NormQuest

Reference Values for ROM Instruments and Questionnaires

Yvonne WM Schulte-van Maaren

Het onderzoek is tot stand gekomen met financiële steun van Rivierduinen, organisatie voor geestelijke gezondheidszorg, Leiden.

Lay Out by: Denise Admiraal, Amsterdam
Cover design: Denise Admiraal
Afbeelding cover: Rene Magritte
Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress
Published by: Uitgeverij BOXPress, 's-Hertogenbosch

ISBN: 978-90-8891-788-2

NormQuest

Reference Values for ROM Instruments and Questionnaires

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 21 januari 2014
klokke 13.45 uur

door
Yvonne Wilhelmina Maria Schulte-van Maaren
geboren te Oosterhout
in 1954

Promotiecommissie

Promotores
Prof. dr. F.G. Zitman
Prof. dr. A.M. van Hemert

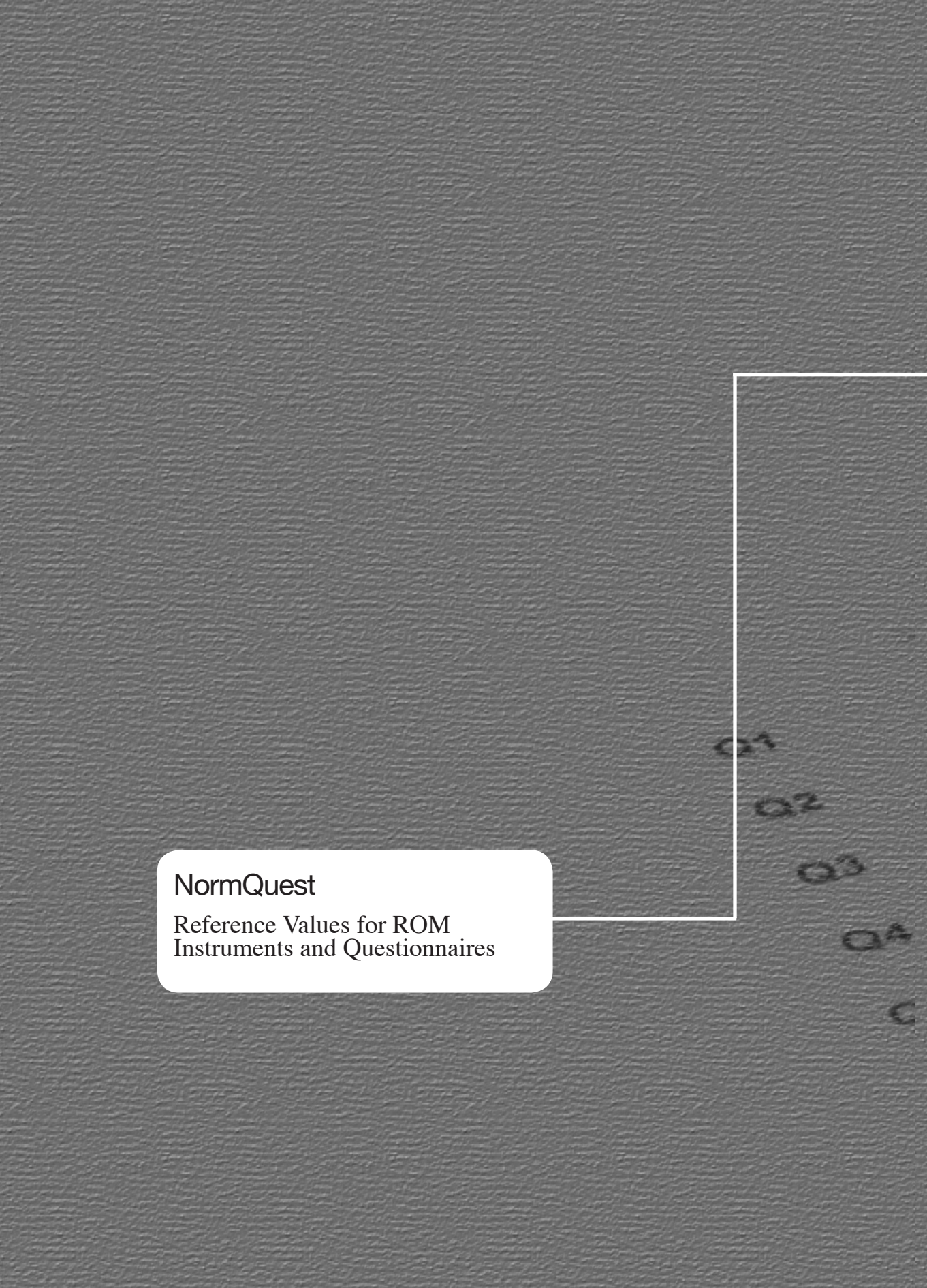
Copromotores
Dr. I.V.E. Carlier
Dr. E.J. Giltay

Overige leden
Prof. dr. J. Gussekloo
Prof. dr. W. Heiser
Dr. E. de Beurs

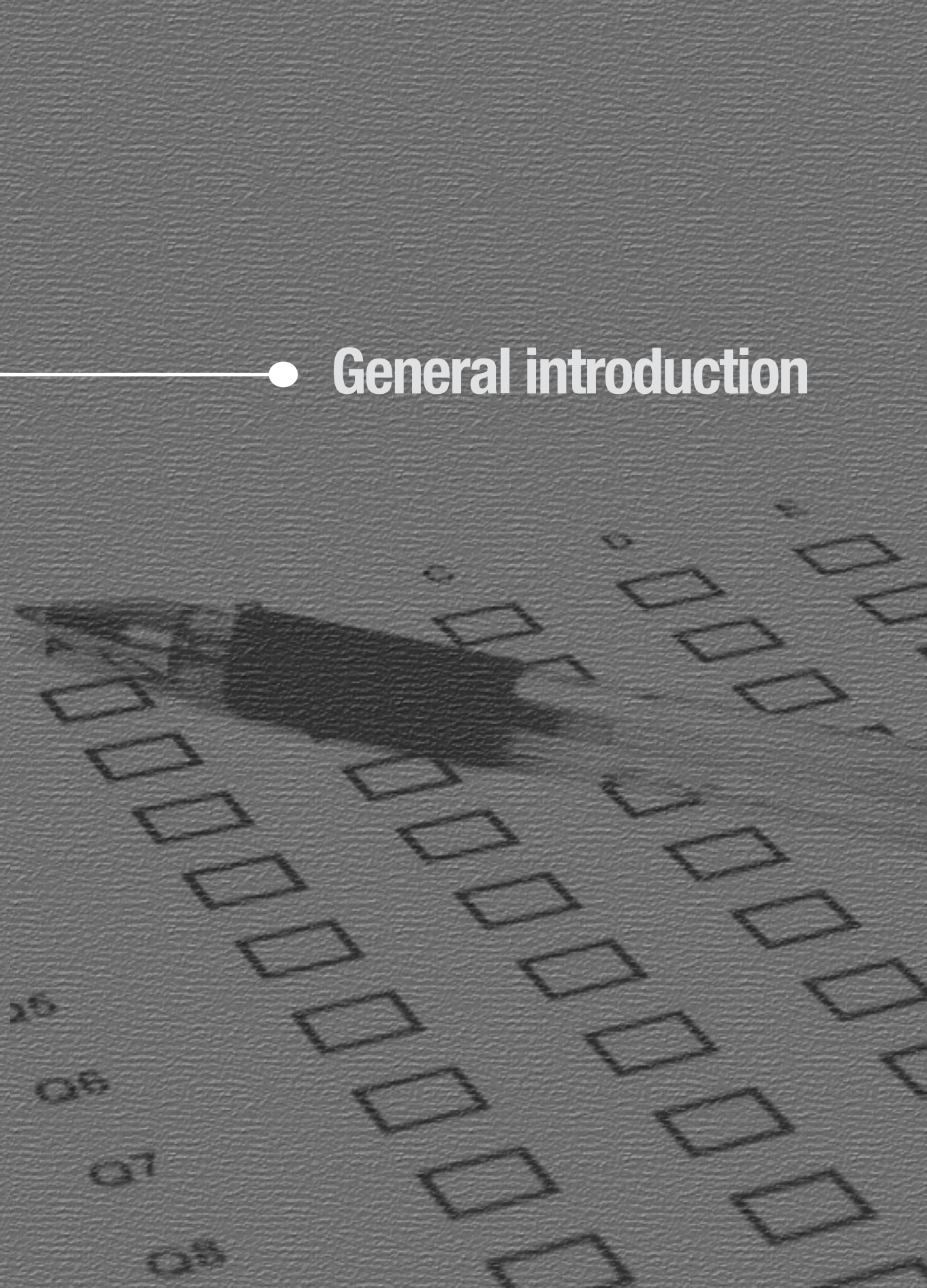
Een van de meest oprechte vormen van respect is echt te luisteren naar wat een ander te zeggen heeft.

