Universiteit Utrecht

ROM in anxiety disorders

Diagnosis-specific and generic assessment instruments

Malindi van der Mheen

Supervised by prof. dr. M.A. van den Hout

ROM in anxiety disorders: Diagnosis-specific and generic assessment instruments

Masterthesis

Clinical and Health Psychology

by

Malindi van der Mheen
3645878

Supervised by prof. dr. M.A. van den Hout

July, 2014

Table of Contents

Abstract

This study examined the sensitivity of the assessment of treatment progress and outcome in anxiety disorders through diagnosis-specific versus generic routine outcome monitoring (ROM) instruments. Adults (N = 160) diagnosed with obsessive-compulsive disorder (OCD), posttraumatic stress disorder (PTSD), panic disorder with (PDA) or without agoraphobia (PD), or social anxiety disorder (SAD) received outpatient treatment at the Altrecht Academic Anxiety Centre. Each patients' treatment outcome was assessed through two generic assessment instruments and two diagnosis-specific assessment instruments. The results show a subtle pattern, such that the differences between pre- and post-treatment assessments appeared to be larger for diagnosis-specific assessment instruments and the generic Brief Symptom Inventory (BSI) than for the generic Outcome Questionnaire (OQ-45). Implications for the current findings are discussed.

Preface

This thesis is written in the context of the final research of the master's degree of Clinical and Health Psychology at Utrecht University. Without the valuable contribution of some people I would not have been able to complete this project. I would like to thank Prof. Dr. M.A. van den Hout, my supervisor, for his enthusiasm and for providing critical comments on my work. Furthermore, I would like to thank the Altrecht Adacemic Anxiety Centre, particularly Mathilde Huisman, for allowing me to use their database and for providing statistical advice. Finally, I would like to thank you, my reader, for showing interest in my thesis.

Introduction

Anxiety disorders are among the most common mental disorders, are associated with an early age of onset, and are often chronic (Kessler et al., 2005). Anxiety disorders are considered as a group of related but distinct mental disorders which are characterized by the experience of extreme anxiety disproportionate to the situation at hand (Olatunji, Cisler, & Daecon, 2010). Anxiety is a response to a signal that warns a person of expected danger, as opposed to fear, which is a response to an immediate external threat (Gray & McNaughton, 2003). Symptoms of anxiety are frequently reported among the population and can be functional by warning an individual of emergencies and subsequently preparing the organism to act accordingly (Clark & Beck, 2010). However, an individual diagnosed with an anxiety disorder reports experiencing excessive amounts of irrational anxiety which substantially impairs the individual's daily functioning (Rachman, 2013). The diagnostic and statistical manual of mental disorders (DSM-IV-TR; American Psychiatric Association [APA], 2000) distinguishes several types of anxiety disorders. The present study relates to obsessive-compulsive disorder (OCD), panic disorder with (PDA) or without agoraphobia (PD), posttraumatic stress disorder (PTSD), and social anxiety disorder (SAD).

In order to treat anxiety disorders, cognitive-behavioral therapy (CBT) is frequently used (Hofmann & Smits, 2008). CBT encloses interventions that are based on the assumption that mental disorders, such as anxiety disorders, are maintained by cognitive and behavioral factors. Therefore, CBT attempts to treat anxiety disorders through the use of psychological treatment consisting of cognitive and behavioral techniques (Beck, 2011). CBT's efficacy in the treatment of anxiety disorders has been well established (e.g., Olatunji et al., 2010; Hofmann & Smits, 2008). Aside from CBT, pharmacotherapy may be used in the treatment of anxiety disorders. The use of selective serotonin reuptake inhibitors (SSRI's) is recommended (Trimbos Instituut, 2013). SSRI's inhibit serotonin reuptake, which results in greater levels of serotonin within synapses. In turn, this leads to the inhibition of neuronal activity in certain brain structures, such as the amygdalae, which are involved in the pathology of anxiety disorders (Prus, 2013).

The effectiveness of such therapeutic interventions can be determined through the use of assessment instruments (e.g., Noom et al., 2012; Hermans, Eelen, & Orlemans, 2007). Determining the effectiveness enables clinicians to reflect on and improve the mental health care they offer (Laane & Luijk, 2012; Walburg & Brinkmann, 2001). A patient's scores on assessment measures may aid a clinician in the decision-making with regard to the patient's

further treatment (Wiger & Solberg, 2001). For example, the use of assessment instruments enables clinicians to identify poorly responding patients and alter the treatment the patient is receiving (Carlier et al., 2012). Moreover, it enables mental health care institutions to make group-level comparisons of treatment outcome and treatment progress both within and between institutions (Hermann, 2005). Also, the use of assessment instruments allows for scientific research to examine the effectiveness of therapeutic interventions (Gilbody, House, & Sheldon, 2002).

In order to systematically monitor treatment progress, mental health institutions increasingly apply routine outcome monitoring (ROM) instruments (Nugter & Buwalda, 2012). In ROM, clinical assessment measures are routinely applied in order to measure both treatment progress and treatment outcome (Bilsker & Goldner, 2002). Interest in such assessment methods has grown as a consequence of changes in the healthcare insurance system (Hoenders et al., 2013; Miller, Duncan, Sorrell, & Brown, 2005). Healthcare insurance companies demand information on the quality of the treatments they are paying for. ROM provides healthcare insurance companies with the opportunity to gain insight in the ambiguous 'black box' of mental health care institutions (Miller et al., 2005). Furthermore, clinical interest in ROM has increased as research has shown that monitoring treatment progress can improve treatment outcome (Lambert, 2007; Miller et al., 2005).

Nugter and Buwalda (2012) state that scientific interest in ROM has primarily increased due to the realization that the results of randomized controlled trials (RCTs) have limited external validity. That is, in order to obtain sufficient internal validity, RCTs require that strict experimental conditions are met. However, these conditions are not consistent with those in practice. Therefore, results acquired through RCTs appear to lack generalizability (Nugter & Buwalda, 2012; Gilbody et al., 2002). Schat and colleagues (2013) reason that the use of extensive exclusion criteria and the focus on a strictly specified patient group limit the generalizability of research findings, such as those of RCTs. Conclusions based on assessments acquired through ROM are thought to be more generalizable than the results of RCTs. According to Nugter and Buwalda (2012) increasing evidence exists that the findings from RCTs are more favorable than treatment results in clinical practice. Shafran and colleagues (2009), on the other hand, argue that "the gap between clinical practice and research trials may not be as wide as many perceive" (p. 903). Increasing research has shown that applying evidence-based treatments to clinical practice settings yields similar results to those in RCTs (e.g., Franklin, Abramowitz, Kozak, Levitt, & Foa, 2000; Stuart, Treat, & Wade, 2000). Moreover, many contemporary research trials use few exclusion criteria (Lambert, 2013). The

exclusion criteria which are used resemble those that clinical institutions apply to ascertain the effects of a psychological treatment (Lambert, 2013; Shafran et al., 2009). Therefore, it can be questioned whether conclusions based on assessments acquired through ROM are more generalizable than the results of RCTs.

