
≥16%	0.49	0.76	0.62	0.65	0.64	0.25
≥17%	0.48	0.78	0.63	0.65	0.64	0.26
≥18%	0.48	0.81	0.67	0.66	0.66	0.29
≥19%	0.43	0.83	0.67	0.64	0.65	0.27
≥20%	0.42	0.83	0.67	0.64	0.64	0.25
% reduction in PSWQ-A Score	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Agreement**	***k***
≥10%	0.57	0.64	0.56	0.65	0.61	0.22
≥11%	0.54	0.67	0.57	0.64	0.61	0.21
≥12%	0.52	0.70	0.59	0.64	0.62	0.23
≥13%	0.52	0.74	0.62	0.66	0.64	0.27
≥14%	0.51	0.74	0.61	0.65	0.64	0.26
≥15%	0.50	0.74	0.61	0.65	0.63	0.25
≥16%	0.49	0.76	0.62	0.65	0.64	0.25
≥17%	0.48	0.78	0.63	0.65	0.64	0.26
≥18%	0.48	0.81	0.67	0.66	0.66	0.29
≥19%	0.43	0.83	0.67	0.64	0.65	0.27
≥20%	0.42	0.83	0.66	0.64	0.64	0.25

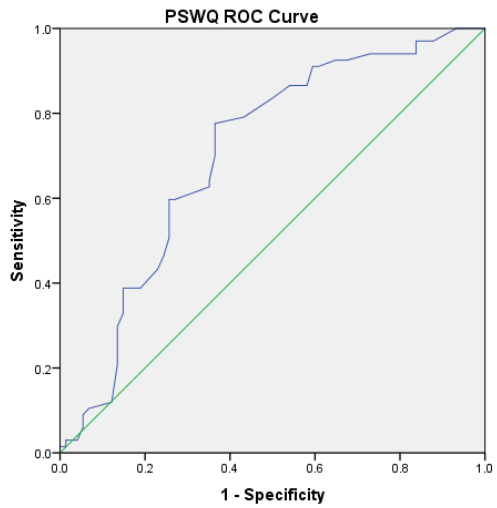
K = kappa, NPV = Negative Predictive Value, PPV = Positive Predictive Value, PSWQ = Penn State Worry Questionnaire; PSWQ-A = Penn State Worry Questionnaire – Abbreviated

Table 4. Prediction of treatment response based on reliable change in GAD diagnostic severity from raw score reduction in PSWQ score from pre to post treatment.

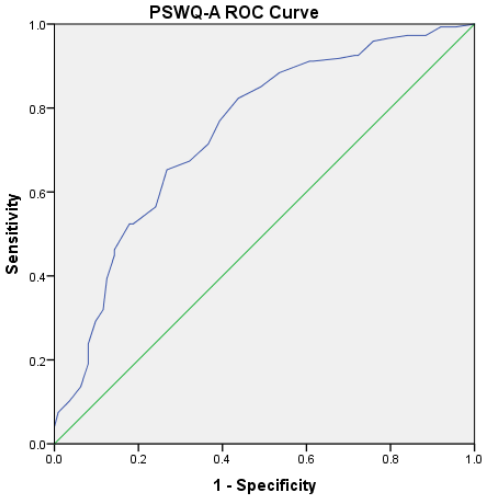
Raw score reduction in PSWQ Score	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Agreement**	***k***
≥1	0.73	0.51	0.60	0.65	0.62	0.24
≥2	0.72	0.55	0.62	0.66	0.64	0.27
≥3	0.68	0.58	0.62	0.63	0.63	0.26
≥4	0.65	0.64	0.65	0.64	0.64	0.29
≥5	0.59	0.68	0.66	0.62	0.64	0.27
≥6	0.54	0.70	0.64	0.59	0.61	0.23
≥7	0.51	0.72	0.65	0.59	0.61	0.23
≥8	0.51	0.78	0.71	0.61	0.64	0.29
≥9	0.45	0.81	0.71	0.59	0.63	0.26
≥10	0.42	0.81	0.70	0.58	0.61	0.23
≥11	0.38	0.84	0.71	0.57	0.61	0.22
Raw score reduction in PSWQ-A Score	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Agreement**	***k***
≥1	0.65	0.43	0.48	0.61	0.53	0.09
≥2	0.62	0.54	0.52	0.64	0.57	0.16
≥3	0.57	0.62	0.55	0.64	0.59	0.19
≥4	0.52	0.64	0.54	0.63	0.59	0.17
≥5	0.46	0.74	0.59	0.63	0.62	0.21
≥6	0.39	0.78	0.59	0.61	0.61	0.18
≥7	0.33	0.82	0.60	0.60	0.60	0.16
≥8	0.32	0.88	0.68	0.61	0.63	0.21
≥9	0.23	0.92	0.70	0.60	0.61	0.16
≥10	0.23	0.92	0.70	0.60	0.61	0.16
≥11	0.21	0.93	0.72	0.60	0.61	0.16