Bryant H. McGill (2001).

Contents


NormQuest
Reference Values for ROM
Instruments and Questionnaires

General introduction



INTRODUCTION

In clinical psychiatry it is common practice that the clinical effectiveness of a treatment is judged by the health care professionals and patients. Routine Outcome Monitoring (ROM) can provide exact and valuable additional information about this clinical effectiveness. ROM is a measurement and feedback system, facilitating the systematic evaluation of a psychiatric patient's treatment response during the course of treatment in routine clinical practice. Measuring progress and providing feedback is beneficial to the treatment, both for the clinician and the patient. This feedback is facilitated by the application of reference values in combination with ROM scores. Reference values may quantify the patient progress in therapy and support decisions on continuing, altering or terminating treatment can be considered.

A case

A 64-year old female inpatient was diagnosed with a 7 year history of depression and anxiety. Her problems had started rather abrupt after marital problems that resulted in divorce. Her past medical history included agoraphobia and orthostatic hypotension. Several times she was treated for anxiety and depression with psychotherapy and several antidepressants, either as inpatient or outpatient. Because of severe depression with psychotic features and resistance to antidepressant treatment she was admitted to the Leiden University Medical Centre (LUMC). She was treated with Electroconvulsive Therapy (ECT) unilaterally and her depression went into remission. Depression severity was monitored during the treatment through clinical judgement and ROM. Depression symptom scores are depicted in the graph, showing a slow but steady decline of the symptom severity, assessed through the observer-rated Montgomery-Äsberg Depression Rating Scale (MADRS), where a higher score means more psychopathology (see Figure 1.1).

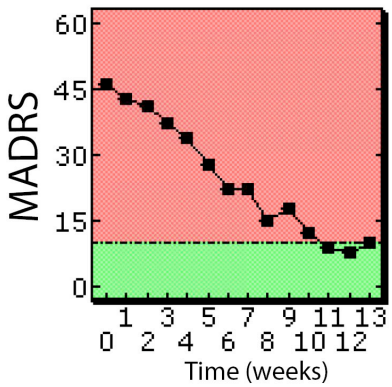


Figure 1.1. ROM graph of MADRS scores of 12 consecutive assessments of an ECT treated patient diagnosed with major depressive disorder.

The provided ROM scores in the above case need interpretation. The baseline MADRS score (week 0, first consultation or admission) matches a diagnosis of depression that was previously established by a clinical interview in combination with the Mini-International Neuropsychiatric Interview-Plus (MINI-Plus; [1]): a severe depression in this case. The consecutive scores (week 1 through 12) depict the course of the symptom severity, supporting the evaluation of the treatment effect (outcome): has the patient deteriorated, improved, not changed, or recovered? In this case a steady improvement can be seen. A key question for the therapist is: when is the patient sufficiently recovered to make the next step in the treatment? One approach that can support such a decision is that the ratings can be compared to those of a normal population. When scoring below a certain cut-off value, the patient is no longer dissimilar from the reference population, and it could be argued that it is legitimate to make a start shifting treatment towards interventions aimed at relapse prevention and ultimately to refer the patient back to her general practitioner (GP). Evidence based cut-off values for commonly used ROM questionnaires, such as the MADRS, can support clinical decisions. These cut-off values can be derived from the distributions of scores from the healthy general population and from patient populations. Cut-off values and additional measures of score distributions are referred to here as reference values.

To provide empirical based reference values for ROM questionnaires, the NormQuest [i.e., quest for norms] study was initiated in 2008 by the LUMC and the regional mental health care provider Rivierduinen. This thesis aims to present these reference values that can be used to support clinical evaluations in the referral and treatment of patients with mood, anxiety, and somatoform (MAS) disorders. Reference values comprise cut-off values, marking the difference between the patient population ('psychiatrically ill') and the reference population ('healthy').

Currently, it is common practice that the clinical effect of an individual treatment is judged qualitatively by the health care professionals and patients. The application of ROM in combination with reference values may facilitate decision making. Ideally, they provide standardized yard-sticks to assess whether the patient's severity of symptoms has been reduced, whether the patient's level of functioning has improved over time and whether therapy has moved someone outside the range of the patient population and within the range of the reference population.

ROUTINE OUTCOME MONITORING (ROM)

ROM provides health care professionals and patients with information relevant to the patient's progress [2]. Diagnosis, monitoring of treatment, and communication between clinician and patient can be improved by ROM [3]. A range of objective, standard outcome measures (self-report questionnaires and observational instruments) are an essential part of ROM. A practical ROM-strategy was implemented in the department of psychiatry of the Leiden University Medical Center (LUMC) and in the outpatient department of the regional mental health care provider Rivierduinen from 2002 onwards (see Box 1).

ROM questionnaires should be clinically relevant, sensitive to change, and minimally burdensome to patient, staff and organization [4]. Therefore, the selection of questionnaires should be based on validity, reliability, availability of reference data, but also on costs. With test characteristics being equal, public domain questionnaires that are free of charge are preferred over copyrighted questionnaires that are commercially exploited. In the context of ROM, there can be serious economic obstacles to the required frequent assessments that are intended for all patients. So, there is an urgent need for the development of public domain questionnaires [5,6].

Questionnaires for ROM comprise both generic and specific ones. Generic measures are used for the assessment of general psychopathology, distress, or general functioning. Since they are, in principle, applicable to all patients, they allow for comparison of treatment outcomes among all patients, irrespective of specific disorders. Generic questionnaires allow statements about the therapy effect regardless of the diagnosis and they are applicable for patients with more than one condition. Furthermore, they facilitate comparisons between different patient groups [7]. Disease-specific measures focus on particular symptoms relevant to a single disorder and are administered only to those patients meeting criteria for the disorder at hand. They are more sensitive to changes in outcome due to treatment as they assess the intensity of the symptoms that the patient suffers from and the specific treatment targets [4,8].

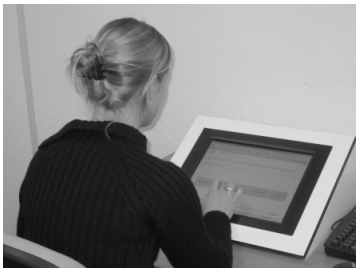
In addition to clinical applications, treatment outcome data can also be relevant to researchers and managers. Research is constantly searching to develop new treatments and these treatments require clinical effectiveness research, which can be facilitated by outcome data. Additionally, researchers can use outcome data for basic research into factors impacting upon outcomes [9] and psychometric research [8,10-14]. For managers, data can provide insight in the quality level of the mental health care by comparing outcomes on differential effectiveness of various treatment programs, locations, departments or even therapists (benchmarking).

BOX 1. ROM in the Leiden University Medical Center & Rivierduinen (courtesy M. van Noorden)

In spring 2002, the Regional Mental Health Provider 'Rivierduinen' (an institute serving a region with more than 1 million inhabitants) and the Department of Psychiatry of the Leiden University Medical Center (LUMC) started collaboration for routine assessment of the DSM-IV diagnosis as well as the symptom severity, well-being and health status at time of the first interview of outpatients referred to Rivierduinen.

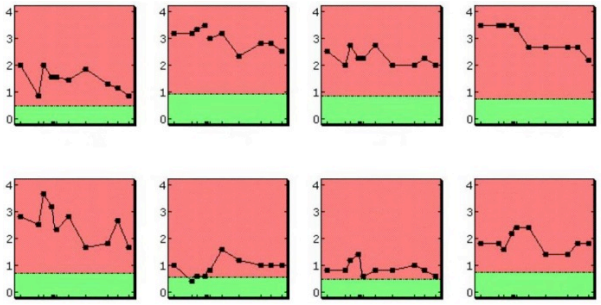
At the start, ROM was restricted to patients referred for treatment of mood, anxiety, and somatoform (MAS) disorders. These patients form a relatively homogenous group with substantial mutual comorbidity (Kessler et al., 1996) and they mainly receive outpatient care. To be eligible, patients had to have sufficient mastery of the Dutch language and had to be able to complete self-report instruments. Patients who are considered (by their clinician) to be too ill to complete questionnaires or refuse to be assessed are excluded from ROM assessment.