Both diagnosis-specific and generic instruments (e.g., Hoyer et al., 2002) can be applied in ROM (Buwalda et al., 2012). Diagnosis-specific instruments purport to assess diagnosis-specific, mainly axis-I, symptomatology (Luteijn et al., 2011). Diagnosis-specific instruments can be applied in order to measure progress in specific symptoms within diagnostically homogenous groups (Nugter & Buwalda, 2012). Such instruments are commonly used in clinical trials (e.g., Hofmann et al., 2012). Generic assessment instruments, on the other hand, measure general functioning and well-being (Patrick & Deyo, 1989). The use of generic assessment instruments enables comparisons of treatment progress and outcome between different diagnostic groups. Also, individual results on a generic assessment instrument may be compared to group-level results (Nugter & Buwalda, 2012). However, the distinction between diagnosis-specific and generic assessment instruments should not be seen as a strict dichotomy as instruments may differ in the degree to which they assess axis-I symptomatology. Wiger and Solberg (2001) recommend the use of an efficient combination of both types of ROM-instruments. In this way, all areas of possible improvement or regression are assessed.

Some critical remarks with regard to the use of ROM should be made. Firstly, ROM is labor intensive and expensive (Carlier et al., 2012). Furthermore, conclusions based on ROM should be drawn with prudence. As noted above, ROM enables healthcare insurance companies to gain insight in the previously ambiguous world of mental health care institutions. However, the awareness of the fact that assessment results will be available for third parties may influence the manner in which both clinicians and patients fill out assessment instruments. That is, clinicians may report more favorably about their patients' treatment progress than would be objectively justifiable (Bilsker & Goldner, 2002). Patients may under-report their symptoms in order to please their treatment provider or, on the other hand, may over-report their symptoms in order to stay in therapy (Bilsker & Goldner, 2002). Due to this possible reporter bias, the accuracy of the assessment results can be questioned. Moreover, the application of different assessment instruments may lead to a detection bias as the sensitivity to change differs across instruments (Hoenders et al., 2013). It should be noted, however, that a number of trials have failed to provide evidence for both the reporter bias and the detection bias in ROM (e.g., MacDonald & Trauer, 2010). Finally, factors other than the outcome score

retrieved through ROM should be taken into consideration when evaluating the effectiveness of an intervention. For example, when comparisons between institutions are made, specific characteristics of the patients for whom the institutions offer care, such as prognosis and comorbidity, should be taken into account.

The current study examines the sensitivity of the assessment of treatment progress and outcome in anxiety disorders through diagnosis-specific versus generic instruments. To the author's knowledge, no such research has been previously conducted. As CBT targets the specific anxiety disorder(s) for which a patient seeks treatment, it is hypothesized that treatment progress is more accurately represented by diagnosis-specific assessment measures than by generic assessment measures. Thus, the differences between pre- and post-treatment assessments are predicted to be larger for diagnosis-specific instruments than for generic instruments. In order to test this hypothesis we took the opportunity to use the database of the Altrecht Academic Anxiety Centre. First, treatment progress as assessed by the OQ-45, BSI, and merged diagnosis-specific instruments was compared through the use of (averaged) standardized scores. Second, main diagnosis was taken into account and the analysis was repeated. Finally, treatment progress as assessed by the OQ-45, BSI, and separate diagnosis-specific instruments was compared per diagnostic group.

Method

Participants

The original sample consisted of 172 patients. Of this sample, 12 patients were excluded because at least one pre- or post-treatment assessment was incomplete. The final sample consisted of 160 patients ranging in age from 18 to 59 years (M = 34.9, SD = 10.0), 97 (60.6%) females and 63 (39.4%) males. As all assessment instruments were completed in Dutch, patients selected in this study had to have adequate command of the Dutch language. Moreover, patients were required to meet the criteria specified in the DSM-IV-TR (American Psychiatric Association [APA], 2000) for one or more of the following conditions: obsessive-compulsive disorder (OCD), panic disorder with (PDA) or without agoraphobia (PD), posttraumatic stress disorder (PTSD), or social anxiety disorder (SAD). Diagnoses were determined through the use of the Dutch version (Groenestijn, Akkerhuis, Kupka, Schneider, & Nolen, 1998) of the Structured Clinical Interview for DSM-IV Axis-I Disorders (SCID-I; First, Gibbon, Spitzer, & Williams, 1996). The distribution of main diagnoses across the sample is shown in Table 1. Of the final sample, 131 (81%) participants had at least one co-morbid DSM-IV-TR-diagnosis. All patients received treatment at the Altrecht Academic Anxiety Centre (AAA), a specialized mental healthcare provider in the Netherlands. The subjects were outpatients and suffered from severe anxiety symptoms. At the AAA, patients received CBT. Some patients suffering from PTSD received EMDR. Patients received 3 to 65 therapy sessions (M = 19.3, SD = 10.6). A total of 81 (50.6%) patients received medication.

Table 1

Distribution of Main Diagnoses across the Sample.

Main diagnosis	*N*	%
OCD	44	27.5
PTSS	16	10.0
PD & PDA	56	35.0
SAD	44	27.5

Procedure

All data were obtained by the AAA between 2008 and 2013. The patients completed a pretest and posttest, which consisted of two generic assessment measures and two diagnosis-specific assessment measures. All patients completed the questionnaires in order to enable the assessment of treatment outcome. The pretest was administered shortly before the commencement of treatment. The posttest was administered directly after the last therapy session. All questionnaires were filled out on a computer, either at home or at the AAA. All data were anonymized and their use in this study was approved by the AAA.

Measures

Several demographic variables were ascertained, such as age and nationality. Also, all patients filled out the Outcome Questionnaire (OQ-45) and the Brief Symptom Inventory (BSI) as generic measures. Furthermore, patients were assessed with diagnosis-specific measures. Patients diagnosed with OCD filled out the Yale-Brown Obsessive Compulsive Scale (Y-BOCS) and the Obsessive-Compulsive Inventory-Revised (OCI-R). Patients diagnosed with PD or PDA filled out the Mobility Inventory (MI) and the Body Sensation Questionnaire (BSQ). PTSD patients completed the PTSD Symptom Scale–Self-Report Version (PSS-SR) and the Posttraumatic Cognitions Inventory (PTCI). Both the Social Phobia and Anxiety Inventory (SPAI) and the Self-Consciousness Scale (SCS) were completed by patients diagnosed with SAD. Mental health care institutions frequently apply the assessment instruments analyzed in this study.

1. *Generic measures*

OQ-45. The Dutch version (De Jong et al., 2008) of the OQ-45 (Lambert et al., 1996) is a self-report generic assessment measure which consists of 45 items. Each item is scored on a 5-point Likert scale ranging from 0 (never) to 4 (almost always). The OQ-45 is comprised of three subscales. The subscale Symptomatic Distress concerns symptoms of common psychiatric disorders, such as depression and anxiety disorders. The subscale Interpersonal Relations measures an individual's functioning with regard to personal relationships with their partner, family, and friends. The subscale Social Role measures an individual's functioning in school, at work, and at one's leisure. In the Dutch OQ-45 an additional subscale is distinguished, namely Anxiety and Somatic Distress. The psychometric properties of the

Dutch OQ-45 have shown to be adequate both for a general population and for a clinical population (De Jong et al., 2008). The Cronbach's alpha of the total scale of the Dutch version is .96, which indicates excellent internal consistency. The test-retest reliability is adequate (r = .79). The validity of the questionnaire has also been found to be adequate. Moreover, the OQ-45's sensitivity to change is considered to be very good.