K = kappa, NPV = Negative Predictive Value, PPV = Positive Predictive Value, PSWQ = Penn State Worry Questionnaire; PSWQ-A = Penn State Worry Questionnaire – Abbreviated

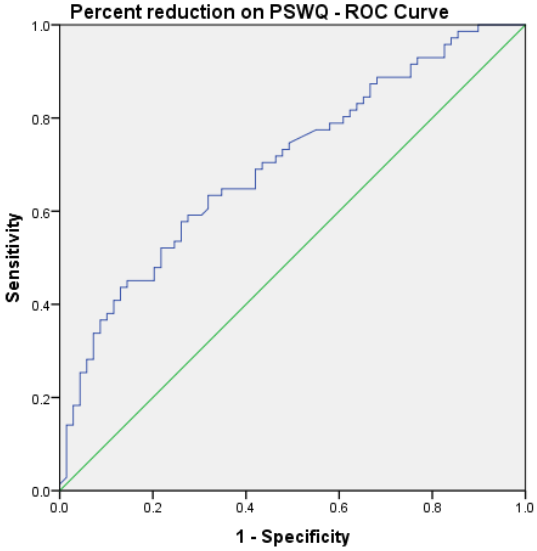
Figure 1. ROC curves for predicting remission from Generalised Anxiety Disorder for PSWQ (a) and PSWQ-A (b). ROC curve for predicting treatment response based on percent reduction in scores on PSWQ (c) and PSWQ-A (d) and raw score reduction in scores on PSWQ (e) and PSWQ-A (f).


PSWQ ROC Curve
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

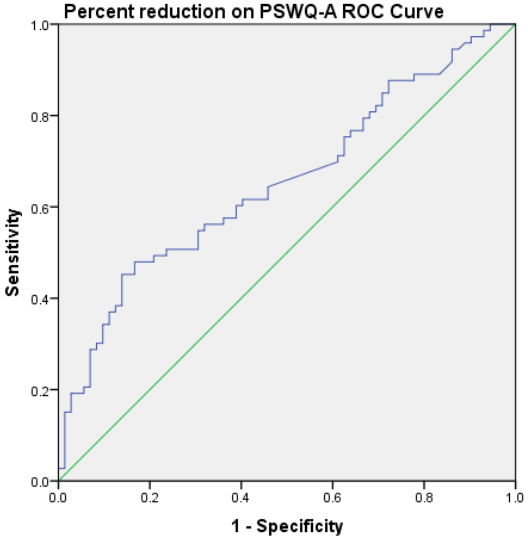
a)


PSWQ-A ROC Curve
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

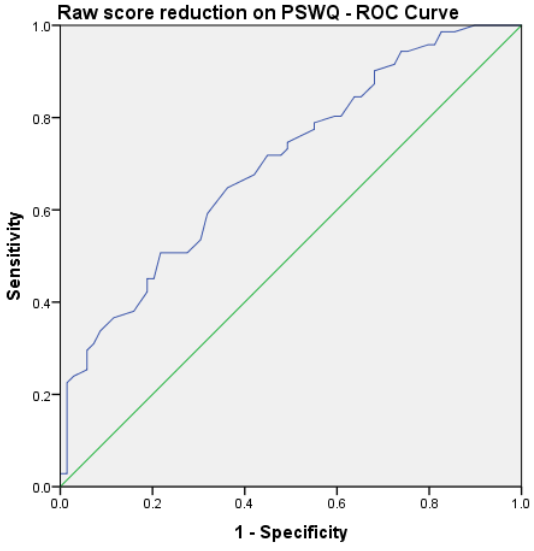
b)


Percent reduction on PSWQ - ROC Curve
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

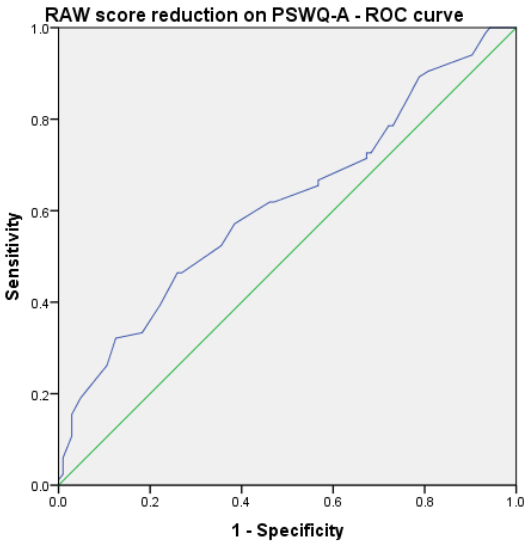
c)


Percent reduction on PSWQ-A ROC Curve
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

d)


Raw score reduction on PSWQ - ROC Curve
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

e)


RAW score reduction on PSWQ-A - ROC curve
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

f)