Journal of Anxiety Disorders 64 (2019) 55–63

Contents lists available at ScienceDirect

Journal of Anxiety Disorders

journal homepage: www.elsevier.com/locate/janxdis

Measuring treatment outcome in patients with anxiety disorders: A comparison of the responsiveness of generic and disorder-specific instruments

S.J. Schawo[a,b], I.V.E. Carlier[c], A.M. van Hemert[c], E. de Beurs[a,d,*]

[a] Institute of Psychology, Faculty of Social Sciences, Leiden University, Pieter De La Court Building, P.O. Box 9555, 2300 RB Leiden, the Netherlands
[b] Institute for Medical Technology Assessment, Erasmus University, Bayle Building, Burgemeester Oudlaan 50, 3062 PA Rotterdam, the Netherlands
[c] Department of Psychiatry, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, the Netherlands
[d] Stichting Benchmark GGZ, Rembrandtlaan 46, 3723 BK Bilthoven, the Netherlands

ARTICLE INFO

Keywords:
Outcome measure
Anxiety
MASQ
BSI
Generic
Disorder-specific

ABSTRACT

Background: For routine outcome monitoring, generic (i.e., broad-based) and disorder-specific instruments are used to monitor patient progress. While disorder-specific instruments may be more sensitive to therapeutic change, generic measures can be applied more broadly and allow for an assessment of therapeutic change, irrespective of a specific anxiety disorder. Our goal was to investigate whether disorder-specific instruments for anxiety disorders are a valuable (or even necessary) addition to generic instruments for an appropriate assessment of treatment outcome in groups of patients.

Methods: Data were collected from 2002 to 2013 from psychiatric outpatients in treatment for Social Phobia (SP; n = 834), Generalized Anxiety Disorder (GAD; n = 661), Panic Disorder (PD; n = 944), Obsessive-Compulsive Disorder (OCD; n = 460), and Posttraumatic Stress Disorder (PTSD; n = 691). Instruments used were the generic Brief Symptom Inventory (BSI), The Mood and Anxiety Symptoms Questionnaire (MASQ), and several disorder-specific instruments (e.g., Social Interaction Anxiety Scale, Social Phobia Scale, Panic Appraisal Inventory, etc.). Responsiveness (i.e., sensitivity to therapeutic change) was examined through correlational analyses, effect sizes (ES), and analysis of variance for repeated measures.

Results: The MASQ appeared generally more responsive than the BSI, except for the BSI Anxiety subscale for PD. Disorder-specific measures equaled the MASQ and BSI in responsiveness. When statistically significant differences occurred, the ES was small.

Discussion/conclusions: For most anxiety disorder groups (i.e., SP, PD and OCD), the MASQ or BSI was equally suited as disorder-specific instruments to detect change at group level. Exceptions are GAD and PTSD. These findings suggest limited incremental information value of disorder-specific instruments over the MASQ and BSI for measuring change.

1. Introduction

In clinical practice, information on the progress and outcome of treatment is needed to make informed treatment choices and to properly determine the course of treatment. Clinical judgment regarding the outcome of mental health interventions is fallible and prone to biases and needs to be supplemented by standardized measurements. Hence, there is a need to make individual patients' improvements measurable and to increase our knowledge regarding the benefits of various mental health interventions.

In recent years, there have been efforts to routinely monitor patient progress before, during, and after treatment in several countries, i.e. the United Kingdom (Clark, 2011), the United States (Lambert & Finch, 1999), and the Netherlands (de Beurs et al., 2011). Recent Dutch efforts are referred to as routine outcome monitoring. With routine outcome monitoring, symptoms and functioning of patients are measured at regular intervals with a battery of instruments. The goals of routine outcome monitoring are to direct patients towards optimal treatment,

* Corresponding author at: Institute of Psychology, Faculty of Social Sciences, Leiden University, Pieter De La Court Building, P.O. Box 9555, 2300 RB Leiden, the Netherlands.
E-mail addresses: saskia.schawo@eshpm.eur.nl (S.J. Schawo), I.V.E.Carlier@lumc.nl (I.V.E. Carlier), A.M.van_Hemert@lumc.nl (A.M. van Hemert), e.de.beurs@fsw.leidenuniv.nl (E. de Beurs).

https://doi.org/10.1016/j.janxdis.2019.04.001
Received 21 August 2018; Received in revised form 25 March 2019; Accepted 2 April 2019
Available online 06 April 2019
0887-6185/

to provide information on treatment progress of individual patients, and to add to the effectiveness research of mental healthcare interventions (de Beurs et al., 2011).

The choice of instruments used for such monitoring is crucial for a proper assessment of treatment progress and results. Within routine outcome monitoring, a distinction is made between generic (i.e. broad-based) and disorder-specific instruments. To be more precise, one could place instruments on a spectrum from a broad focus on general symptoms of psychopathology to a more specific focus on the symptoms of a single mental disorder, such as panic disorder or obsessive-compulsive disorder. Yet, it has to be noted that some generic instruments comprise subscales for specific symptom dimensions, germane to specific anxiety disorders (e.g., the Brief Symptom Inventory [BSI; Derogatis & Melisaratos, 1983]). Thus, some scales can clearly be categorized as generic (e.g. the total score on the BSI) or disorder-specific (the Social Interaction Anxiety Scale [SIAS; Mattick & Clarke, 1998], Social Phobia Scale [SPS; Mattick & Clarke, 1998] or the Panic Appraisal Inventory [PAI; Telch, Brouillard, Telch, Agras, & Taylor, 1989;] total scores and subscale scores). Other scales cannot clearly be labeled as either generic or disorder-specific as they assess generic as well as specific aspects of mental disorders (e.g., the anxiety subscale of the BSI, the total score of the Mood Anxiety Symptoms Questionnaire [MASQ; Clark & Watson, 1991]). See Fig. 1 for an illustration.

An advantage of the use of generic instruments is that outcomes can be assessed and compared across various diagnostic subgroups of patients. Yet, disorder-specific instruments may be more sensitive to detect clinically relevant changes when compared to generic instruments (Engel, Adair, Hayas, & Abraham, 2009; Hay & Mond, 2005; Patrick & Deyo, 1989). If this is true, disorder-specific instruments would be more useful to monitor individual patients as they may better detect small and specific changes, while such changes may remain undetected with generic instruments. A simple solution to obtain comparable, sensitive and clinically-relevant information, would be to administer both generic and disorder-specific instruments. However, it is desirable to limit the burden to patients, especially when monitoring large groups of patients. This makes it opportune to investigate how both types of instruments compare in the assessments of treatment effect and whether using both may be superfluous. When investigating outcomes of groups of patients, the relevant question is how generic and disorder-specific instruments compare with respect to responsiveness (i.e., sensitivity to therapeutic change).