BSI. The Dutch version (De Beurs & Zitman, 2006) of the BSI (Derogatis, 1975) is a self-report generic assessment measure which consists of 53 items. Respondents are asked to give an indication of the extent to which certain problems have bothered them the past week, including the present day. Each item is scored on a 5-point Likert scale ranging from 0 (not at all) to 4 (very much). The Dutch BSI measures nine dimensions of psychopathology, namely: somatization, cognitive difficulties, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. Each dimension is comprised of four to six items. The overall score indicates the severity of an individual's psychopathology. The Dutch BSI exhibits sound psychometric properties (De Beurs & Zitman, 2006). The total scale of the questionnaire has demonstrated excellent internal consistency (α = .96) and the test-retest reliability is good (r = .90). Moreover, the validity of the Dutch BSI is considered to be sufficient. The questionnaire's sensitivity to change is adequate.

2. Diagnosis-specific measures

2.1 OCD

Y-BOCS. The Dutch version (Arrindell, Albersnagel, & Van Oppen, 1990) of the Y-BOCS (Goodman et al., 1989a) is an assessment instrument developed to measure the type and severity of obsessive-compulsive symptoms. The inventory is comprised of a semi-structured clinical interview and a self-report symptom checklist. In this study, only the clinical interview was used. This interview consists of 10 items and assesses the severity of obsessions and compulsions. Each item is scored on a 5-point Likert scale ranging from 0 (no symptoms) to 4 (extreme symptoms). Through the interview an indication is obtained of the time occupied by obsessive-compulsive symptoms, the extent to which the symptoms interfere with an individual's functioning, the degree of distress, the amount of offered resistance, and an individual's success in resistance. The psychometric properties of the Dutch Y-BOCS have not been examined. However, the original version of the Y-BOCS appears to have adequate reliability and validity coefficients (Goodman et al., 1989a; Goodman et al., 1989b).

Furthermore, the Dutch Y-BOCS has shown to be sensitive to change (Van Oppen, Emmelkamp, Van Balkom, & Van Dyck, 1995).

OCI-R. The Dutch translation (Cordova-Middelbrink, Dek, & Engelbarts, 2007) of the OCI-R (Foa et al., 2002) is a diagnosis-specific self-report questionnaire which measures OCD characteristics. The OCI-R consists of 18 items. Respondents are asked to give an indication of the extent to which they have been distressed by certain OCD characteristics during the past month. Each item is scored on a 5-point Likert scale ranging from 0 (not at all) to 4 (extremely). The inventory is comprised of six subscales, namely: washing, obsessing, hoarding, ordering, checking, and neutralizing. The psychometric properties of the Dutch translation of the OCI-R have been found to be good (Cordova-Middelbrink, Dek, & Engelbarts, 2007).

2.2 PTSD

PSS-SR. The Dutch translation (Arntz, 1993) of the PSS-SR (Foa, Riggs, Dancu, & Rothbaum, 1993) is a diagnosis-specific self-report questionnaire developed to measure PTSD symptomatology. The PSS-SR contains 17 items corresponding to each of the 17 DSM-IV (APA, 1994) symptoms of PTSD. Respondents are asked to rate the severity of the PTSD-symptoms during the past two weeks. Each item is scored on a 4-point Likert scale ranging from 0 (not at all) to 3 (almost always). The PSS-SR is comprised of three subscales, parallel to the DSM-IV criteria, namely reexperiencing, avoidance, and arousal symptoms. The questionnaire's psychometric qualities appear to be good (Wohlfarth, Van den Brink, Winkel, & Ter Smitten, 2003). The Cronbach's alpha of the total scale is .93. The Dutch PSS-SR's validity has also been found to be good.

PTCI. The Dutch translation (Van Emmerik, Schoorl, Emmelkamp, & Kamphuis, 2006) of the PTCI (Foa, Tolin, Ehlers, Clark, & Orsillo, 1999) is a diagnosis-specific self-report assessment instrument designed to assess trauma-related thoughts and beliefs. The PTCI consists of 36 items that are rated on a 7-point Likert scale ranging from 1 (totally disagree) to 7 (totally agree). The questionnaire is comprised of three subscales, namely: negative cognitions about self (21 items), negative cognitions about the world (7 items), and self-blame (5 items). The psychometric properties of the Dutch version of the PTCI have been found to be adequate. The Cronbach's alpha of the total scale is .94, which indicates excellent internal consistency. The test-retest reliability has been found to be adequate (r = .79) and the validity

of the Dutch PTCI is good. Moreover, scores on the questionnaire have demonstrated to covary with PTCI symptomatology, which emphasizes the PTCI's utility as an outcome measure.

2.3 PD & PDA

MI. The Dutch version (De Beurs, 2001) of the MI (Chambless, Caputo, Jasin, Gracely, & Williams, 1985) is a diagnosis-specific self-report assessment measure which measures agoraphobic avoidance. The inventory is comprised of three parts. The first part consists of 26 items which are scored on a 5-point Likert scale ranging from 1 (never) to 5 (always). Respondents are asked to give an indication of the degree to which they avoid certain situations due to anxiety, when they are alone or when they are with a trusted person. Secondly, respondents are asked to circle five situations that most negatively influence their lives. The final part concerns panic attacks. Respondents are asked to give an indication of both the severity and the amount of panic attacks they have experienced the past seven days and the past three weeks. The psychometric qualities of the Dutch version of the MI are unknown. However, the original version of the MI appears to have sound psychometric properties (Chambless et al., 1985).

BSQ. The Dutch version (Bouman, 1998) of the BSQ (Chambless, Caputo, Bright, & Gallagher, 1984) is a diagnosis-specific self-report inventory which assesses anxiety for physical sensations which can occur when an individual is anxious. The BSQ consists of 17 items which are scored on a 5-point Likert scale ranging from 1 (not at all) to 5 (extremely). Respondents are asked to give an indication of the degree to which they are frightened by certain sensations. A definitive factor solution has not been determined. The Dutch BSQ has sound psychometric properties (Bouman, 1998). The Cronbach's alpha of the total scale is .91, which indicates excellent internal consistency. Furthermore, the validity of the questionnaire is good.

2.4 SAD

SPAI. The Dutch version (Scholing, Bögels, & Van Velzen, 1995) of the SPAI (Turner, Beidel, Dancu, & Stanley, 1989) is a diagnosis-specific self-report questionnaire which assesses social phobia symptomatology. The questionnaire contains 45 items. Each item is scored on a 7-point Likert scale ranging from 1 (never) to 7 (always). Ten items concern agoraphobia and 32 items measure social phobia. Some items assessing social phobia are answered four times, namely with regard to strangers, the opposite sex, authority figures, and

people in general. The reliability and validity of the Dutch SPAI appear to be good (Scholing et al., 1995).