All patients are assessed by an independent psychiatric research nurse at the start, and during follow up at intervals of three to four months, at the beginning of a new treatment step and at the end of the treatment.



During the first session, a standardised diagnostic interview is administered and observer- and self-reported ratings are determined. At baseline the Axis-I diagnosis according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) is established using the Mini-International Neuropsychiatric Interview-plus (MINI-plus, Sheehan et al., 1998). The interviews are performed by psychiatric research nurses who have been extensively trained and supervised. The Dimensional Assessment of Personality Pathology (DAPP-SF) is administered to assess maladaptive personality traits (Livesley et al., 2006; van Kampen et al., 2008). Until now, in ROM no detailed treatment information is available.

Subsequently, a number of symptom severity rating scales are administered at baseline and are also completed at each re-assessment to allow for the evaluation of treatment outcome. Together, these instruments cover change in three areas of functioning: symptom reduction, increased wellbeing, and improvement in general life functioning (Sperry et al., 1996). They are commonly used in treatment-outcome research and have good psychometric properties as evidenced by national and international publications (an overview of instruments used is available at http://www.lumc.nl/psychiatry/ROM-instruments). Outcome is assessed by patients' self-report and by an independent assessor (observer-rated), and includes both generic and disorder-specific measures. Clinicians receive a report on the results of the baseline assessments as well as follow-up reporting on treatment outcome in the above mentioned domains. Results of the assessments are provided in detail by the research nurses as well as in a summarised form. The summaries facilitate clinicians to discuss the results with their patients and use them as a tool to evaluate the treatment. Results are also used, in an anonymous form, for scientific purposes.



Since ROM-data are primarily being used by clinicians and patients to monitor treatment progress, no specific informed consent is needed. The use of anonymized data for research purposes has been approved by the Medical Ethical Committee of the LUMC.

MOOD, ANXIETY AND SOMATOFORM (MAS) DISORDERS

There are many different categories of psychiatric disorders for which ROM could be used to systematically evaluate a patient's treatment. We focused on mood, anxiety, and somatoform (MAS) disorders. The majority of patients of the LUMC and a substantial number in Rivierduinen are treated for these disorders. Estimates of different prevalence proportions for mood and anxiety disorders are relatively high [15-19], as can be seen in Table 1.1. Unfortunately no data are available for somatoform disorders.

Table 1.1. Lifetime-, 12 month, and point prevalence rates of common mood and anxiety disorders* in the Netherlands in weighted percentages.

	Prevalence rates		
	Lifetime*	**12-month***	**Point**
Any mood disorder	19.6	6.9	4.1
- Major Depression	17.0	5.5	2.9
- Dysthymic Disorder	3.9	1.6	0.8
Any anxiety disorder+	19.4	11.3	5.5
- Panic Disorder	3.8	1.7	2.7
- Social Phobia	8.5	4.3	0.8
- Obsessive-Compulsive Disorder	0.9	0.5	0.5
- Generalized Anxiety Disorder	3.4	1.4	0.8

Lifetime- and 12-month prevalence rates based on the Netherlands Mental Health Survey and Incidence Studies NEMESIS-1 and NEMESIS-2 [15,16]
Point prevalence rates in a GP consulting population based on De Waal et. al., 2004 [17]
* No data were ascertained for somatoform disorders
+ No data available for post-traumatic stress disorder PTSD

MAS disorders are the most frequently observed mental disorders in primary health care [20,21]. The disease burden is very large, with depression as the most important single contributor to the global burden of disease [22]. MAS disorders frequently occur as comorbid disorders [23-25], possibly more frequently than often assumed [26]. The Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revision (DSM-IV-TR) provides standard criteria for the classification of mental disorders and is used in (specialized) mental health care [27,27]. Table 1.2 shows the DSM-IV-TR criteria for a selection of MAS disorders.

For MAS disorders various questionnaires are available, for nearly every diagnostic category a separate one. Although standardization of psychiatric assessments and their reference values are essential for patient care, for various MAS instruments reference values are not available.

Table 1.2. Examples of prevalent MAS disorders: DSM-IV-TR criteria of Major Depressive Episode, Panic Disorder with Agoraphobia, and Hypochondriasis

Major Depressive Episode	**Panic Disorder With Agoraphobia**	**Hypochondriasis**
A. ≥5 of the following symptoms present ≥2 weeks, representing a change from previous functioning; at least one of the symptoms is either 1 or 2	A. Both (1) and (2):	A. Preoccupation with fears of having, or the idea that one has, a serious disease based on the person's misinterpretation of bodily symptoms
1. depressed mood	1.recurrent unexpected panic attacks	
2. markedly diminished interest or pleasure	2.≥1 attack has been followed by ≥1 month of ≥1 of the following:	
3. significant weight loss or weight gain, or decrease or increase in appetite	a. persistent concern about having additional attacks	
4. insomnia or hypersomnia	b. worry about the implications of the attack or its consequences	
5. psychomotor agitation or retardation	c. a significant change in behavior related to the attacks	
6. fatigue or loss of energy		
7. feelings of worthlessness or excessive or inappropriate guilt		
8. diminished ability to think or concentrate, or indecisiveness		
9. recurrent thoughts of death or suicide		
B. The symptoms do not meet criteria for a Mixed Episode	B. The presence of agoraphobia	B. The preoccupation persists despite appropriate medical evaluation and reassurance

(continued)

Major Depressive Episode	Panic Disorder With Agoraphobia	Hypochondriasis
C. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning	C. The Panic Attacks are not due to the direct physiological effects of a substance or a general medical condition	C. The belief in Criterion A is not of delusional intensity (as in Delusional Disorder, Somatic Type) and is not restricted to a circumscribed concern about appearance (as in Body Dysmorphic Disorder)
D. The symptoms are not due to the direct physiological effects of a substance (e.g., drug of abuse, medication) or a general medical condition	D. The panic attacks are not better accounted for by another anxiety disorder	D. The preoccupation causes clinically significant distress or impairment in social, occupational, or other important areas of functioning
E. The symptoms are not better accounted for by bereavement		E. The duration of the disturbance is at least 6 months
		F. The preoccupation is not better accounted for by Generalized Anxiety Disorder, Obsessive-Compulsive Disorder, Panic Disorder, a Major Depressive Episode, Separation Anxiety, or another Somatoform Disorder

MAS denotes Mood Anxiety Somatoform; DSM-IV-TR denotes Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revision.

REFERENCE VALUES

Reference values are used for variables that can be assessed quantitatively, such as body temperature or depression severity. They are assessed in a reference population, i.e. a population not selected on pathology with respect to that variable. Reference values can be used to assess whether, for instance, a person suspected of influenza has a body temperature increased above a certain level (called 'fever') or whether somebody treated for depression still has a score increased above a certain level on a depression severity scale like the MADRS. The term 'reference values' was introduced by Gräsbeck and Saris [28]. They did so to replace the older, more ambiguous, terminology of 'normal values' by a well-defined nomenclature and recommended procedure in the field [28,29]. The term 'normal values' caused confusion because the word 'normal' has multiple, rather different connotations (e.g., statistical, epidemiological, psychological, or clinical).

The selection of the reference population and the definition of reference values are important. The reference population should consist of individuals with a well-defined state of health [29,30]. Health can be operationalized in different ways: medically and statistically. The medical approach considers health as absence of pathology, in absolute terms, or at least of a certain type of pathology. Thus, individuals with that disorder are excluded from the reference population. For instance, in the medical approach, to obtain reference values for depression, depressed patients are excluded from the reference population. The statistical approach is based on the distribution of scores of a quantifiable variable in a population, the reference population, not selected on certain values of that variable. For instance, in the statistical approach of reference values for depression severity, the latter is assessed in a population not selected on certain scores of depression, for instance a sample of the general population. In the statistical approach the middle range of scores of the distribution of that variable is considered as healthy and extreme high or low scores as deviant [31]. Healthy values usually are based on the middle 95% of the reference population. However, extreme high and low variables are not always deviant. For many variables used in ROM, like depression severity, only one extreme, mostly the highest score, is considered deviant. In such cases, deviancy is restricted to the top 5%. Individuals with current elevated levels of psychopathology (i.e., who display characteristics similar to those being addressed in the treatment) are not excluded from the reference group, because otherwise, a 'supernormal' sample would be created. Resulting reference values would be overly stringent [32]. Similarly, the bottom 5% of the psychiatrically-ill population can be considered "deviant"; their symptoms may have become subsyndromal. Deviancy at the top of the distribution is clinically meaningless (i.e., too ill). In this study the statistical approach was followed.