The responsiveness of instruments measuring symptoms of a disorder has rarely been investigated in the field of psychiatry and clinical psychology (Carlier et al., 2017). However, recently we reported on a study comparing generic and disorder specific self-report measures for depression (de Beurs et al., 2018). Comparative studies of health-related quality of life measures do exist for other fields, and these studies have reported mixed results (Bessette et al., 1998; Bombardier et al., 1995; Damiano et al., 1995; Stucki, Liang, Fossel, & Katz, 1995; Angst et al., 2008; Mitchison et al., 2013). To our knowledge, there are a small number of studies on the responsiveness of generic versus disorder-specific instruments for anxiety disorders (Matthey, Fisher, & Rowe, 2013; Schibbye et al., 2014; van der Mheen, ter Mors, van den Hout, & Cath, 2017). The studies by Matthey et al. (2013) and Schibbye et al. (2014) did not constitute direct head-to-head comparisons of anxiety measures. The recent study of van der Mheen et al. (2017) compared disorder-specific anxiety measures with the Outcome Questionnaire (OQ-45; Lambert et al., 1996) and with the BSI. The authors found that disorder-specific measures were more responsive than the OQ-45 total score. In order to strengthen the research base on this important topic, in this study we set out to compare generic and disorder-specific instruments for five specific anxiety disorders: Social Phobia (SP), Generalized Anxiety Disorder (GAD), Panic Disorder (PD), Obsessive-Compulsive Disorder (OCD), and Post-Traumatic Stress Disorder (PTSD). Based on the literature, which suggests disorder-specific instruments to be more sensitive to detect clinically relevant change (Engel et al., 2009; Hay & Mond, 2005; Patrick & Deyo, 1989), our hypothesis is that disorder-specific instruments are more responsive compared to generic scales.

2. Methods

2.1. Research design

The current study used a dataset from the Dutch mental health care institutions GGZ Rivierduinen and from the Department of Psychiatry of the Leiden University Medical Center (LUMC). Data were collected from 2002 to 2013 as part of routine outcome monitoring. Data collection took place at the start of treatment and subsequently every three to four months during treatment (de Beurs et al., 2011).

2.2. Participants

Patients could participate in the study if they were literate and had sufficient mastery of the Dutch language. They were recruited after referral for treatment at the LUMC or mental health care provider GGZ Rivierduinen. The present study is limited to patients who met diagnostic criteria for SP, GAD, PD, OCD, and/or PTSD, based on a structured clinical diagnostic interview, the Mini International Psychiatric Interview (MINI-Plus, Sheehan et al., 1998). Patients with a primary diagnosis of substance use or DSM-IV Axis II disorders were excluded, as were patients with missing data on any of the relevant scales (i.e. the

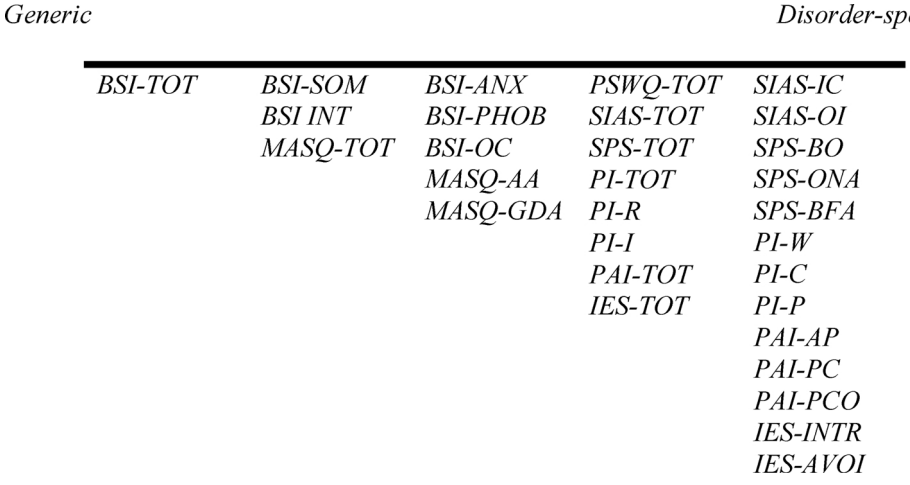

Fig. 1. Position of the different instruments on a dimension from generic to disorder-specific.

subscales of the generic or disorder-specific instruments administered). In the intake interview with a psychiatrist (held prior to the MINI-plus diagnostic interview), the primacy of a Substance Disorder or Personality Disorder was determined by clinical judgment and these patients were referred to other departments of GGZ Rivierduinen. Participants of the study were treated with psychotherapy (mostly, cognitive behavior therapy) frequently in combination with pharmacotherapy (predominantly antidepressants). Before participating, patients were informed that routine outcome monitoring was part of their treatment. Furthermore, patients learned that there was an extensive protocol, which safeguarded the anonymity of patients and participants and ensured proper handling of the data. The Medical Ethical Committee of the LUMC approved this protocol. Patients could object to the use of their data, which resulted in the deletion of their data from the research database. A comprehensive account of the procedures of routine outcome monitoring can be found in de Beurs et al. (2011).

The proportion of male respondents in our sample ranged from 25.6%–42.0% across disorders. The mean age of respondents across disorders was between 33.8 and 39.1 years. BSI total scores at baseline ranged from $M = 0.81$ to $M = 0.97$ across disorders, and MASQ total scores ranged from $M = 203.8$ to $M = 218.9$ across disorders. The

highest mean scores on both instruments occurred in patients with PTSD (see Table 1). Some patients suffered from more than one anxiety disorder, according to the MINI-plus: 2856 (81,2%) had a single diagnosis, 563 (16,0%) suffered from two diagnoses, and 98 (2,8%) from three or four anxiety disorders. Due to comorbidity of anxiety disorders, the data of some patients were included in more than one specific disorder group.

2.3. Procedure

Questionnaires were administered by computer at the clinic or at home via the Internet. The initial assessment, including the MINI-Plus diagnostic interview, was performed by specially trained research nurses or psychologists. Bi-weekly 4-hour sessions were organized in which research nurses were trained in mood, anxiety and somatoform disorders by experts in the field. Furthermore, they were trained in administering and interpreting the MINI-plus by watching video-taped interviews of patients with colleagues and discussing the rating of diagnostic questions until consensus about a proper score had been reached. Also, they all did their first interviews conjointly and closely supervised by a more experienced research nurse. Patients meeting the criteria for more than one DSM-IV disorder were allocated to more than one diagnostic group and hence contributed data to two or more data sets. Respondents completed all questionnaires within one single session. Patients were reassessed every three to four months. As treatment duration varied among patients, the number of reassessments also varied, with some patients being reassessed only once, whereas others were reassessed repeatedly (up to four times). There was a substantial decrease in respondents after each measurement interval (by about 50%). Patients were no longer assessed if they had completed their treatment some time before the re-assessment session (about half of the missing reassessments) or if they were no longer willing to participate in reassessments (the other half; see de Beurs et al., 2011). Thus, for half of the patients the last assessment coincided with their last treatment session. For each pair of instruments that were compared, we considered the first and the last available assessment per patient. Due to the substantial loss of data over successive assessments, we censored measurement trajectories after the fourth assessment. Furthermore, participants were restricted to those with a pre-to-posttest interval of at least three months, but less than three years.

2.4. Instruments

Axis I diagnoses according to the diagnostic and statistical manual of mental disorders (DSM-IV-TR, 2000) were established using the Mini-International Neuropsychiatric Interview-Plus (MINI-Plus; Sheehan et al., 1998). The MINI-Plus has good psychometric properties, with inter-rater reliability ranging from 0.88 to 1.00, test-retest reliability ranging from 0.76 to 0.93, and adequate validity compared to the composite international diagnostic interview (Lecrubier et al., 1997).