SCS. The Dutch translation (Vleeming & Engelse, 1981) of the SCS (Fenigstein, Scheier, & Buss, 1975) is a diagnosis-specific self-report inventory designed to measure self-consciousness. The questionnaire consists of 23 items and is comprised of three subscales: private self-consciousness (10 items), public self-consciousness (7 items), and social anxiety (6 items). Each item is rated on a 5-point Likert scale ranging from 0 (extremely uncharacteristic/strongly disagree) to 4 (extremely characteristic/strongly agree). The psychometric properties of the Dutch translation of the SCS are unknown. However, the reliability and validity of the original version of the questionnaire have been found to be good (Fenigstein et al., 1975).

Results

In preparation of the analyses, mean differences of treatment progress were calculated. First, total scores for each patient on the assessment instruments were computed. Next, as all assessment instruments consist of different scales, the total scores were standardized into z-scores. That is, for each questionnaire the mean of the standardized pre- and post-treatment assessments taken together was 0, with a standard deviation equal to 1. In order to increase statistical power one overall variable for the diagnosis-specific instruments was created. This variable was computed by averaging the standardized total scores on the completed diagnosis-specific instruments per patient. Subsequently, mean differences were calculated per patient for the OQ-45, BSI, and diagnosis-specific instruments by subtracting standardized post-treatment assessments from standardized pre-treatment assessments. Therefore, the mean differences give an indication of treatment progress.

The analyses were completed in three steps. First, mean differences of treatment progress as assessed by the OQ-45, BSI, and averaged diagnosis-specific instruments were compared. Second, main diagnosis was taken into account and the analysis of the first step was repeated. Finally, mean differences as assessed by the OQ-45, BSI, and separate diagnosis-specific instruments were compared per diagnostic group. The results are presented below.

OQ-45, BSI, and averaged diagnosis-specific instruments

To test the hypothesis that the differences between pre- and post-treatment assessments are larger for diagnosis-specific instruments than for generic instruments a one-way repeated-measures analysis of variance (ANOVA) was completed. The independent variable was type of questionnaire consisting of three categories, namely the OQ-45, BSI, and averaged diagnosis-specific instruments. Mean differences of treatment progress per type of questionnaire are illustrated in Figure 1.

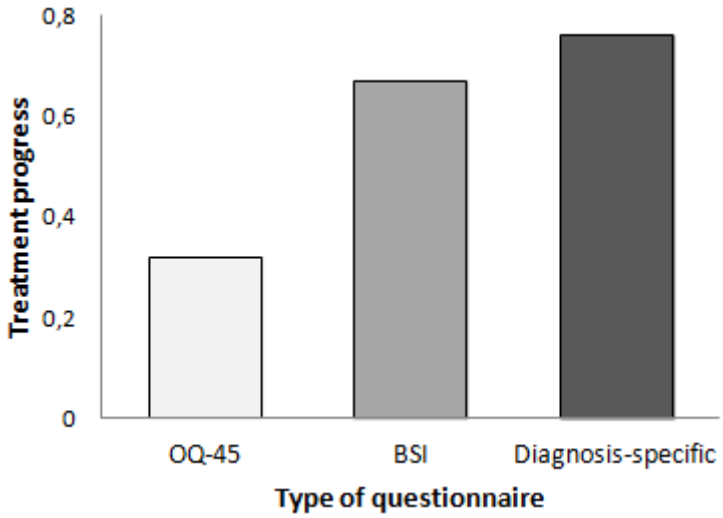

Figure 1. Standardized mean differences of treatment progress as assessed by generic assessment instruments (OQ-45 and BSI) and the averaged diagnosis-specific instruments.

Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(2) = 52.06$, $p < .001$. Therefore, degrees of freedom were corrected applying Huynh-Feldt estimates of sphericity ($\varepsilon = .79$). The results showed that patients' mean differences of treatment progress significantly differ across type of assessment instrument, $F(1.57, 250.32) = 9.96$, $p < .001$. The effect size was moderate (Pallant, 2010), $\eta_p^2 = .06$. Post hoc comparisons using the Bonferroni test indicated that treatment progress as assessed by the BSI ($M = 0.67$, $SD = 0.06$) was significantly larger than treatment progress as assessed by the OQ-45 ($M = 0.32$, $SD = 0.10$), $p = .01$. Moreover, treatment progress as assessed by diagnosis-specific instruments ($M = 0.76$, $SD = 0.06$) was larger than that assessed by the OQ-45, $p < .001$. No significant differences were found between treatment progress assessed by the BSI and treatment progress assessed by diagnosis-specific instruments, $p = .61$.

OQ-45, BSI, and averaged diagnosis-specific instruments per main diagnosis

A 3 (type of assessment instrument) x 4 (main diagnosis) repeated-measures ANOVA was computed in order to test whether mean differences of treatment progress as measured by diagnosis-specific and generic assessment instruments differ among main diagnoses. Mean differences of treatment progress per type of questionnaire and per diagnostic group are illustrated in Figure 2.

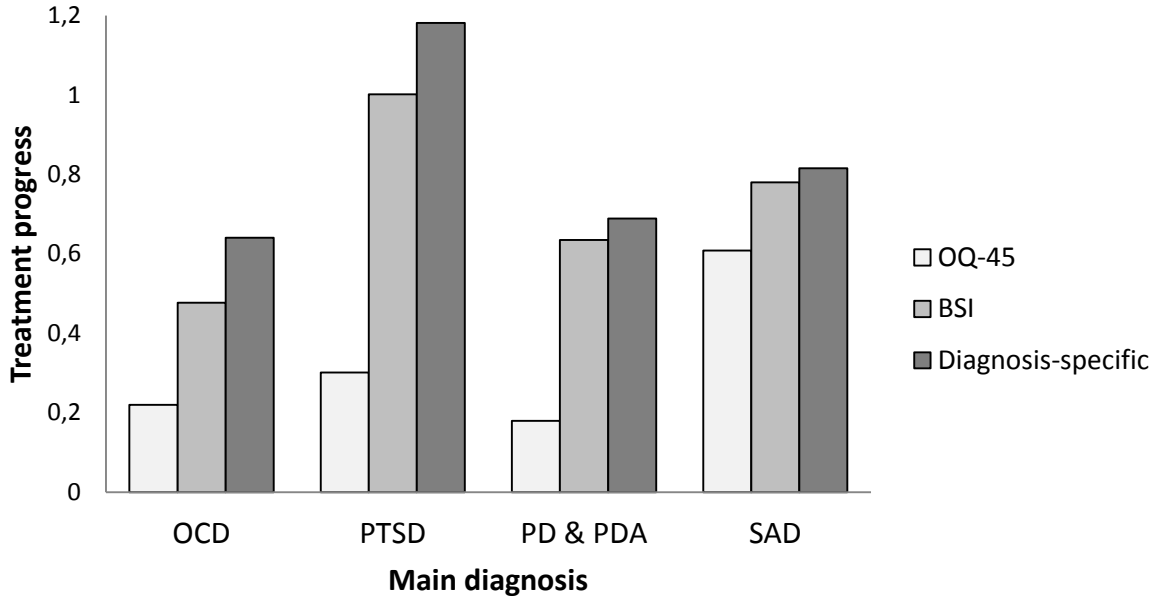

Figure 2. Standardized mean differences of treatment progress as assessed by generic assessment instruments (OQ-45 and BSI) and the averaged diagnosis-specific instruments per main diagnosis.

Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(2) = 50.30$, $p < .001$. Therefore, degrees of freedom were corrected applying Huynh-Feldt estimates of sphericity ($\varepsilon = .81$). The main effect of type of assessment instrument was significant, $F(1.61, 251.09) = 10.35$, $p < .001$. Mean differences of treatment progress appear to differ among the three types of assessment instruments similarly as in the previous analysis. Moreover, a significant main effect was found for main diagnosis, $F(3, 156) = 3.06$, $p = .03$. The four groups of main diagnoses seem to differ in treatment progress independent of type of assessment instrument. The effect size was moderate, $\eta_p^2 = .06$. Post hoc comparisons using the least significant difference test indicated that, independent of assessment instrument, treatment progress was significantly larger for PTSD patients ($M = 0.83$, $SD = 0.80$) than for OCD patients ($M = 0.45$, $SD = 0.51$), $p = .03$. Furthermore, overall treatment progress appears to be significantly larger for SAD patients ($M = 0.73$, $SD = 0.54$) than for OCD patients, $p = .02$. The interaction between type of assessment instrument and main diagnosis was not significant, $F(4.83, 251.09) = 0.69$, $p = .63$. Hence, the sensitivity of the different types of assessment instruments does not appear to differ per main diagnosis.

OQ-45, BSI, and separate diagnosis-specific instruments per main diagnosis

Four repeated-measures ANOVA's were computed to test whether mean differences of treatment progress differ between the OQ-45, BSI, and all completed diagnosis-specific instruments per main diagnosis. All mean differences are illustrated in Figure 3.

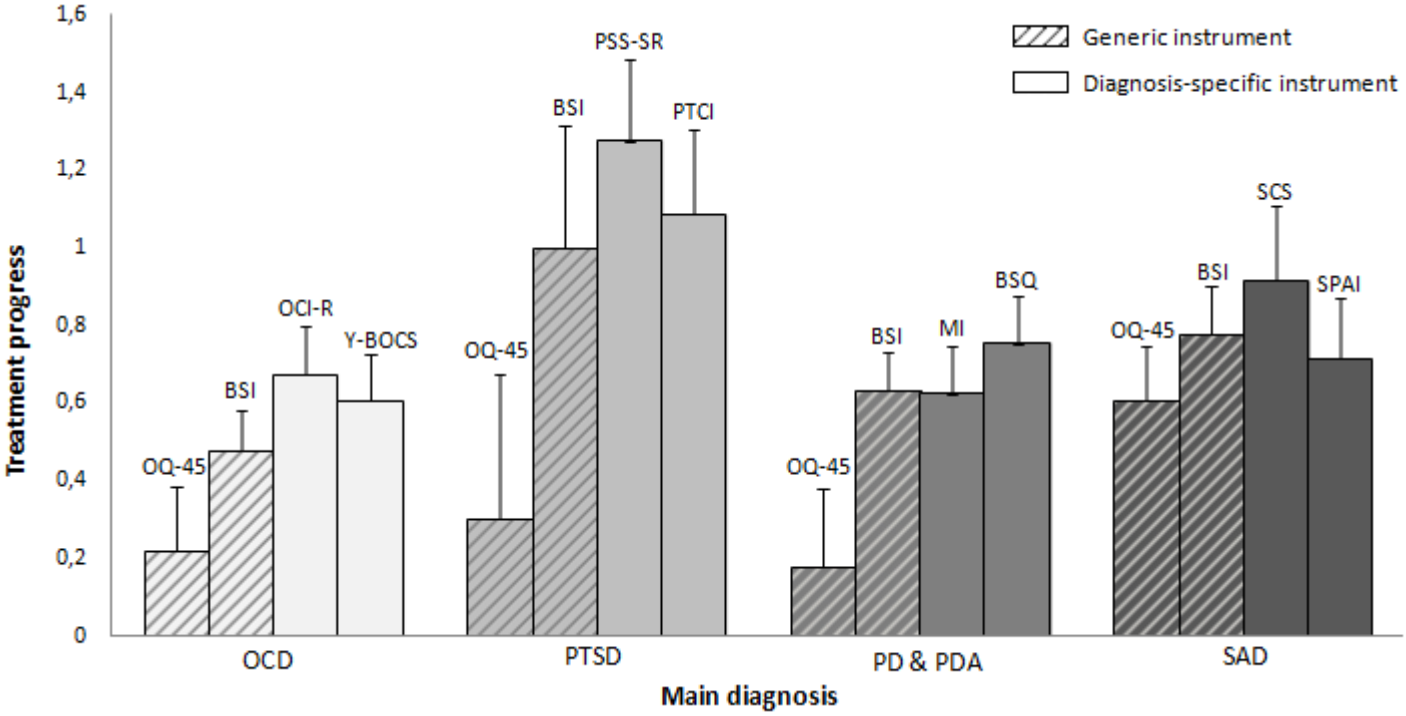

Figure 3. Standardized mean differences of treatment progress per main diagnosis as assessed by several assessment instruments. Error bars represent standard errors of the mean. The generic assessment measures are the Outcome Questionnaire (OQ-45) and the Brief Symptom Inventory (BSI). The diagnosis-specific instruments are the Obsessive-Compulsive Inventory-Revised (OCI-R), Yale-Brown Obsessive Compulsive Scale (Y-BOCS), PTSD Symptom Scale–Self-Report Version (PSS-SR), Posttraumatic Cognitions Inventory (PTCI), Mobility Inventory (MI), Body Sensation Questionnaire (BSQ), Self-Consciousness Scale (SCS), and the Social Phobia and Anxiety Inventory (SPAI).

OCD (N = 44). Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 37.71$, $p < .001$. Therefore, degrees of freedom were corrected applying Greenhouse-Geisser estimates of sphericity ($\varepsilon = .64$). The results indicated that OCD patients' mean differences of treatment progress do not significantly differ between assessments by the OQ-45, BSI, OCI-R, and Y-BOCS, $F(1.91, 82.27) = 2.78$, $p = .07$. However, a marginal trend

towards significance can be seen, such that the OCI-R (M = 0.67, SD = 0.83) appears to show more treatment progress than the OQ-45 (M = 0.22, SD = 1.07), p = .10.

PTSD (N =16). Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 24.60$, $p < .001$. Therefore, degrees of freedom were corrected applying Greenhouse-Geisser estimates of sphericity (ε = .48). The results indicated that PTSD patients' mean differences of treatment progress do not significantly differ between assessments by the OQ-45, BSI, PSS-SR, and PTCI, $F(1.44, 21.66) = 2.94$, p = .09. However, a marginal trend towards significance can be seen, such that the PSS-SR (M = 1.27, SD = 0.84) seems to show more treatment progress than the OQ-45 (M = 0.30, SD = 1.49), p = .09.