If it is clinically relevant, partition criteria can be used to characterize subgroups from the reference population, which could be based for example on gender and age categories, as clinically important differences in reference values may be present in these subgroups [9].

Methods of comparison

Reference values will be used to assess clinical efficacy of a treatment. To assess a change from pre-test to post-test as clinically meaningful, the proposal of Jacobson and colleagues [33] is followed in ROM. They proposed two criteria for clinically significant change: (1) the change must be greater than the measurement error of the instrument (statistically reliable change), and (2) the treated patient displays a severity of symptomatology that is equivalent to or beyond levels found in the general population. The transition from illness to health signifies recovery, the transition vice versa signifies relapse. When only the first criterion is met there is reliable improvement or deterioration, but no recovery or relapse yet. When only the second criterion is met there is indeed a transition from illness to health or vice versa, but both the pre-test score and the post-test score is so close to the cut-off value that the change is not clinically significant.

The Jacobson method is based on the assumption that the distribution of psychopathology scores in a patient population is Gaussian (normal). However, psychopathology scores like many biological data are often not symmetrically distributed in the general population [30] and the distribution is non-Gaussian. Indeed, psychopathology questionnaires measure the severity of symptoms, not the level of healthy functioning. The analytical procedures need to take these non-Gaussian distributions into account through nonparametric methods [34]. Therefore, the Jacobson method is not directly appropriate for the ROM reference group scores. Percentile scores (5^{th}, 25^{th}, 50th, 75^{th}, and 95^{th}) however can be used as a modification for both Gaussian and non-Gaussian distributions. They are introduced in this thesis for both the ROM reference group and the ROM patient group as is discussed in the section about percentile scores.

Sensitivity and specificity

Sensitivity and specificity are statistical performance characteristics of a test. Sensitivity refers to the ability of a test or a questionnaire to correctly identify those patients with psychopathology.

$$Sensitivity = \frac{True\,Positives}{True\,Positives + False\,Negatives}$$

Specificity refers to the ability of a test to correctly identify those clients without psychopathology.

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positives}$$

The terms positive predictive value and negative predictive value are used when considering the value of a test to a clinician: they answer the questions "How likely is it that the patient has the disease given that the test result is positive" and "How likely is it that the patient does not have the disease given that the test result is negative". The relationship among the terms is depicted in the following crosstab.

	Condition positive	**Condition negative**	
Test outcome positive	True Positive	Fals Positive	Positive Predictive Value
Test Outcome Negative	False Negative	True Negative	Negative Predictive value
	Sensitivity = True positive/ Condition Positive	Specificity= True Negative/ Condition Negative	

If a test results in a completely correct separation of healthy and diseased individuals, there would be no overlap between a reference group and a patient group, and sensitivity and specificity of a test would be 1. But in reality there is virtually always some overlap: i.e., there are people in the reference group who are ill and persons in the patient group who are not ill. However, in psychiatric disorders the situation is more complicated: absolute definitions of having a psychiatric disorder or not do not exist. They have to be defined on the basis of cut-off scores. In fact, it would be more correct to speak of cut-off scores indicating a severity necessitating treatment. When the cut-off scores are changed the sensitivity and specificity of the test will change. By studying several cut-off scores, optimal cut-off scores for both high sensitivity and specificity can be computed.

The sensitivity and specificity of a test are dependent on the cut-off value above which the test is considered positive and when the cutoff value is changed, the two test characteristics will change complementary: for a higher cutoff value, the specificity will increase and the sensitivity will decrease, and vice versa [35]. A cautious, high cut-off point results in a high specificity with a high percentage of true negative results in non-diseased individuals, but at the cost of a lower sensitivity, with more diseased subjects being rated as false negatives. A strict, low cut-off point will result in a high sensitivity (i.e., few false negatives at the cost of more false positives). When false negatives and false positives are equally undesirable (and the disease is not uncommon), a trade-off is commonly proposed where sensitivity and specificity are equal. Two important factors that determine the optimal balance between high sensitivity and high specificity are: a) the prevalence or a priori probability of the disorder; and b) the relative cost or undesirability of errors [36]. First, testing for low-frequency diseases is always problematic. It is relevant whether you use a test in the general population, in the primary care population, or in the psychiatric population. Given the same sensitivity and specificity, the positive and negative predictive values are very different for the different prevalence rates. Second, the 'costs' depend on the kind and prevalence of the disorder and differ for false negatives and false positives. High sensitivity is sought when the questionnaire is used to identify a serious but treatable disorder. The test will not be very specific, however, with a high proportion of clients with a positive test result who are subsequently found to have no underlying pathology (false positives). After initial screening with a sensitive test, a second test with higher specificity could identify nearly all of the false positives as disorder negative [35].

In sum, we use sensitivity and specificity because they are characteristics of the test; they are independent on the prevalence of the disease in the population of interest. This is in contrast to the use of positive and negative predictive values, which are characteristics of the usefulness of the test in different populations: they are affected by the prevalence of the disease.

Receiver Operating Characteristics (ROC)

A Receiver Operating Characteristic (ROC) is a classification model that illustrates, by way of a graphical plot, the diagnostic performance of a questionnaire as its discrimination threshold (the cut-off value) is varied. It is created by plotting the sensitivity versus the specificity, for all possible cut-off values. The Area Under the ROC Curve (AUC) is equal to the probability that the questionnaire will rank a randomly chosen positive instance higher than a randomly chosen negative one, i.e., will discriminate illness from health. ROM questionnaires, which are used to assess the level of (dys-) functionality both in the reference group and the

patient group need to have good discriminatory power. By means of ROC analyses and subsequent AUC analyses, the discriminative power that is illustrative of the diagnostic capability of the ROM questionnaires can be investigated.

Percentile scores

Reference values are used to describe and interpret the treatment outcomes, operationalized as questionnaire scores. Percentile scores (e.g., 5th, 25th, 50th [i.e., median], 75th, and 95th) are appropriate reference values for all types of distributions, including non-Gaussian distributions of reference group scores and Gaussian distributed patient scores. Indeed, this non-parametric method makes no specific assumption regarding the distribution of the scores [34]. Firstly, percentile scores facilitate norm-referenced testing, so as to determine how the tested person scores compared to other persons from a certain population, e.g., with a similar disorder or of similar gender. Secondly, percentile scores allow cut-off-referenced testing where the questionnaire score is interpreted absolutely, by comparing the score with a clinical threshold (i.e., cut-off value).

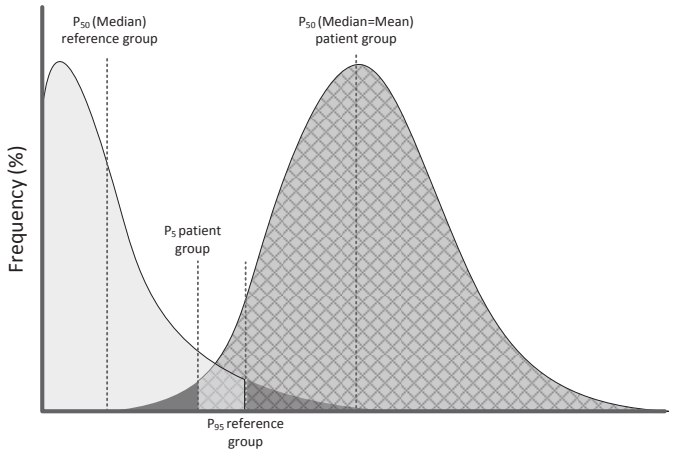


Figure 1.2. Hypothetical distribution of the scores of a questionnaire measuring psychopathology within the reference population and within the patient population. Two cut-off values are depicted: the 95th percentile score (P_{95}) of the reference group and the 5th percentile score of the patient group. The median scores (P_{50}) of the groups are depicted as well (which is equal to the mean only in case of a normal Gaussian distribution). A commonly used definition is that 1 out of 20 (or 5%) results will fall outside the established reference range in random samples from the reference population.