In accordance with the Axis-I diagnoses, a personalized selection of self-report instruments and rating scales was administered (de Beurs et al., 2011). A battery of generic and disorder-specific measures was composed based on the outcome of the MINI-plus diagnostic interview: each additional diagnosis implied additional disorder-specific measures. Thus, two generic measures could be compared to six disorder specific instruments for the five anxiety disorders. See Table 2 for an overview of the instruments and their subscales. All instruments and their Dutch translations possess good psychometric properties. The BSI has demonstrated good internal consistency (Cronbach's α ranging from 0,71 to 0,84), sufficient test-retest reliability, inter-item correlations showed sufficient concordance, and convergent and discriminant validity were sufficient as well (de Beurs, Smit, & Comijs, 2005). The MASQ has shown excellent internal consistency (Cronbach's α 0.91 − 0.96), sufficient concordance of inter-item correlations, and sufficient criterion-related validity (de Beurs, den Hollander-Gijsman, Helmich, & Zitman, 2007). The SIAS demonstrated good to excellent internal consistency (Cronbach's α 0.88 − 0.93), sufficient concordance, good test-retest reliability, and good discriminant validity (Brown et al., 1997; de Beurs, Tielen, & Wollmann, 2014; Heimberg, Mueller, Holt, Hope, & Liebowitz, 1993; Mattick & Clarke, 1998). The SPS showed good to excellent internal consistency (Cronbach's α 0.89 − 0.94), sufficient concordance, good test-retest reliability and sufficient discriminant validity (Brown et al., 1997; de Beurs et al., 2014; Heimberg et al., 1993; Mattick & Clarke, 1998). The PSWQ demonstrated good to excellent internal consistency (Cronbach's α 0,90 − 0,95), good convergent and discriminant validity (Kerkhof, Hermans, Figee, & Laeremans, 2000; van Rijsoort, Emmelkamp, & Vervaeke, 1999). The PAI showed good internal consistency (Cronbach's α 0,80 – 0,92), sufficient concordance, and good convergent validity (de Beurs et al., 2005). The PI demonstrated satisfactory to excellent internal consistency (Cronbach's α 0,77 − 0,93), sufficient discriminant and construct validity (Van Oppen, Hoekstra, & Emmelkamp, 1995). Finally, the IES showed good to excellent internal consistency (Cronbach's α 0,90 − 0,95) and good convergent validity (van der Ploeg, Mooren,

Table 1
Participant characteristics.

% (n)	Social phobia	GAD	Panic disorder	OCD	PTSD
N	828	651	944	453	641
Gender (male) n (%)	348 (42.0)	231 (35.5)	369 (39.1)	168 (37.1)	164 (25.6)
Mean age (sd)	33.8 (11.9)	38.9 (13.5)	37.5 (12.4)	35.1 (12.5)	39.1 (12.9)
Mean length pre-post interval in days (sd)	329.8 (196.0)	320.3 (184.4)	318.2 (202.1)	347.3 (209.9)	325.0 (202.4)
MASQ-TOT M (sd)	203.8 (62.8)	205.5 (66.1)	207.2 (70.2)	208.0 (64.5)	218.9 (71.0)
BSI-TOT M (sd)	0.81 (0.66)	0.82 (0.68)	0.84 (0.75)	0.86 (0.71)	0.97 (0.81)

SP = Social Phobia; GAD = Generalized Anxiety Disorder; PD = Panic Disorder; OCD = Obsessive-Compulsive Disorder; PTSD = Post Traumatic Stress Disorder, MASQ-TOT = Total score on the Mood and Anxiety Questionnaire, BSI-TOT = Total score on the Brief Symptom Inventory.

Table 2
Overview of measures used and their subscales.

Instrument	Description	# items	Likert scale	Reference:
Generic				
BSI	Brief Symptom Inventory		0-4	Derogatis & Melisaratos, 1983; de Beurs & Zitman, 2006
BSI-TOT	Total score	53		
BSI-SOM	Somatic Complaints	7		
BSI-INT	Interpersonal Sensitivity	4		
BSI-ANX	Anxiety	6		
BSI-PHOB	Phobic Avoidance	5		
BSI-OC	Obsessive Compulsive			
MASQ	Mood Anxiety Symptoms Questionnaire		0-4	Clark & Watson, 1991; de Beurs et al., 2007
MASQ-TOT	Total score	90		
MASQ-GDA	General Distress Anxiety	11		
MASQ-AA	Anxious Arousal	17		
Disorder specific				
Social Phobia				
SIAS	Social Interaction Anxiety Scale		0-4	Mattick & Clarke, 1998; de Beurs et al., 2014
SIAS-TOT	Total score	20		
SIAS-IC	Initiating Contact	7		
SIAS-OI	Ongoing Interaction	13		
SPS	Social Phobia Scale		0-4	Mattick & Clarke, 1998; de Beurs et al., 2014
SPS-TOT	Total score	20		
SPS-BO	Being Observed	6		
SPS-ONA	Others notice anxiety	6		
SPS-BFA	Being focus of attention	8		
Generalized Anxiety Disorder				
PSWQ	Penn State Worry Questionnaire		1-5	Meyer, Miller, Metzger, & Borkovec, 1990; Kerkhof et al., 2000
PSWQ-TOT	Total score	16		
Panic disorder				
PAI	Panic Appraisal Inventory		0-10	Telch et al., 1989; de Beurs et al., 2005
PAI-TOT	Total score	45		
PAI-AP	Anticipated Panic	15		
PAI-PC	Panic consequences	15		
PAI-PCP	Physical	5		
PAI- PCS	Social	5		
PAI-PCL	Loss of control	5		
PAI-COP	Coping with panic	15		
Obsessive-Compulsive Disorder				
PI	Padua Inventory		0–4	Sanavio, 1988; Van Oppen et al., 1995
PI-TOT	Total Score	60		
PI-I	Impulses	7		
PI-W	Washing	10		
PI-C	Checking	7		
PI-R	Rumination	11		
PI-P	Precision	6		
Posttraumatic Stress Disorder				
IES	Impact of Event Scale		1-3	Horowitz, Wilner, & Alvarez, 1979; Brom & Kleber, 1985
IES-TOT	Total score	15		
IES-INTR	Intrusions	6		
IES-AVOI	Avoidance	7		

Kleber, van der Velden, & Brom, 2004).

2.5. Composition of study samples

Based on the MINI-plus interview, patients were selected who had received one or more diagnoses of SP, GAD, PD, OCD and/or PTSD. Diagnoses could be primary or secondary (e.g. a primary diagnosis of PD or a primary diagnosis of major depressive disorder with a secondary diagnosis of PD). There was overlap in patients between the diagnostic groups due to comorbid anxiety disorders, i.e. patients with more than one anxiety disorder. Thus, having any anxiety disorder made a patient eligible for inclusion in the study and some patients were included in more than one group. Of the patients diagnosed with one or more anxiety disorders, cases with incomplete scores on the BSI, the MASQ, or on the disorder-specific instrument at baseline or at the last available posttest were excluded. To minimize data loss due to exclusion of incomplete cases and to preserve statistical power to find differences, we created two datasets per anxiety disorder: one with complete BSI and disorder-specific total scores, and one with complete MASQ and disorder-specific total scores per anxiety disorder. General characteristics of respondents are presented in Table 1.