PD & PDA (N = 56). Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 31.24$, $p < .001$. Therefore, degrees of freedom were corrected applying Greenhouse-Geisser estimates of sphericity (ε = .74). The results indicated that PD(A) patients' mean differences of treatment progress significantly differ across assessment instruments, $F(2.22, 122.29) = 3.49$, p = .03. The effect size was moderate, η_p^2 = .06. Post hoc comparisons using the least significant difference test indicated that mean differences of treatment progress as assessed by the OQ-45 (M = 0.18, SD = 1.50) were significantly smaller than those assessed by the MI (M = 0.62, SD = 0.91), p = .02. Also, treatment progress as assessed by the OQ-45 was smaller than progress as assessed by the BSQ (M = 0.75, SD = 0.92), p = .01.

SAD (N = 44). Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 24.15$, $p < .001$. Therefore, degrees of freedom were corrected applying Greenhouse-Geisser estimates of sphericity (ε = .74). The results indicated that OCD patients' mean differences of treatment progress do not significantly differ between assessments by the OQ-45, BSI, SPAI, and SCS, $F(2.22, 95.51) = 0.73$, p = .50.

Discussion

The purpose of this study was to examine the sensitivity of diagnosis-specific versus generic assessment instruments which are applied in ROM for anxiety disorders. At pre- and post-treatment all participants filled out two generic assessment instruments, namely the OQ-45 and BSI, and two diagnosis-specific instruments. It was expected that the differences between pre- and post-treatment assessments are larger for diagnosis-specific instruments than for generic instruments. The results show a subtle pattern. As indicated by Figure 1, treatment progress as assessed by the (generic) BSI was larger than treatment progress as assessed by the (generic) OQ-45. Also, effects of treatment as assessed by diagnosis-specific instruments in general appeared to be larger than treatment progress assessed by the OQ-45. However, although a non-significant trend was found in favor of diagnosis-specific instruments, the sensitivity of the BSI and diagnosis-specific instruments did not seem to differ.

Moreover, as can be derived from Figure 2, the sensitivity of the generic and diagnosis-specific assessment instruments did not differ per main axis-I diagnosis. Furthermore, as shown in Figure 3, the sensitivity of the generic and all separate diagnosis-specific instruments was compared per diagnostic group. Remarkably, through this analysis differences in sensitivity were only found for the questionnaires filled out by patients with PD(A). Treatment progress as assessed by both the MI and BSQ (diagnosis-specific instruments) appeared to be larger than treatment progress as assessed by the OQ-45 (generic assessment instrument). Nevertheless, the sensitivity of the (generic) BSI did not seem to differ from the accuracy of the MI or BSQ. For the samples of OCD and PTSD patients trends towards significance were found. The OCI-R appeared to show more treatment progress for OCD than the OQ-45 did. For PTSD patients treatment progress as assessed by the PSS-SR seemed to be larger than progress assessed by the OQ-45. Interestingly, when each diagnostic group was considered separately, the sensitivity of the BSI did not appear to differ from the sensitivity of the OQ-45.

It may seem to be counter-intuitive that the examined generic assessment instruments differ in assessment sensitivity when all diagnostic groups are merged. However, the BSI assesses general symptom reduction (e.g. “feeling anxious to travel by bus, train, or tram”, “having hot flashes or cold shivers”; De Beurs & Zitman, 2006; Derogatis, 1975), whereas the OQ-45 assesses an individual’s overall functioning (e.g. “I am satisfied with my life”, “I am not working/studying as well as I used to”; De Jong et al., 2008; Lambert et al., 1996). Moreover, two out of the nine subscales of the BSI measure aspects of anxiety. The OQ-45, on

the other hand, does not include such subscales. As previously stated, CBT targets the specific symptomatology for which a patient seeks treatment. Therefore, it may not be surprising that the BSI shows more treatment progress in anxiety disorders than does the OQ-45. This may also explain the fact that the sensitivity of the BSI did not appear to differ from the sensitivity of the diagnosis-specific instruments. It logically follows that it may also account for the fact that treatment progress assessed by the MI and BSQ was larger than progress assessed by the OQ-45, but did not differ from treatment progress determined by the BSI.

The BSI and OQ-45 do not appear to differ in sensitivity when each assessment instrument is considered separately. This seems to reflect a power problem. Sample size influences statistical power such that smaller sample sizes decrease the probability of detecting an existing difference and larger sample sizes increase this chance (Cohen, 1988). As stated, for the samples of OCD patients and PTSD patients trends towards significance were found. It is likely that if larger sample sizes had been obtained, significant differences would have been found. The fact that differences in the sensitivity of the OQ-45 and BSI were found when all main diagnoses were considered together (i.e., larger sample size) but failed to be found when each main diagnosis was considered separately (i.e., smaller sample sizes) supports this reasoning.

However, the differences in sensitivity of the questionnaires completed by the sample of SAD patients were far from significant. Furthermore, the sample size of SAD patients was relatively large. Hence, it is unlikely that a lack of statistical power explains the fact that no differences in sensitivity were found for the questionnaires assessing SAD patients' treatment progress. It is tempting to speculate that the OQ-45 is a relatively specific assessment measure for SAD. Figure 3 shows that the differences in treatment progress as assessed by the OQ-45 and the other questionnaires are relatively small for SAD patients. Two out of the four subscales of the OQ-45 measure social role and interpersonal relations. Individuals diagnosed with SAD suffer from intense fear of the scrutiny of others. As a consequence, an individual with SAD feels anxious in social or performance occasions, or may even avoid such interpersonal encounters (Stein & Stein, 2008). Accordingly, it is likely that SAD patients' scores on the subscales social role and interpersonal relations, and therefore their scores on the OQ-45, will improve following treatment.

Another notable outcome is that PTSD patients' treatment progress appears to be larger than that of OCD, PD(A), and SAD patients (see Figure 3). This result seems to be in line with observations made earlier by Bisson and colleagues (2007). Through a meta-analysis, they found that PTSD patients improve more following CBT than do other anxiety patients. This may

be explained by the fact that in the treatment of PTSD the typical course of processing a traumatic experience is set in motion (Pool, Heuvel, Ranchor, & Sanderman, 2004). For other anxiety disorders, the treatment may be more complex.

One might perhaps assume that assessment instruments with larger effect sizes provide a more accurate estimate of treatment progress. Though this assumption may seem plausible, in order to determine an instruments' accuracy the purpose of psychological treatment should be taken into consideration. It could be stated that patients seek help in order to reduce certain symptoms. Therefore, the primary goal of treatment would be to diminish these symptoms and treatment progress should be determined by assessing the degree to which a patient experiences specific symptoms. This is in accordance with Freud's notion that "much will be gained if [psychoanalysis] can succeed in transforming hysterical misery into common unhappiness. With a mental life that has been restored to health [an individual] will be better armed against that unhappiness" (Breuer & Freud, 1893/1955, p. 305). According to Freud, the intention of treatment is not to induce happiness, but to reduce pathological suffering to the level of ordinary suffering. The above implies that diagnosis-specific assessment instruments more accurately depict treatment progress. However, why would one choose to treat an individual's symptoms if doing so would not improve their quality of life? Otherwise stated, if a reduction in an individual's scores on diagnosis-specific instruments does not indicate an improvement of scores on generic quality of life assessment instruments, the value of treatment is disputable. Perhaps the goals of treatment, and therefore the measure of treatment progress, should be individually determined. Correspondingly, the assessment instruments applied in ROM should be individually selected.