The 5th percentile score (P_5) of the patient population, marking the bottom 5%, would be the clinical threshold for referral from primary care to specialized mental health care (see Figure 1.2): i.e., persons enter treatment when they are no longer part of the reference population, but belong to the patient population instead (see Figure 1.3). A second clinically relevant cut-off point is the point that the patient has to cross at the time of the post-treatment assessment in order to be classified as changed to a clinically significant degree of functionality or health [34]. As can be seen in Figure 1.2, the cut-off, marking the top 5%, would be the 95th percentile score (P_{95}) of the reference population. Below this value, the patient in specialized mental health care is more similar to the reference population than to the patient population, and referral back to primary care is indicated (see Figure 1.4).

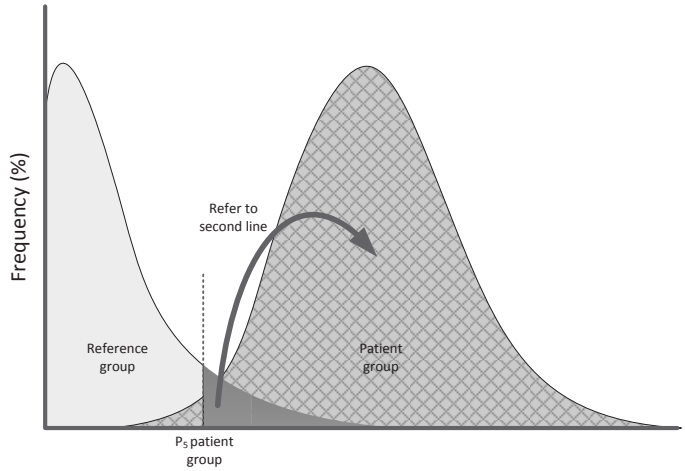


Figure 1.3. Cut-off values relevant for referral from primary care to secondary care. Patients enter treatment when they are no longer part of the reference population, but belong to the patient population instead, above the cut-off value P_5 of the patient group.

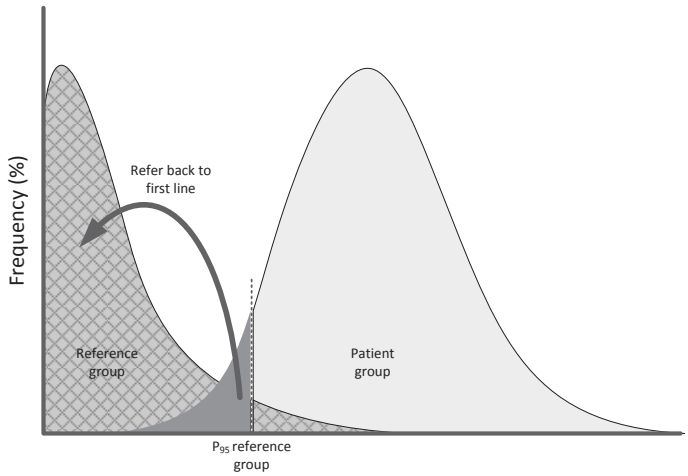


Figure 1.4. Cut-off values relevant for referral from secondary care to primary care. Patients depart from treatment when they no longer belong to the patient population, but belong to the reference population instead, below the cut-off value P_{95} of the reference group.

Considerations on the use of reference values

When interpreting differences between observed values and reference values, it is important to realize that statistical significance is only descriptive: it does not imply clinical importance per se [30]. Individual patient factors can affect the clinical meaning: overall level of functioning and the ability to carry out activities of daily living. In addition, the best-possible result of treatment is not necessarily statistically meaningful. Decision limits (i.e., cut-off values) based on reference values should not be used as a single decision criterion, but they can be an important adjunct to the clinical treatment. Clinicians are in the best position to judge the unique characteristics of their patients. A treatment strategy is most likely to succeed when it combines effective therapy, ROM and its reference values, and a strong therapeutic relationship. We do not recommend a rigid system of treatment and referral that eliminates the ability to respond to individual needs of the patient.

AIMS AND OUTLINE OF THIS THESIS

Aims of this thesis

As specified above, ROM is a measurement feedback system that facilitates systematic evaluation of a patient's treatment response during the course of treatment in routine clinical practice. ROM comprises a comprehensive assessment battery, including both generic and disorder-specific measures. The first aim of the study in this thesis (referred to as the NormQuest [i.e., quest for norms] study) was to provide empirical based, valid reference values for patients with one or more MAS-disorders. We aimed to generate reference values for both 'healthy' and 'clinically ill' MAS populations. We chose to define health statistically (as opposed to medically). To enable norm-referenced testing, percentile scores were calculated for each of the measures. To facilitate cut-off-referenced testing, we aimed to calculate cut-off values based on percentile scores and Receiver Operating Characteristics (ROC). The P_5 ROM patient group cut-off values can be used by primary care physicians as decision indicator for referral to the specialized mental health care. The P_{95} ROM reference group cut-off values can be used by specialized mental health care as decision indicator for referral back to primary care physicians. For comparability with the international literature, we also report means and standard deviations. We calculated reference values in separate strata of gender and age to study the strata effects. Also, we assessed the discriminative power of the questionnaire scores by means of Receiver Operating Characteristics (ROC) analyses. Additionally, internal consistency reliabilities were calculated.

The second aim of the NormQuest study concerned the development of public domain questionnaires. In this study, the Symptom Questionnaire-48 (SQ-48) was developed as a public domain alternative for the frequently used Brief Symptom Inventory (BSI), which is not free of charge.

Thesis outline

Chapter 2 describes the objectives, design, and methods of the NormQuest study in detail. The extensive process of recruitment and baseline characteristics of the reference group versus the patient group are reported.

In *Chapter 3*, reference values for four generic questionnaires were calculated: the Brief Symptom Inventory (BSI), the Mood & Anxiety Symptom Questionnaire -30 (MASQ-D30), the Short Form Health Survey 36 (SF-36), and the Dimensional Assessment of Personality Pathology-Short Form (DAPP-SF). Gender- and age effects were studied.

In *Chapter 4*, we focused on the reference values for three disorder-specific questionnaires concerning depression: the Beck Depression Inventory-II (BDI-II), the Inventory of Depressive Symptoms (self-report) (IDS-SR), and the Montgomery-Äsberg Depression Rating Scale (MADRS). Again gender- and age effects were assessed.

In *Chapter 5*, we calculated reference values for eight anxiety questionnaires: the Brief Scale for Anxiety (BSA), the PADUA Inventory Revised (PI-R), the Panic Appraisal

Inventory (PAI), the Penn State Worry Questionnaire (PSWQ), the Worry Domains Questionnaire (WDQ), the Social Interaction, the Anxiety Scale (SIAS), the Social Phobia Scale (SPS), and the Impact of Event Scale-Revised (IES-R). These questionnaires cover most of the anxiety disorders.

Chapter 6 provides reference values for three disorder-specific questionnaires concerning somatoform disorders: the Body Image Concern Inventory (BICI), the Checklist Individual Strength (CIS20R), and the Whitely Index (WI). These questionnaires assess symptom severity in patients with body dysmorphic disorder, hypochondriasis and chronic fatigue syndrome.

Chapter 7 describes the development, validation and reference values of our newly developed public domain questionnaire, the 48-item Symptom Questionnaire (SQ-48). This questionnaire was developed as a psychological distress instrument, including measures of vitality and work functioning, to be used as a screening / monitoring tool in clinical settings (psychiatric and non-psychiatric), as a benchmark tool, or for research purposes.

Finally, in *Chapter 8*, we summarized the main results of this study. We discussed these results, the clinical implications, and provided recommendations for further improvement of ROM as well as suggestions for future research.

Reference List

1. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC. (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J *Clin Psychiatry*, 59 Suppl 20, 22-33.

2. Carlier IVE, Meuldijk D, Van Vliet IM, Van Fenema EM, Van der Wee NJ, Zitman FG. (2012) Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *J Eval Clin Pract,* 18 (1), 104-110.

3. Knaup C, Koesters M, Schoefer D, Becker T, Puschner B. (2009) Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. *Br J Psychiatry,* 195 (1), 15-22.