2.6. Statistical analysis

Statistical analyses were performed to investigate the overall question whether responsiveness diverged dependent on the specificity of instruments. We distinguished four levels of instrument-specificity: the total score on the generic instruments BSI and MASQ, the anxiety-specific subscale scores of the BSI and MASQ, the total score of disorder-specific instruments, and (if applicable) subscale scores of disorder-specific instruments (see Fig. 1).

Before the analyses were performed, all scores were standardized (on the pretest variance) to arrive at a common metric, resulting in z-scores with a pretest mean score of $M = 0$; $SD = 1$.

For each subtype of anxiety disorder, three types of analyses were performed to investigate differences in responsiveness between pairs of instruments. First, correlations between instruments were calculated to evaluate concordance of scales at each time point (Angst et al., 2008; de Beurs et al., 2012). Second, within group effect sizes (ES) were determined by dividing the difference between pre- and post-test

measurement by the pretest standard deviation (Seidel, Miller, & Chow, 2013). Third, a repeated measures analysis of variance (generalized linear model: GLM) was performed to test for a significant difference in decline of the mean scores over time between instruments (de Beurs et al., 2012) with $MeanZscore = \beta_0 + \beta_1 time + \beta_2 instrument + \beta_3 time * instrument + \varepsilon$ (1). In this model, instrument (a dichotomous variable) indicated a generic or disorder-specific instrument, and time referred to the measurement moment. The coefficient of the interaction term of time*instrument is most relevant to the research question, as it indicates a difference in change over time and, thus, a difference in responsiveness between instruments. The size of this interaction effect was expressed in partial η^2. According to Cohen (1988), $\eta^2 = .02$ corresponds with a small effect, $\eta^2 = .13$ with a medium effect, and $\eta^2 = .26$ with a large effect.

All analyses were performed in SPSS (version 22) with a two-sided significance level < .05, based on scores of the total generic scale, generic subscales, total disorder-specific scores and (if applicable) subscales.

3. Results

3.1. Correlations between generic and disorder-specific instruments

The results of the Pearson correlations of the generic and disorder-specific instruments and the corresponding subscales showed a positive relation between the different instruments (see Table 3). All correlations were significantly different from zero at the .01-level (two-sided). The correlation coefficients of the total instrument scales ranged from $r = 0.43$ to $r = 0.58$ for the baseline measurement, from $r = 0.64$ to $r = 0.74$ for the final measurement, and from $r = 0.46$ to $r = 0.61$ for the difference scores, indicating that these scales measure distinct but substantially associated concepts. At the subscale level, we observed mostly intermediate correlation coefficients with more variation compared to total scales.

Additional analyses revealed, that correlation coefficients tended to increase from initial to final measurement, both on total score and subscale level. This was likely due to an increase in variance over consecutive assessments (some patients improved or recovered, whereas others remained stable). This increase in association over time was most evident in the scores on the disorder-specific instruments.

3.2. Comparative responsiveness between generic and disorder-specific instruments

Next, effect sizes (ES) of the pre-to-post change (first to final assessment) on the various instruments were determined and differential responsiveness was evaluated. To this end, with GLM, the statistical significance of the coefficient of the interaction term between time and instrument was evaluated per instrument pair. The coefficient of the interaction term time * instrument was used as indicator of the difference in responsiveness between instruments. A statistically significant interaction between instrument and time implied a difference between

Table 3
Correlation coefficients (Pearson's r) for the association between change scores from initial to final measurement.

	Generic scales								
	BSI-TOT	BSI-SOM	BSI-INT	BSI-ANX	BSI-PHOB	BSI- OC	MASQ-TOT	MASQ-GDA	MASQ-AA
Disorder-specific scales									
Social phobia									
SIAS-TOT	.59	.30	.59	.45	.54	.50	.54	.44	.33
SIAS-IC	.52	.27	.52	.42	.50	.45	.51	.40	.29
SIAS-OI	.56	.28	.57	.42	.51	.49	.50	.41	.31
SPS-TOT	.58	.37	.56	.49	.55	.49	.54	.48	.41
SPS-BO	.50	.37	.44	.46	.47	.40	.44	.41	.42
SPS-ONA	.52	.33	.49	.42	.53	.41	.48	.42	.35
SPS-BFA	.55	.31	.54	.44	.49	.49	.52	.45	.35
Generalized Anxiety Disorder									
PSWQ-TOT	.50	.28	.43	.43	.39	.53	.55	.49	.35
Panic disorder									
PAI-TOT	.61	.50	.45	.59	.65	.42	.60	.54	.54
PAI-AP	.55	.45	.38	.56	.65	.40	.53	.49	.48
PAI-PC	.50	.40	.40	.43	.47	.40	.47	.42	.45
PAI-PCP	.39	.39	.28	.34	.38	.29	.36	.34	.41
PAI-PCS	.42	.28	.38	.35	.42	.33	.39	.35	.34
PAI-PCL	.47	.34	.35	.42	.39	.40	.44	.39	.38
PAI-PCO	.44	.36	.32	.44	.46	.35	.45	.39	.37
Obsessive-Compulsive Disorder									
PI-TOT	.59	.38	.46	.47	.45	.58	.58	.50	.43
PI-I	.44	.33	.29	.35	.38	.35	.44	.37	.41
PI-W	.26	.18	.20	.18	.20	.26	.26	.21	.22
PI-C	.45	.27	.32	.39	.33	.51	.43	.35	.29
PI-R	.61	.38	.51	.51	.44	.59	.63	.57	.42
PI-N	.36	.24	.28	.23	.28	.35	.33	.29	.26
Post Traumatic Stress Disorder									
IES-TOT	.46	.34	.34	.44	.36	.41	.50	.45	.36
IES-INTR	.48	.36	.35	.45	.37	.44	.52	.47	.38
IES-AVOI	.36	.27	.27	.35	.29	.32	.41	.37	.29

Note. All correlations are significant at the .01 level (2-tailed).
BSI-TOT = BSI total score; BSI-SOM = somatization; BSI-INT = interpersonal sensitivity; BSI-ANX = anxiety; BSI-PHOB = phobic anxiety; BSI-OC = Obsessive Compulsive; MASQ- TOT = MASQ total score; MASQ- GDA = general distress anxiety; MASQ- AA = anxious arousal; SIAS-TOT = SIAS total score; SIAS-IC = initiating contact; SIAS-OI = anxiety in ongoing interaction; SPS-TOT = SPS total score; SPS-BO = being observed; SPS-ONA = others notice anxiety; SPS-BFA = becoming focus of attention; PSWQ-TOT = PSWQ total score; PAI-TOT = PAI-TOT score; PAI-AP = anticipated panic; PAI-PC = total panic consequences; PAI-PCP = panic consequences (physical); PAI-PCS = panic consequences (social); PAI-PCL = panic consequences (loss of control); PAI-PCO = panic coping; PI-TOT = Padua Inventory Total score; PI-I = impulses; PI-W = washing; PI-C = checking; PI-R = rumination; PI-P = precision; IES-TOT = IES total score; IES-INTR = intrusion; IES-AVOI = avoidance.

Table 4
Effect sizes (ES) for the pre-post change on each scale (in bold typeface) and eta2 for pair wise comparisons of BSI versus MASQ per anxiety disorder.