Some limitations should be considered when interpreting the present findings. As previously discussed, the modest sample sizes used in this study may have limited statistical power. Simple power analysis using GPower (Faul, Erdfelder, Lang, & Buchner, 2007) teaches that each sample size per main diagnosis should increase to approximately $N = 50$ in order for differences to reach statistical significance at the .05 level. Moreover, the relatively high comorbidity rates may have influenced the results. However, statistical analyses show that comorbidity was not associated with any of the outcomes.

In summary, this study examined the sensitivity of the assessment of treatment progress and outcome in anxiety disorders through diagnosis-specific versus generic ROM instruments. Indications were found that the BSI and diagnosis-specific assessment instruments show larger effect sizes of treatment progress in anxiety disorders than the OQ-45 does. Therefore, the justifiability of the widespread use of the OQ-45 in ROM (Van Beljouw & Verhaak, 2010) can

be questioned. However, when selecting an assessment instrument for ROM, one should take into consideration the intended goal of treatment.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Arntz, A. (1993). *Dutch translation of the PSS-SR*. Universiteit Maastricht, Nederland: Author.

Arrindell, W.A., Albersnagel, F.A., & Oppen, P. van. (1990). Yale-Brown Obsessive Compulsive Scale (Y-BOCS). Nederlandse bewerking, verkorte versie. Retrieved 13 November 2013 from http://www.tijdschriftvoordepsychiatrie.nl/assets/measuringinstruments/meetinstrumenten_130pdf.pdf.

Beck, J. S. (2011). *Cognitive behavioral therapy: the basics and beyond* (2nd ed.). New York, NY: The Guilford Press.

Beljouw, I. M. J., van, & Verhaak, P. F. M. (2010). *Geschikte uitkomstmaten voor routinematige registratie door eerstelijnspsychologen*. Utrecht: Nivel.

Beurs, E. de (2001). *Mobility Inventory, Nederlandse vertaling*. Leiden: Leids Universitair Medisch Centrum.

Beurs, E. de & Zitman, F. G. (2006). De Brief Symptom Inventory (BSI): de betrouwbaarheid en validiteit van een handzaam alternatief voor de SCL-90. Maandblad *Geestelijke Volksgezondheid, 61*, 120-141.

Bilsker, D., & Goldner, E. M. (2002). Routine outcome measurement by mental health-care providers: Is it worth doing? *Lancet*, *360*, 1689-1690.

Bisson, J. I., Ehlers, A., Matthews, R., Pilling, S., Richards, D., & Turner, S. (2007). Psychological treatments for chronic post-traumatic stress disorder: Systematic review and meta-analysis. *British Journal of Psychiatry, 190*, 97-104.

Bouman, T.K. (1998). Kort instrumenteel: De Body Sensation Questionnaire (BSQ). *Gedragstherapie, 31*, 165-168.

Breuer, J., & Freud, S. (1955). Studies on hysteria. In J. Strachey (Ed. & Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 2, pp. 1-305). London: Hogarth Press. (Original work published 1893-1895).

Buwalda, V. J. A., Nugter, M. A., Van de Brug, S. Y., Draisma, S., Smit, J. H., Swinkels, J. A., & Van Tilburg, W. (2012). Rom in de volwassenenpsychiatrie: een evaluatie van meetinstrumenten. *Tijdschrift voor Psychiatrie*, *54*, 147-150.

Carlier, I. V. E., Meuldijk, D., Van Vliet, I. M., Van Fenema, E., Van der Wee, N. J. A., & Zitman, F. G. (2012). Routine outcome monitoring and feedback on physical or mental health status: Evidence and theory. *Journal of Evaluation in Clinical Practice*, *18*, 104–110.

Chambless, D. L., Caputo, G. C., Bright, P., & Gallagher, R. (1984). Assessment of fear of fear in agoraphobics: the body sensations questionnaire and the agoraphobic cognitions questionnaire. *Journal of Consulting and Clinical Psychology*, *52*, 1090-1097.

Chambless, D. L., Caputo, G. C., Jasin, S. E., Gracely, E. J., & Williams, C. (1985). The mobility inventory for agoraphobia. *Behaviour Research and Therapy*, *23*, 35-44.

Clark, D. A. & Beck, A. T. (2010). *Cognitive Therapy of Anxiety Disorders*. New York, NY: The Guilford Press.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cordova-Middelbrink, J. A., Dek, E. C. P., & Engelbarts, M. M. B. (2007). *Automatisering bij subklinische obsessief-compulsieve stoornis* [Automatic processes in subclinical obsessive compulsive disorder] (Unpublished master's thesis). Utrecht University, Utrecht, Netherlands.

Derogatis, L.R. (1975). *The Brief Symptom Inventory*. Baltimore, MD: Clinical Psychometric Research.

Emmerik, A. A. P. van, Schoorl, M., Emmelkamp, P. M. G., & Kamphuis, J. H. (2006). Psychometric evaluation of the Dutch version of the posttraumatic cognitions inventory (PTCI). *Behaviour Research and Therapy*, *44*, 1053-1065.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods,39*, 175-191.

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology, 43*, 522-527.

First, M.B., Gibbon, M., Spitzer, R.L., Williams, J.B.W. (1996). *User's guide for the Structured Clinical Interview for the DSM-IV Axis-I disorders - research version*. New York: Biometrics Research Department, New York State Psychiatric Institute.

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., et al. (2002). The Obsessive-Compulsive Inventory: Development and validation of a short version. *Psychological Assessment, 14*, 485-495.

Foa, E. B., Riggs, D. S., Dancu, C. V., Rothbaum, B. O. (1993). Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress, 6*, 459–473.

Foa, E. B., Tolin, D. F., Ehlers, A., Clark, D. M. & Orsillo, S. M. (1999). The posttraumatic cognitions inventory (PTCI): Development and validation. *Psychological Assessment, 11*, 303-314.

Franklin, M. E., Abramowitz, J. S., Kozak, M. J., Levitt, J. T., & Foa, E. B. (2000). Effectiveness of exposure and ritual prevention for obsessive-compulsive disorder: Randomized compared with nonrandomized samples. *Journal of Consulting and Clinical Psychology, 68*, 594–602.

Gilbody, S. M., House, A. O., & Sheldon, T. A. (2002). Outcomes research in mental health: Systematic review. *British Journal of Psychiatry*, *181*, 8-16.

Gray, J. A. & McNaughton, N. (2003). *The Neuropsychology of Anxiety* (2nd ed.). Oxford: Oxford University Press.

Groenestijn, M. A. C. van, Akkerhuis, G. W., Kupka, R. W., Schneider, N., & Nolen, W. A. (1998). *SCID-I: Gestructureerd klinisch interview voor de vaststelling van DSM-IV as-I stoornissen*. Lisse: Swets en Zeitlinger.

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., Heninger, G. R., & Charney, D. S. (1989a). The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. *Archives of General Psychiatry, 46*, 1006-1011.

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Delgado, P., Heninger, G. R., & Charney, D. S. (1989b). The Yale-Brown Obsessive Compulsive Scale. II. Validity. *Archives of General Psychiatry, 46*, 1012-1016.