4. De Beurs E, Den Hollander-Gijsman ME, Van Rood YR, Van der Wee NJ, Giltay EJ, Van Noorden MS, Van der Lem R, Van Fenema EM, Zitman FG. (2011) Routine outcome monitoring in the Netherlands: practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clin Psychol Psychother,* 18, 1-12.

5. Carlier I, Giltay E, Vergeer P. (2012) Development and validation of the 48-item Symptom Questionnaire (SQ-48) in patients with depressive, anxiety and somatoform disorders. *Psychiatry Res,* 200 (2-3), 904-910.

6. Moessner M, Gallas C, Haug S, Kordy H. (2011) The clinical psychological diagnostic system (KPD-38): sensitivity to change and validity of a self-report instrument for outcome monitoring and quality assurance. *Clin Psychol Psychother,* 18 (4), 331-338.

Loge JH, Kaasa S. (1998) Short form 36 (SF-36) health survey: normative data from the general Norwegian population. *Scand J Soc Med,* 26 (4), 250-258.

8. McKay R, Coombs T, Pirkis J. (2012) A framework for exploring the potential of routine outcome measurement to improve mental health care. *Australas Psychiatry*, 20 (2), 127-133.

9. Van Noorden MS, Giltay EJ, Den Hollander-Gijsman ME, Van der Wee NJ, Van Veen T, Zitman FG. (2010) Gender differences in clinical characteristics in a naturalistic sample of depressive outpatients: the Leiden Routine Outcome Monitoring Study. *J Affect Disord,* 125 (1-3), 116-123.

10. De Beurs E, Rinne T, Van Kampen D, Verheul R, Andrea H. (2009) Reliability and validity of the Dutch Dimensional Assessment of Personality Pathology-Short Form (DAPP-SF), a shortened version of the DAPP-Basic Questionnaire. *J Pers Disord,* 23 (3), 308-326.

11. Schulte-van Maaren YWM, Carlier IVE, Zitman FG, Hemert AM, De Waal MW, Van Noorden MS, Giltay EJ. (2012) Reference values for generic instruments used in Routine Outcome Monitoring: the Leiden Routine Outcome Monitoring Study (in press). *BMC Psychiatry.*

12. Schulte-van Maaren YWM, Carlier IVE, Giltay EJ, Van Noorden MS, De Waal MW, Van der Wee NJ, Zitman FG. (2012) Reference values for mental health assessment instruments: objectives and methods of the Leiden Routine Outcome Monitoring Study. *J Eval Clin Pract*.

13. Schulte-van Maaren YWM, Carlier IVE, Zitman FG, Van Hemert AM, De Waal MW, Van der Does AJW, Van Noorden MS, Giltay EJ. (2012) Reference values for major depression questionnaires: the Leiden Routine Outcome Monitoring Study. *Journal of Affective Disorders*.

14. Wardenaar KJ, Van Veen T, Giltay EJ, Den Hollander-Gijsman ME, Penninx BW, Zitman FG. (2010) The structure and dimensionality of the Inventory of Depressive Symptomatology Self Report (IDS-SR) in patients with depressive disorders and healthy controls. *J Affect Disord,* 125 (1-3), 146-154.

15. Bijl RV, Ravelli A, Van Zessen G. (1998) Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc Psychiatry Psychiatr Epidemiol,* 33 (12), 587-595.

16. De Graaf R, Ten Have M, Van Gool C, Van Dorsselaer S. (2012) Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. *Soc Psychiatry Psychiatr Epidemiol,* 47 (2), 203-213.

17. De Waal MW, Arnold IA, Eekhof JA, Van Hemert AM. (2004) Somatoform disorders in general practice: prevalence, functional impairment and comorbidity with anxiety and depressive disorders. *Br J Psychiatry,* 184, 470-476.

18. Escobar JI. (2009) Somatoform Disorders. Sadock BJ, Sadock VA, Ruiz P, editors. Kaplan & Sadock's Comprehensive Textbook of Psychiatry. 9th. Lipincott Williams & Wilkins.

19. Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler KS. (1994) Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Arch Gen Psychiatry,* 51 (1), 8-19.

20. Roca M, Gili M, Garcia-Garcia M, Salva J, Vives M, Garcia CJ, Comas A. (2009) Prevalence and comorbidity of common mental disorders in primary care. *J Affect Disord*, 119 (1-3), 52-58.

21. Toft T, Fink P, Oernboel E, Christensen K, Frostholm L, Olesen F. (2005) Mental disorders in primary care: prevalence and co-morbidity among disorders. results from the functional illness in primary care (FIP) study. *Psychol Med,* 35 (8), 1175-1184.

22. Wittchen HU, Jacobi F, Rehm J, et al. (2011) The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol,* 21 (9), 655-679.

23. Ansseau M, Dierick M, Buntinkx F, Cnockaert P, De Smedt J, Van den Haute M, Van der Mijnsbrugge D. (2004) High prevalence of mental disorders in primary care. *J Affect Disord,* 78 (1), 49-55.

24. Gili M, Comas A, Garcia-Garcia M, Monzon S, Antoni SB, Roca M. (2010) Comorbidity between common mental disorders and chronic somatic diseases in primary care patients. *Gen Hosp Psychiatry,* 32 (3), 240-245.

25. Hanel G, Henningsen P, Herzog W, Sauer N, Schaefert R, Szecsenyi J, Lowe B. (2009) Depression, anxiety, and somatoform disorders: vague or distinct categories in primary care? Results from a large cross-sectional study. *J Psychosom Res,* 67 (3), 189-197.

26. Mergl R, Seidscheck I, Allgaier AK, Moller HJ, Hegerl U, Henkel V. (2007) Depressive, anxiety, and somatoform disorders in primary care: prevalence and recognition. *Depress Anxiety,* 24 (3), 185-195.

27. American Psychiatric Association. (2000) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision. Washington,DC, American Psychiatric Association.

28. Gräsbeck R, Saris N-E. (1969) Establishment and use of normal values. *Scan J Clin Lab Invest,* 26 (110), 62-63.

29. Geffre A, Friedrichs K, Harr K, Concordet D, Trumel C, Braun JP. (2009) Reference values: a review. *Vet Clin Pathol,* 38 (3), 288-298.

30. Solberg HE. (1987) International Federation of Clinical Chemistry. Scientific committee, Clinical Section. Expert Panel on Theory of Reference Values and International Committee for Standardization in Haematology Standing Committee on Reference Values. Approved recommendation (1986) on the theory of reference values. Part 1. The concept of reference values. *Clin Chim Acta,* 165 (1), 111-118.

31. Zimmerman M, Chelminski I, Posternak M. (2004) A review of studies of the Montgomery-Asberg Depression Rating Scale in controls: implications for the definition of remission in treatment studies of depression. *Int Clin Psychopharmacol,* 19 (1), 1-7.