		BSI-TOT	SOM	INT	ANX	PHOB	OC
Social phobia	**ES**	**.82**	**.44**	**.82**	**.69**	**.66**	**.60**
MASQ	TOT	**.91**	.02	.21	.01	.07	.07	.13
	GDA	**.68**	.04	.08	.02	ns	ns	.01
	AA	**.49**	.16	.01	.09	.05	.03	.01
GAD	**ES**	**.81**	**.49**	**.63**	**.82**	**.49**	**.66**
MASQ	TOT	**.92**	.03	.19	.10	.02	.17	.09
	GDA	**.73**	.02	.07	.01	.01	.06	.01
	AA	**.51**	.13	Ns	.01	.11	**ns**	**.03**
Panic disorder	**ES**	**.76**	**.63**	**.51**	**.98**	**.73**	**.58**
MASQ	TOT	**.84**	.02	.06	.13	.03	.02	.11
	GDA	**.76**	Ns	.02	.06	.08	ns	.04
	AA	**.64**	.03	Ns	.02	.16	.01	.01
OCD	**ES**	**.63**	**.39**	**.43**	**.72**	**.44**	**.62**
MASQ	TOT	**.68**	.01	.10	.09	ns	.09	.01
	GDA	**.63**	Ns	.08	.05	.02	.05	ns
	AA	**.42**	.09	Ns	ns	.11	ns	.05
PTSD	**ES**	**.72**	**.49**	**.52**	**.69**	**.46**	**.67**
MASQ	TOT	**.86**	.06	.18	.15	.05	.18	.06
	GDA	**.66**	.ns	.05	.03	ns	.06	ns
	AA	**.53**	.07	Ns	ns	.04	ns	.03

Note. ns = no significant time*instrument interaction at 0.05 level (2-sided). BSI-TOT = BSI Total score; BSI-SOM = somatization; BSI-INT = interpersonal sensitivity; BSI-ANX = anxiety; BSI-PHOB = phobic anxiety; BSI-OC = Obsessive-Compulsive; MASQ-TOT = MASQ total score; MASQ-GDA = general distress anxiety; MASQ-AA = anxious arousal.

instruments in decline of symptom severity. First, we compared the responsiveness of the BSI-total score and the MASQ-total score in the entire sample, irrespective of specific anxiety disorder. Analysis of variance for repeated measures revealed a significant time effect [$F(1, 1893) = 1016{,}26$, $p < .001$; $eta^2 = .349$], instrument effect [$F(1, 1893) = 0{,}080$, $p = .778$; $eta^2 = .000$] and, most importantly, a significant time by instrument effect [$F(1, 1873) = 68{,}91$, $p < .001$; $eta^2 = .035$], with a standardized mean difference of ES = 0.64 for the MASQ and ES = .46 for the BSI, indicating that the MASQ was significantly more responsive than the BSI.

Table 4 summarizes the findings regarding the comparison between the BSI and the MASQ (and their embedded subscales) for the five anxiety disorders. The ES values indicate that the MASQ-total scale was more responsive than the BSI-total scale and all its subscales in all samples, with two exceptions: In PD, the BSI-ANX was more responsive (*ES* = .98 vs. .84, $p < .05$) and in OCD, the BSI-ANX equaled the responsivity of the MASQ-TOT (*ES* = .72 and .68, $p = .26$). Whenever differences in responsiveness were statistically significant, they were usually small.

Table 5 presents pre-post ES in combination with the significance levels of the GLM interaction terms. Fig. 2 depicts mean pretest and final scores on generic and disorder-specific scales for the five samples. The responsiveness of the MASQ total scale equaled or exceeded the responsiveness of the disorder-specific instruments or their subscales for all anxiety disorder groups. Some (subscales of) disorder-specific instruments showed responsiveness similar to that of the MASQ total scale: SPS-BFA in SP, PSWQ-total score in GAD, the PAI-PCO and PAI-total score in PD, the PI-R scale in OCD, and the IES-INTR and IES-total scale in PTSD. Subscales of the MASQ performed worse than the MASQ-total scale in terms of responsiveness, and for most comparisons disorder-specific instruments were more responsive than the MASQ-subscales. The findings for the BSI were more equivocal. For two disorders (SP and OCD), the BSI total score was equal or more responsive than the total or subscales of disorder-specific instruments; in PD the responsiveness of BSI total scores fell in between the total scale and most of the subscales of the disorder-specific instrument. For two disorders (GAD and PTSD), the BSI-TOT was less responsive than the disorder-specific total and less or equally responsive to the disorder-specific subscales. Finally, results of the BSI-subscales were mixed, ranging from more to less responsive when compared to the disorder-specific instruments. In SP, the SPS-BFA subscale seemed to capture change over time equally well as the MASQ or the BSI total score. In GAD, the MASQ and PSWQ total scales were equally responsive and more responsive than the BSI. In Panic Disorder, the BSI-ANX and MASQ-TOT equal the PAI-PCO in responsiveness. In PTSD, the IES-total scale and its IES-INTR subscale appeared most responsive, yet the MASQ showed comparable responsiveness.

4. Discussion

In this study, we examined the responsiveness of generic and disorder-specific instruments in patients with five types of anxiety disorders. Our goal was to investigate whether disorder-specific instruments for anxiety disorders are a valuable (or even necessary) addition to generic instruments for an appropriate assessment of treatment outcome in groups of patients.

Based on ES and significance of GLM interaction effects for the head-to-head comparison of scales, we found small differences in responsivity between the BSI and the MASQ. In all samples, the total score of the MASQ appeared somewhat more responsive compared to the total score of the BSI, which might be explained by the more generic nature of the BSI (more diverse item content and multidimensionality) compared to the more specific MASQ (which specifically measures symptoms of depression and anxiety).

Furthermore, we found that the responsivity of the MASQ and BSI almost equaled the responsivity of disorder-specific scales. The MASQ total scale in particular was equally responsive as disorder-specific instruments and in some instances even more responsive. The BSI total scale or one of the BSI subscales was equally or more responsive than disorder-specific instruments in SP, PD and OCD. Yet, disorder-specific instruments were more responsive than the BSI or its subscales in GAD and PTSD. Overall, we find the MASQ to be most responsive when compared to the BSI and disorder-specific instruments. Furthermore, in two of the five anxiety disorders, the disorder-specific instruments appear slightly more responsive than the BSI. Taken together, these findings suggest limited incremental information value of disorder-specific instruments over the MASQ and BSI for measuring change. Yet, it has to be noted that the MASQ, may best be described as falling in-between disorder-specific and purely generic instruments. Hence, the instrument, which performed best in terms of responsiveness, appears to be neither purely disorder-specific nor purely generic.

Furthermore, Pearson correlations indicated substantial association between change scores on subscales purported to measure similar concepts, which provides evidence for convergent validity [for example: PAI physical concerns (PAI-PCP), BSI somatization scale (BSI-SOM), and the MASQ Anxious Arousal scale (MASQ-AA); PAI social concerns (PAI-PCS), and BSI interpersonal sensitivity symptoms (BSI-INT); BSI anxiety scale (BSI-ANX) and MASQ General Distress Anxiety scale (MASQ-GDA)]. Correlation coefficients appear lower for change scores on scales measuring distinct aspects of symptomatology, thus demonstrating discriminant validity. Both groups of findings are supportive of the construct validity of the measures and their subscales. Furthermore, correlations appear higher for total scales than for subscales (i.e. high correlations of PA-TOT vs. BSI-TOT and MASQ-TOT; IES-TOT vs. BSI-TOT and MASQ-TOT). Convergent and divergent validity of the measures is supported, as correlation coefficients appear higher for subscales which intent to measure similar concepts compared to subscales measuring dissimilar concepts.