Hermann, R. C. (2005). Linking outcome measurement with process measurement for quality improvement. In: Ishak, W. W., Burt, T., Sederer, L. I., (Red.). *Outcome measurement in psychiatry, a critical review* (p. 23-34). Washington: American Psychiatric Publishing.

Hermans, D., Eelen, P., & Orlemans, H. (2007). *Inleiding tot de gedragstherapie* (6e druk). Houten: Bohn Stafleu van Loghum.

Hoenders, R. H. J., Bos, E. H., Bartels-Velthuis, A. A., Vollbehr, N. K., Ploeg, K., van der, Jonge, P., de, & Jong, J. T. V. M., de. (2013). Pittfalls in the assessment, analysis, and interpretation of routine outcome monitoring (ROM) data: Results from an outpatient

clinic for integrative mental health. *Administration and Policy in Mental Health and Mental Health Services Research*, DOI 10.1007/s10488-013-0511-7.

Hofmann, S. G., Asnaani, A., Imke, J. J., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research*, *36*, 427-440.

Hofmann, S. G., & Smits, J. A. J. (2008). Cognitive-behavioral therapy for adult anxiety disorders: a meta-analysis of randomized placebo-controlled trials. *Journal of Clinical Psychiatry*, *69*, 621– 632.

Hoyer, J., Becker, E. S., Neumer, S., Soeder, U., & Margraf, J. (2002). Screening for anxiety in an epidemiological sample: predictive accuracy of questionnaires. *Anxiety Disorders*, *16*, 113-134.

Jong, K. de, Nugter, M. A., Polak, M. G., Wagenborg, J. E. A., Spinhoven, P., & Heise, W. J. (2008). De Nederlandse versie van de Outcome Questionnaire (OQ-45): een crossculturele validatie. *Psychologie en Gezondheid*, *36*, 35-45.

Kessler, D. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 593-602.

Laane, R. & Luijk, R. (2012). ROM en de positie van zorgverzekeraars. *Tijdschrift voor psychiatrie*, *54*, 135-139.

Lambert, M. J. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*, *17*, 1-14.

Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C., Christopherson, C., & Burlingame, G. M. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology & Psychotherapy*, *3*, 249-258.

Lambert, M. J. (Ed.). (2013). *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (6th ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Luteijn, F., Barelds, D. P. H., Arrindell, W. A., Deelman, B. G., Kamphuis, J. H., & Vertommen, H. (2011). *Psychologische diagnostiek in de gezondheidszorg* (2nd ed.). The Hague: Boom Lemma uitgevers.

Macdonald, A. J. D., & Trauer, T. (2010). Objections to routine clinical outcomes measurement in mental health services: any evidence so far? *Journal of Mental Health*, *19*, 517-522.

Miller, S. D., Duncan, B. L., Sorrel, R., & Brown, G. S. (2005). The partners for change

outcome management system. *Journal of Clinical Psychology*, *186,* 199-208.

Noom, M. J., De Jong, K., Tiemens, B., Kamsteeg, F., Markus, M. T., Pot, A. M., Schippers, G. M., Swildens, W., Sytema, S., Theunissen, J., Van der Vlist, H. P., Vuyk, J., & Zondervan, T. (2012). Routine outcome monitoring en benchmarking: hoe kunnen we behandelresultaten op een zorgvuldige manier vergelijken? *Tijdschrift voor Psychiatrie*, *54*, 141-145.

Nugter, M. A. & Buwalda, V. J. A. (2012). Achtergronden en gebruiksmogelijkheden van ROM in de ggz. *Tijdschrift voor Psychiatrie*, *54*, 111-118.

Olatunji, B. O. Cisler, J. M., & Deacon, B. J. (2010). Efficacy of cognitive behavioral therapy for anxiety disorders: a review of meta-analytic findings. *Psychiatric Clinics of North America*, *33*, 557-577.

Oppen, P. van, Emmelkamp, P. M. G., Balkom, A. J. L. M. van, & Dyck, R. van. (1995). The sensitivity to change of measures for obsessive-compulsive disorder. *Journal of Anxiety Disorders*, *9*, 241-248.

Pallant, J. (2010). *SPSS Survival Manual: A step by step guide to data analysis using SPSS* (4th ed.). New York, NY: Open University Press.

Patrick, D. L. & Deyo, R. A. (1989). Generic and disease-specific measures in assessing health status and quality of life. *Medical Care*, *27*, 217-232.

Pool, G., Heuvel, F., Ranchor, A. V., & Sanderman, R. (2004). *Handboek psychologische interventies bij chronisch-somatische aandoeningen.* Assen: Koninklijke Van Gorcum BV.

Prus, A. J. (2013). *An Introduction to Drugs and the Neuroscience of Behavior* (International Edition). Wadsworth: Cengage Learning.

Rachman, S. (2013). *Anxiety* (3rd ed.). East Sussex: Psychology Press.

Schat, A., Noorden, M. S., van, Noom, M. J., Giltay, E. J., Wee, N. J. A., van der, Vermeiren, R. R. J. M., & Zitman, F. G. (2013). Predictors of outcome in outpatients with anxiety disorders: The Leiden routine outcome monitoring study. *Journal of Psychiatric Research*, *47*, 1876-1885.

Scholing, A., Bögels, S. M., & Van Velzen, C. (1995). The Dutch Social Phobia and Anxiety Inventory (SPAI). Authorized translation.

Shafran, R., Clark, D. M., Fairburn, C. G., Arntz, A., Barlow, D. H., Ehlers, A., Freeston, M., Garety, P. A., Hollon, S. D., Ost, L. G., Salkovskis, P. M., Williams, J. M. G., & Wilson, G. T. (2009). Mind the gap: Improving the dissemination of CBT. *Behaviour Research and Therapy*, *47*, 902-909.

Stein, M. B., & Stein, D. J. (2008). Social anxiety disorder. *Lancet, 371*, 1115-1125.

Stuart, G. L., Treat, T. A., & Wade, W. A. (2000). Effectiveness of an empirically based treatment for panic disorder in a service clinic setting: 1-year follow-up. *Journal of Consulting and Clinical Psychology, 68*, 506–512.

Trimbos Instituut. (2013). *Multidisciplinaire richtlijn angststoornissen* (3e revisie). Utrecht: Trimbos Instituut.

Turner, S. M., Beidel, D. C., Dancu, C. V. & Stanley, M. A. (1989). An empirically derived inventory to measure social fears and anxiety: The social phobia and anxiety inventory. *Journal of Consulting and Clinical Psychology, 1*, 35-40.

Vleeming, R. G. & Engels, J. A. (1981). Assessment of private and public self-consciousness: A Dutch replication. *Journal of Personality Assessment, 45*, 385-389.

Walburg J.A. & Brinkmann, J. (2001*). Kwaliteit is geen toeval: de praktijk van de kwaliteitszorg in de ggz*. Utrecht: GGZ Nederland.

Wiger D. E., & Solberg K. B. (2001). *Tracking mental health outcomes: A therapist's guide to measuring client progress, analyzing data, and improving your practice*. New York: Wiley.

Wohlfarth, T. D., Brink, W. van den, Winkel, F. W., & Smitten, M. ter. (2003). Screening for posttraumatic stress disorder: An evaluation of two self-report scales among crime victims. *Psychological Assessment, 15*, 101-109.