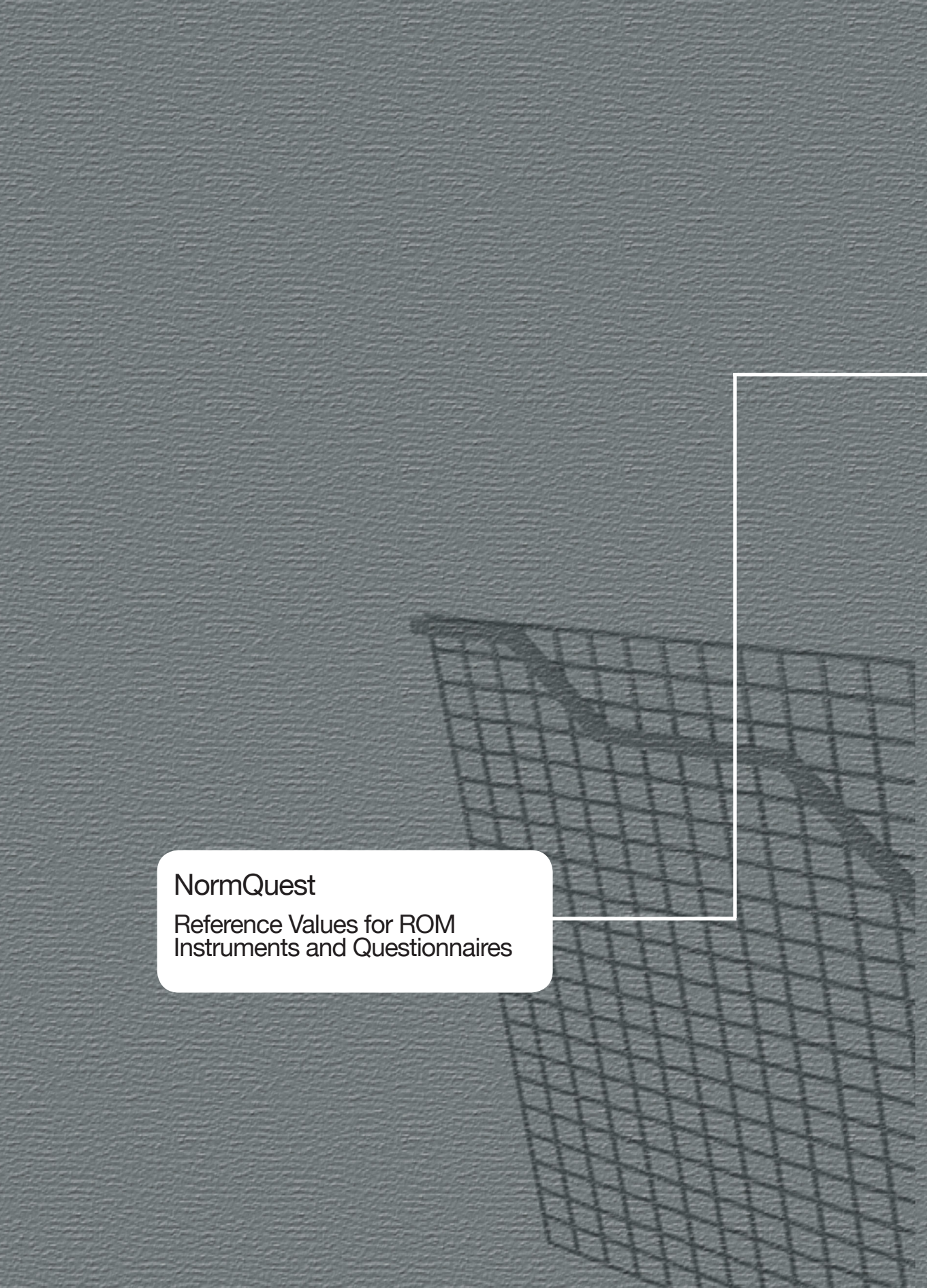
32. Kendall PC, Marrs-Garcia A, Nath SR, Sheldrick RC. (1999) Normative comparisons for the evaluation of clinical significance. J *Consult Clin Psychol,* 67 (3), 285-299.

33. Jacobson NS, Roberts LJ, Berns SB, McGlinchey JB. (1999) Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *J Consult Clin Psychol,* 67 (3), 300-307.

34. Solberg HE. (2008) Establishment and use of reference values. Burtis CA, Ashwood ER, Bruns DE, editors. Fundamentals of clinical chemistry. 6[14], 229-238. St. Louis, Missouri, Saunders Elsevier.

35. Bewick V, Cheek L, Ball J. (2004) Statistics review 13: receiver operating characteristic curves. *Crit Care,* 8 (6), 508-512.

36. Marazia S, Barnabei L, De Caterina R. (2008) Receiver operating characteristic (ROC) curves and the definition of threshold levels to diagnose coronary artery disease on electrocardiographic stress testing. Part II: the use of ROC curves in the choice of electrocardiographic stress test markers of ischaemia. *J Cardiovasc Med (Hagerstown),* 9 (1), 22-31.


NormQuest
Reference Values for ROM
Instruments and Questionnaires

Chapter 2

Reference values for mental health assesment instruments: objectives and methods of the Leiden Routine Outcome Monitoring Study

Yvonne W.M. Schulte-van Maaren
Ingrid V.E. Carlier
Erik J. Giltay
Martijn S. van Noorden
Margot W.M de Waal
Nic. J.A. van der Wee
Frans G. Zitman

Journal of Evaluation in Clinical Practice (2013), 19(2), 342-50

ABSTRACT

Background: Routine outcome monitoring (ROM) was developed to establish the outcome of psychotherapeutic and pharmacological treatments through repeated assessments before, during and after treatment. Although standardization of psychiatric assessments and their reference values are essential for patient care, for various ROM instruments reference values are not available. The aim of the Leiden ROM Study is to generate reference values for 22 ROM instruments, covering generic and specific mood, anxiety and somatoform (MAS) disorders, for the general population. This article describes the extensive process of recruitment, as well as baseline characteristics of patient versus non-patient groups.

Method: Cross-sectional study in randomly selected participants aged 18-65 years from the Dutch population, included through general practitioners.

Results: Extensive demographic, psychosocial, mental health, and biological data from 1302 participants, recruited via general practitioners, were collected during a two-hour standardized assessment including observer-rated and self-report scales. These data will be compared with corresponding data from 7840 patients with psychopathology who were referred to secondary care. On-going quality control and calibration ensured maintenance of high quality during data collection.

Discussion: This reference group study for mental health assessments is the first study of this size carried out in the Netherlands.The results of this study are expected to be of value to secondary psychiatric care because they allow the indication of progress in health, treatment effect and possible termination of treatment. Additionally, the reference values can be used by primary care physicians as decision threshold for referral to specialized mental health care and vice versa.

INTRODUCTION

Routine outcome monitoring (ROM) was developed to enhance the effectiveness of psychiatric care. ROM routinely measures treatment outcomes using different outcome measures that are both generic and disorder-specific. It provides clinicians with information on the type and severity of psychopathology and feedback on treatment efficacy. Additional benefits are its use in research and benchmarking [1-3]. However, several ROM instruments lack reference values that provide optimal discrimination between the 'healthy' and the 'diseased', indicating whether the patient has progressed to a range of psychological health similar to non-patients, whilst not necessarily free of all symptoms. Also, with outcome variables often varying between different gender and age groups, reference values are the key to determining whether a group or an individual scores above or below average for their gender and age [4,5]. Anchoring ROM instruments in population-based reference values makes clinical and scientific interpretations more meaningful and is consistent with practice in other areas of medicine [6,7]. Furthermore, reference values are useful to determine when primary care physicians could refer their patients to secondary care and vice versa.

In order to study the relationship between psychosocial factors, genetic variation, the effect of the hypothalamic-pituitary-adrenal (HPA) axis stress system, and the occurrence and course of mood, anxiety and somatoform (MAS) disorders, the Leiden Routine Outcome Monitoring Study was designed to generate a large ROM database [8,9].

The present ROM Reference Group Study was designed to provide reference values for 22 ROM in the general practice population in the Netherlands. This may help to facilitate assessment of a clinically significant change of treatment effects, defined as returning to normal functioning.

A secondary aim was to collect saliva from a large general population control group in order to facilitate research on genetic characteristics (DNA) and the HPA axis stress system in relation to the development and course of MAS disorders. Genetic factors and a deregulated HPA axis are involved in the etiology of MAS disorders. Twin studies [10,11] have shown that mood and anxiety disorders are for 30-40% determined by hereditary factors. Furthermore, dysregulation of the HPA axis is believed to be linked with the pathophysiology of depression [12-14] and anxiety disorders [15,16].

The present study describes the methods and objectives of the ROM Reference Group Study, as well as baseline characteristics of patient versus non-patient groups.

METHODS

Participants

The ROM reference group was recruited to serve as a comparison for the ROM patient group. Therefore, the aim for this reference group was that it be representative of the ROM population referred for suspected (but not necessarily diagnosed with) MAS disorders, treated at the psychiatric outpatient department of the Leiden University Medical Center (LUMC) or at the mental health clinics of Rivierduinen (RD) (hereafter referred to as the 'ROM patient group'). The sample was stratified for gender, age and urbanization level to be representative of the ROM patient group [17].

A total of 1302 participants (18-65 years) was recruited, 1294 of whom provided complete data sets (Figure 2.1). In order to recruit persons reflecting normal functioning with different levels of subthreshold psychopathology, recruitment took place via general practices. In the Netherlands, because 99.9% of the general population is registered with a general practitioner (GP) [18], the practice registers provide a convenient frame for sampling the local general population. Eight university-affiliated general practices with a total of ± 14,000 enlisted patients in the vicinity of Leiden were involved. In order to form a non-patient control group and to secure the reliability and validity of the collected data, four exclusion criteria were formulated: 1) treatment in a secondary psychiatric care centre in the last six months for psychiatric problems and/or dependence on alcohol or drugs; 2) hearing impairment, limited cognitive abilities, such as aphasia, severe dyslexia or dementia; 3) illiteracy or insufficient mastery of the Dutch language, and; 4) a terminal disease.