Strengths of the current study were the large sample sizes and the coverage of five different types of anxiety disorders. Furthermore, data were gathered in a real life setting with patients undergoing treatment as provided in regular clinical practice, strengthening the external validity of the findings. Yet, the study also has limitations. First, respondents were recruited in the Netherlands and locations were limited

Table 5
Effect sizes for the pre-post change (ES in bold typeface) and eta2 for pair wise comparisons of generic and disorder-specific scales per anxiety disorder.

ES/significance (eta^2)		BSI TOT	SOM	INT	ANX	PHOB	OC		MASQ TOT	GDA	AA
Social phobia	**ES:**	**.83**	**.44**	**.82**	**.69**	**.66**	**.60**	**ES**	**.91**	**.68**	**.49**
SIAS-TOT	**.80**	ns	.10	Ns	.01	.02	.03	**.79**	.01	.01	.08
SIAS-IC	**.80**	ns	.09	Ns	.02	.02	.03	**.80**	.01	.01	.07
SIAS-OI	**.72**	.01	.06	.01	.01	.01	.02	**.72**	.03	ns	.04
SPS-TOT	**.74**	.01	.08	.01	ns	.01	.01	**.73**	.03	ns	.06
SPS-BO	**.50**	.11	ns	.09	.03	.03	.01	**.50**	.13	.03	Ns
SPS-ONA	**.55**	.08	.01	.07	.01	.01	Ns	**.55**	.11	.02	Ns
SPS-BFA	**.84**	ns	.11	Ns	.03	.03	.02	**.84**	ns	.02	.09
GAD	**ES**	**.81**	**.49**	**.62**	**.82**	**.49**	**.66**	**ES**	**.92**	**.73**	**.51**
PSWQ-TOT	**.94**	.01	.10	.07	.01	.12	.06	**.93**	ns	.03	.10
Panic disorder	**ES**	**.76**	**.63**	**.50**	**.98**	**.73**	**.58**	**ES**	**.84**	**.76**	**.64**
PAI-TOT	**.87**	.02	.05	.11	.01	.03	.08	**.88**	ns	.01	.05
PAI-AP	**.66**	.01	ns	.02	.01	.01	.01	**.66**	.03	.01	Ns
PAI-PC	**.55**	.04	.01	Ns	.03	.03	Ns	**.55**	.07	.03	.01
PAI-PCP	**.47**	.08	.03	.06	.16	.06	.01	**.46**	.10	.06	.03
PAI-PCS	**.35**	.14	.06	.12	.22	.12	.05	**.34**	.17	.11	.07
PAI-PCL	**.49**	.07	.02	.05	.16	.05	.01	**.48**	.10	.05	.02
PAI-PCO	**.90**	.01	.04	.09	.02	.02	.06	**.90**	ns	.01	.04
OCD	**ES**	**.62**	**.39**	**.43**	**.72**	**.43**	**.62**	**ES**	**.68**	**.63**	**.42**
PI-TOT	**.49**	.03	.01	Ns	.05	ns	.02	**.49**	.05	.02	Ns
PI-I	**.24**	.16	.03	.04	.04	.04	.12	**.24**	.17	.13	.04
PI-W	**.17**	.21	.06	.08	.08	.08	.17	**.17**	.20	.17	.07
PI-C	**.40**	.07	ns	Ns	ns	ns	.06	**.40**	.08	.05	Ns
PI-R	**.59**	ns	.04	.03	.03	.03	Ns	**.60**	.01	ns	.03
PI-P	**.30**	.11	ns	.02	.02	.02	.09	**.31**	.11	.09	.01
PTSD	**ES**	**.71**	**.48**	**.52**	**.68**	**.45**	**.67**	**ES**	**.86**	**.66**	**.53**
IES-TOT	**.90**	.02	.10	.08	.03	.12	.03	**.86**	ns	.03	.07
IES-INTR	**.91**	.03	.11	.09	.03	.12	.04	**.87**	ns	.03	.08
IES-AVOI	**.78**	ns	.05	.04	ns	.06	.01	**.74**	.01	ns	.03

Note. ns = no significant time*instrument interaction at 0.05 level (2-sided).
BSI-TOT = BSI total score; BSI-SOM = somatization; BSI-INT = interpersonal sensitivity; BSI-ANX = anxiety; BSI-PHOB = phobic anxiety; BSI-OC = Obsessive Compulsive; MASQ- TOT = MASQ total score; MASQ- GDA = general distress anxiety; MASQ- AA = anxious arousal; SIAS-TOT = SIAS total score; SIAS-IC = initiating contact; SIAS-OI = anxiety in ongoing interaction; SPS-TOT = SPS total score; SPS-BO = being observed; SPS-ONA = others notice anxiety; SPS-BFA = becoming focus of attention; PSWQ-TOT = PSWQ total score; PAI-TOT = PAI-TOT score; PAI-AP = anticipated panic; PAI-PC = total panic consequences; PAI-PCP = panic consequences (physical); PAI-PCS = panic consequences (social); PAI-PCL = panic consequences (loss of control); PAI-PCO = panic coping; PI-TOT = Padua Inventory Total score; PI-I = impulses; PI-W = washing; PI-C = checking; PI-R = rumination; PI-P = precision; IES-TOT = IES total score; IES-INTR = intrusion; IES-AVOI = avoidance.

to two mental health care institutions. Hence, national or local culture (or translations of the measures into Dutch) may have affected the outcomes of the study and findings may not generalize beyond the Dutch context. Second, only patients with anxiety disorders were included. Possibly, results may differ in other patient populations (i.e. depression or personality disorders) where different instruments are used and different outcomes are central to treatment success. Furthermore, the exclusion criteria of a primary substance disorder or personality disorder were evaluated by clinical judgment with no further check on the reliability of that diagnosis. Moreover, in this study data were obtained in a real-life setting. Treatments were predominantly generic/broad based CBT, which may have influenced the findings of the present study in favor of generic scales. In addition, many patients had comorbid (anxiety) disorders. The composition of the datasets, stemming from a rather mixed group of patients, may have biased the results in favor of generic instruments as well. Nevertheless, the findings have high external validity, as in everyday clinical care comorbidity is the rule rather than the exception. In addition, pre-post intervals of the data were often considerably large. Furthermore, the present findings are limited to the disorder-specific instruments that were evaluated; other disorder-specific instruments may still be more responsive than generic scales. Finally, all our conclusions pertain to an evaluation of treatment outcome at the group level and not at the level of the individual patient. For monitoring of individual patients, which is the primary aim of routine outcome monitoring, disorder-specific instruments may still be of value. Evaluation of outcomes at a higher level of aggregation is a secondary aim of routine outcome monitoring (aimed at benchmarking and effectiveness research). Based on the current findings, we can conclude that the MASQ, and, to a certain extent the BSI as well, suffice for this secondary aim, at least within this subset of anxiety disorders.

The MASQ and BSI appear sufficiently responsive to changes of symptoms during treatment and they appear equally suited as disorder-specific instruments to be used at the group level. In the case of OCD, the MASQ total scale even appears better suited than the total and all subscales of the disorder-specific Padua Inventory. In the other samples, the BSI TOT and the BSI ANX scale appear better suited than five of the six disorder-specific subscales. To limit the burden to patients, merely administering the MASQ or BSI in a group of anxiety disorder patients is feasible. Besides a low burden to patients, use of these instruments provides the advantage of comparability across (mental) disorders while not having to sacrifice responsiveness of the measure or the ability to detect clinical changes.

Based on the current study, we suggest further research to be directed at mental health disorders other than anxiety disorders, to provide additional support for our conclusions and to investigate the generalizability of the present findings to other disorders. Furthermore, it may be interesting to replicate the current study in an international setting with a larger number of mental health care providers involved, to exclude possible effects of local or national culture. In addition, it would be interesting to investigate differences in response over time for both types of measures. Potentially, improvement is sooner detected on disorders-specific measures than on generic measures. Furthermore, based on the calculated effect sizes within this study, one could evaluate the significance of differences between the generic instruments as well. It may be valuable to further investigate these differences in

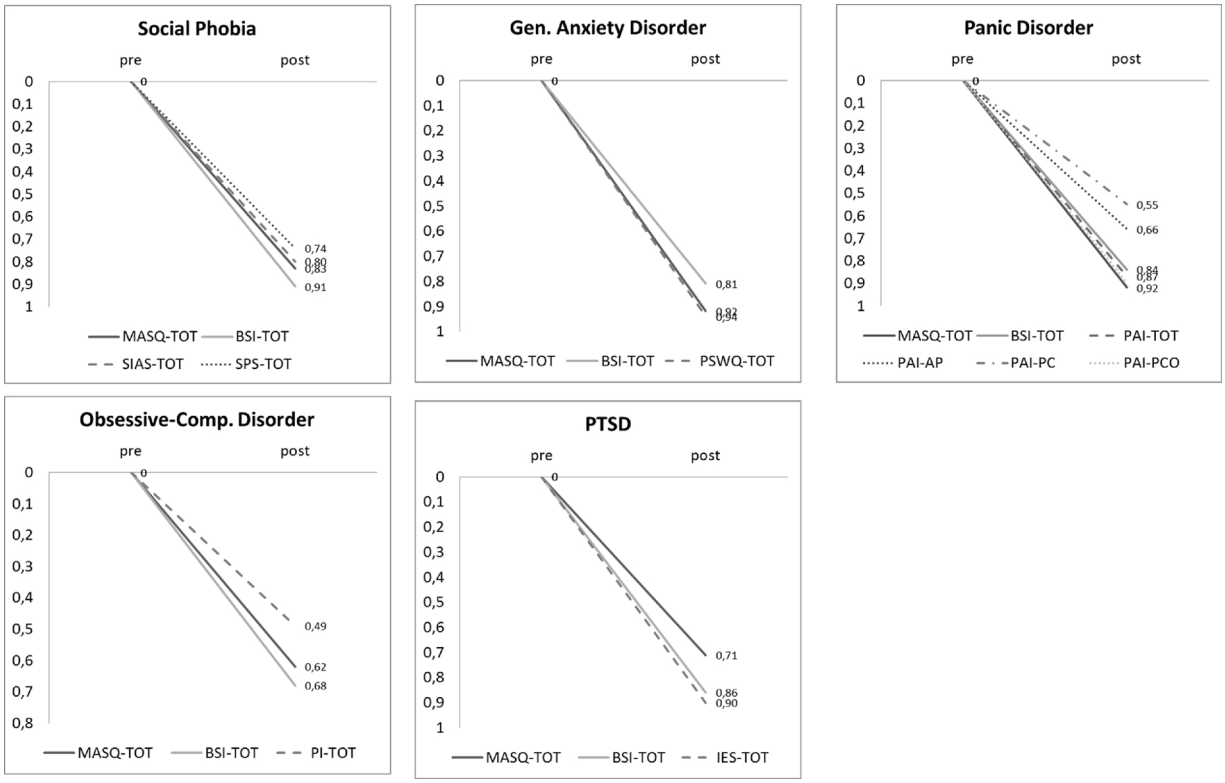

Fig. 2. Change from pretest tot posttest in standardized scores on generic and disorder specific outcome measures for five anxiety disorders.

responsiveness as to being able to express a preference of a generic measure.

By and large the present findings do not fully support our hypothesis of greater responsivity of disorder-specific scales and are not fully in line with common contention. No clear evidence was found for superior responsivity of disorder-specific measures over generic measures when considering groups of patients. Yet, we found the MASQ, which contains generic and disorder-specific characteristics, to perform best in terms of responsiveness. For the assessment of individual patients, disorder specific measures may still yield additional information, which may be overlooked when only the MASQ or BSI are used. In planning research or in evaluation of treatment outcome in everyday clinical practice, one has to appraise whether the increase in information value justifies the extra burden on the patient of administrating several assessment instruments.

Funding

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

Summary declaration of interest

Saskia Schawo (SS), Ingrid Carlier (IC), Albert van Hemert (AvH) and Edwin de Beurs (EdeB) declare to have no conflict of interest.

Acknowledgments

We gratefully acknowledge the mental health care provider GGZ Rivierduinen and the patients participating in this study.

References

Angst, F., Verra, M. L., Lehmann, S., & Aeschlimann, A. (2008). Responsiveness of five condition-specific and generic outcome assessment instruments for chronic pain. *BMC Medical Research Methodology, 8*, 26. https://doi.org/10.1186/1471-2288-8-26.

Bessette, L., Sangha, O., Kuntz, K. M., Keller, R. B., Lew, R. A., Fossel, A. H., ... Katz, J. N. (1998). Comparative responsiveness of generic versus disease-specific and weighted versus unweighted health status measures in carpal tunnel syndrome. *Medical Care, 36*(4), 491–502.

Bombardier, C., Melfi, C. A., Paul, J., Green, R., Hawker, G., Wright, J., ... Coyte, P. (1995). Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. *Medical Care, 33*(4), AS131–AS144.

Brom, D., & Kleber, R. J. (1985). De schokverwerkings lijst [The Impact of Events Scale]. *Nederlands Tijdschrift voor de Psychologie, 40*, 164–168.

Brown, E. J., Turovsky, J., Heimberg, R. G., Juster, H. R., Brown, T. A., & Barlow, D. H. (1997). Validation of the social interaction anxiety scale and the social phobia scale across the anxiety disorders. *Psychological Assessment, 9*(1), 21–27.

Carlier, I. V., Kovács, V., Noorden, M. S., Feltz-Cornelis, C., Mooij, N., Schulte-van Maaren, Y. W., ... Giltay, E. J. (2017). Evaluating the responsiveness to therapeutic change with routine outcome monitoring: A comparison of the symptom questionnaire-48 (SQ-48) with the Brief Symptom Inventory (BSI) and the Outcome Questionnaire-45 (OQ-45). *Clinical Psychology & Psychotherapy, 24*(1), 61–71.

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry, 23*(4), 318–327.

Clark, L. A., & Watson, D. (1991). A tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100*(3), 316–336.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence.

Damiano, A. M., Steinberg, E. P., Cassard, S. D., Bass, E. B., Diener-West, M., Legro, M. W., ... Kolb, M. (1995). Comparison of generic versus disease-specific measures of functional impairment in patients with cataract. *Medical Care, 33*(4), AS120–AS130.

de Beurs, E., Tielen, D., & Wollmann, L. (2014). The dutch social interaction anxiety scale and the social phobia scale: Reliability, validity, and clinical utility. *Psychiatry Journal, 2014*.

de Beurs, E., Barendregt, M., Flens, G., van Dijk, E., Huijbrechts, I., & Meerding, J. W. (2012). Vooruitgang in de behandeling meten-Een vergelijking van vragenlijsten voor zelfrapportage [A comparison of Self-report Questionnaires for treatment outcome]. *Maandblad Geestelijke Volksgezondheid, 67*, 259–270.

de Beurs, E., den Hollander-Gijsman, M. E., van Rood, Y. R., van der Wee, N. J., Giltay, E. J., van Noorden, M. S., & Zitman, F. G. (2011). Routine outcome monitoring in the Netherlands: Practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clinical Psychology & Psychotherapy, 18*, 1–12. https://doi.org/10.1002/cpp.696.

de Beurs, E., den Hollander-Gijsman, M. E., Helmich, S., & Zitman, F. G. (2007). The tripartite model for assessing symptoms of anxiety and depression: Psychometrics of the Dutch version of the mood and anxiety symptoms questionnaire. *Behaviour Research and Therapy, 45*(7), 1609–1617.

de Beurs, E., Meesters, Y., Vissers, E., Schoevers, R. A., Carlier, I. V., & van Hemert, A. M.

(2018). Comparative responsiveness of generic versus disorder-specific instruments for depression: An assessment in three longitudinal datasets. *Journal of Depression & Anxiety, 36*, 93–102.
de Beurs, E., Smit, J. H., & Comijs, H. (2005). De Paniek Opinie Lijst (POL). De betrouwbaarheid en validiteit van een cognitieve maat voor paniekstoornis [The Panic Appraisal Inventory (PAI). The reliability and validity of a cognitive measure for panic disorder]. *Gedragstherapie, 38*, 139–152.
de Beurs, E., & Zitman, F. G. (2006). De Brief Symptom Inventory (BSI): De betrouwbaarheid en validiteit van een handzaam alternatief voor de SCL-90 [The Brief Symptom Inventory: the reliability and validity of an alternative to the SCL-90]. *Maandblad Geestelijke Volksgezondheid, 61*, 120.
Derogatis, L. R., & Melisaratos, N. (1983). The brief symptom inventory: An introductory report. *Psychological Medicine, 13*(3), 595–605.
Engel, S. G., Adair, C. E., Hayas, C. L., & Abraham, S. (2009). Health-related quality of life and eating disorders: A review and update. *The International Journal of Eating Disorders, 42*(2), 179–187.
Hay, P. J., & Mond, J. (2005). How to 'count the cost'and measure burden? A review of health-related quality of life in people with eating disorders. *Journal of Mental Health, 14*(6), 539–552.
Heimberg, R. G., Mueller, G. P., Holt, C. S., Hope, D. A., & Liebowitz, M. R. (1993). Assessment of anxiety in social interaction and being observed by others: The Social Interaction Anxiety Scale and the Social Phobia Scale. *Behavior Therapy, 23*(1), 53–73.
Horowitz, M. J., Wilner, N. R., & Alvarez, W. (1979). Impact of Event Scale: A measure of subjective stress. *Psychosomatic Medicine, 41*(3), 209–218.
Kerkhof, A. J., Hermans, D., Figee, A., & Laeremans, I. (2000). De Penn State Worry Questionnaire en de Worry Domains Questionnaire: Eerste resultaten bij Nederlandse en Vlaamse klinische en poliklinische populaties [The Penn State Worry Questionnaire and the Worry Domains Questionnaire: First results within Dutch and Flamisch inpatient and outpatient populations]. *Gedragstherapie, 33*(2), 135–145.
Lambert, M. J., & Finch, A. E. (1999). The outcome questionnaire. In M. E. Maruish (Ed.). *The use of psychological testing for treatment planning and outcomes assessment* (pp. 831–869). (2nd ed.). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C., & Yanchar, S. C. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology & Psychotherapy: An International Journal of Theory and Practice, 3*(4), 249–258.
Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., ... Dunbar, G. C. (1997). The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Psychiatry, 12*(5), 224–231.
Matthey, S., Fisher, J., & Rowe, H. (2013). Using the Edinburgh postnatal depression scale to screen for anxiety disorders: Conceptual and methodological considerations. *Journal of Affective Disorders, 146*(2), 224–230.
Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy, 36*(4), 455–470.
Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn state worry questionnaire. *Behaviour Research and Therapy, 28*(6), 487–495.
Mitchison, D., Hay, P., Engel, S., Crosby, R., Le Grange, D., Lacey, H., ... Touyz, S. (2013). Assessment of quality of life in people with severe and enduring anorexia nervosa: A comparison of generic and specific instruments. *BMC Psychiatry, 13*(1), 284.
Patrick, D. L., & Deyo, R. A. (1989). Generic and disease-specific measures in assessing health status and quality of life. *Medical Care, 27*(3), 217–232.
Sanavio, E. (1988). Obsessions and compulsions: The Padua inventory. *Behaviour Research and Therapy, 26*(2), 169–177.
Schibbye, P., Ghaderi, A., Ljótsson, B., Hedman, E., Lindefors, N., Rück, C., ... Kaldo, V. (2014). Using early change to predict outcome in cognitive behaviour therapy: Exploring timeframe, calculation method, and differences of disorder-specific versus general measures. *PloS One, 9*(6), e100614.
Seidel, J. A., Miller, S. D., & Chow, D. L. (2013). Effect size calculations for the clinician: Methods and comparability. *Psychotherapy Research, 24*(4), 470–484.
Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., ... Dunbar, G. C. (1998). The Mini-International Neuro-psychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric inter-view for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry, 59*(Suppl. 20), 22–33.
Stucki, G., Liang, M. H., Fossel, A. H., & Katz, J. N. (1995). Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. *Journal of Clinical Epidemiology, 48*(11), 1369–1378.
Telch, M. J., Brouillard, M., Telch, C. F., Agras, W. S., & Taylor, C. B. (1989). Role of cognitive appraisal in panic-related avoidance. *Behavior Research Therapy, 27*(4), 373–383.
van der Mheen, M., ter Mors, L. M., van den Hout, M. A., & Cath, D. C. (2017). Routine outcome monitoring bij de behandeling van angststoornissen: diagnosespecifieke versus generieke meetinstrumenten [Routine outcome monitoring within the treatment of anxiety disorders: disorder-specific versus generic instruments]. *Tijdschrift voor Psychiatrie* (forthcomming).
van der Ploeg, E., Mooren, T., Kleber, R. J., van der Velden, P. G., & Brom, D. (2004). Construct validation of the Dutch version of the impact of event scale. *Psychological Assessment, 16*(1), 16.
Van Oppen, P., Hoekstra, R. J., & Emmelkamp, P. M. (1995). The structure of obsessive-compulsive symptoms. *Behaviour Research and Therapy, 33*(1), 15–23.
van Rijsoort, S., Emmelkamp, P., & Vervaeke, G. (1999). The Penn State Worry Questionnaire ad the Worry Domains Questionnaire: structure, reliability and validity. *Clinical Psychology & Psychotherapy, 6*, 297–307.