The study protocol was approved by the Ethical Review Board (ERB) of the LUMC and all subjects signed informed consent.

Since 2002 the LUMC and RD, serving a region of more than one million people, have implemented ROM [1]. ROM baseline assessments in the ROM patient group started in 2002 and are ongoing. Specially trained psychiatric research nurses assessed 80% of the patients (totaling 8357 ROM patients), 7840 of whom were aged 18-65 years. To facilitate research on genetic characteristics (DNA) and the HPA axis stress system (cortisol day curves) the MASHBANK (biobank for MAS disorders and the HPA-axis) was founded at the LUMC and RD in 2007 after approval by the ERB of the LUMC. In this biobank, saliva samples are stored from ± 1000 consenting MAS patients. Figure 2.1 shows the multi-stage recruitment flow of the ROM reference group, as well as recruitment of the ROM patient group.

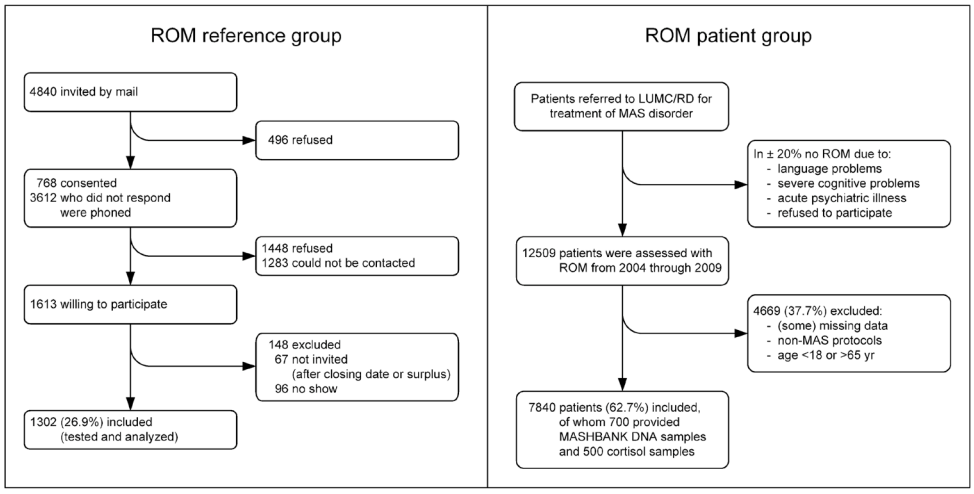


Figure 2.1. Flow chart depicting recruitment of the ROM reference and patient groups

Participants of the ROM reference group were offered the full set of generic instruments. Since the total number of instruments was too extensive and all participants were already asked to complete the depression instruments, random samples of 50% each were asked to complete the anxiety instruments or the somatoform instruments, with even ratios of males and females in each subgroup. Thus, four subgroups were established: males-anxiety; females-anxiety; males-somatoform; females-somatoform. A sample size of at least 120 per subgroup was considered to provide adequate power to yield reference values [19]. In genetic research an adequate sample size is imperative because of the low frequency of several genetic variants and the problem of multiple testing. Furthermore, a sample size of 1000 DNA donating participants was deemed to be required [20]. With an anticipated response rate of 30%, about 4500 people were approached. In order to get a ROM representative sample, four age groups were used: 18-25; 26-40; 41-55; 56-65 years, and the reference group was sampled accordingly.

Procedures

In order to recruit the ROM reference group, the eight participating GPs first screened their patient lists for those that met the inclusion/exclusion criteria. Subsequently, randomly selected appropriate persons were invited to participate by a letter (sent by regular postal service) by their GPs that was followed by an announced telephone call by the research team to ask for their participation. Objections against this call could be indicated on an enclosed reply card. To compensate for possible seasonal influences, recruitment took place all year long (between November 2009 and January 2011). Location was the LUMC clinic site and, if appreciated, at the participant's home or in the GP's practice. Similar to the ROM patient assessment procedures, dedicated web-based computer software was used for the

administration of all instruments and to prevent missing data within instruments. It was also used for data collection and storage, and for creation of summary variables [1]. Touch screens were used to accommodate computer-illiterate participants. A personal data entry program was developed in database software to organize identification codes for general, ROM and MASHBANK data, and to randomly assign the two specific instrument packets (depression and anxiety; depression and somatoform) to participants.

For participants of the ROM reference group the interview started with an explanation of the study, and signing of the informed consent form. This was followed by a check and assessment of personal details and demographic data, general health, cognitive functioning, and physical examination (i.e., body weight, height, and blood pressure). Saliva samples were collected in participants who additionally consented to this biobank substudy. Next, computerised observer-rated and self-report questionnaires were completed. Finally, participants completed an evaluation form and received a gift voucher of €30 (for their time and cooperation) and a travel allowance.

In the ROM Reference Group Study 3 psychiatric research nurses, 3 psychologists (Master's degree level) and 11 Master's students in psychology were extensively trained and tested at the start of and during the reference group study to ensure uniform and adequate quality and reliability. Topics were Mini-International Neuropsychiatric Interview Plus, version 5.0.0-R (MINI Plus 5.0.0.) and abbreviated Comprehensive Psychopathological Rating Scale (vCPRS) interviewing methods, Global Assessment of Functioning Scale (GAF) scoring, use of QuestManager, and additional knowledge about MAS disorders and MASHBANK. Three full days of training (by the primary investigator, SvM, two psychiatrists and two ROM-trained nurses) took place. Each interviewer also observed at least three interviews, and the first two interviews were carried out under supervision (one of which observed by the primary investigator). Supervision regarding interview techniques, problematic behaviour of the participants and scoring rating scales, to improve inter-rater reliability, took place every two months. Videorecordings of interviews were used to further calibrate assessments between interviewers. Using a semi-structured scoring scale a qualitative assessment was done, and was found to be very good in all but one potential interviewer. This latter interviewer with insufficient skills was considered unsuitable and no longer took part. The ROM patient group was assessed by two trained ROM psychiatric research nurses; their training has been decribed in detail elsewhere [1].

Assessments

The ROM reference group assessment comprised measurement of physical health, saliva collection and observer-rated and self-report instruments. Measurement of physical health indicators comprised blood pressure, heart rate and body mass index, and health-related factors (i.e., general health, chronic diseases, smoking status, and alcohol consumption).

From participants who agreed to participate in the MASHBANK substudy, saliva was collected enabling cortisol measurements and DNA isolation. HPA axis activity was

assessed by free cortisol measurements using seven saliva samples per participant, self collected at home. Procedures are similar to that described in detail elsewhere [14,16,21]. Saliva for DNA isolation was collected in DNA Genotek kits (Oragene). Measuring cortisol and DNA concentrations in saliva has many advantages over measurements in blood samples. Saliva collection is non-invasive and can be repeated frequently. Furthermore, storage of the material requires no special treatment because DNA and cortisol levels remain stable at room temperature.

The assessments comprised 25 instruments concerning demographic and personal characteristics, psychosocial function, physical health and psychopathology (Table 2.1), 22 of which require reference values. Except for the 48-item Symptom Questionnaire (SQ-48), all tested ROM instruments are internationally used and validated. The generic self-report instrument SQ-48 was recently developed by our research group in order to assess mood, anxiety, somatoform symptoms, hostility and vitality.

Table 2.1. Instruments used in the ROM reference and patient groups

Instrument	Full name	Domain	No. of items	Time (min)	Type	Public domain	References
Generic							
Personal							
DEMOG	Demographic Inventory	Demography	12	2	SR	Yes	
CTQ	Child Trauma Questionnaire	Traumatic Events Childhood	28	5	SR	Yes	[22]
Psychosocial functioning							
GAF	Global Assessment of Functioning	General Functioning	1	1	Obs	Yes	[23]
LOT-R	Life Orientation Test – Revised	Optimism	10	5	SR	Yes	[24]
SF-36	Short Form Health Survey 36	Physical Health	36	6	SR	Yes	[25]
Psychopathology							
BSI	Brief Symptom Inventory	General Pathology	53	8	SR	No	[26]
DAPP-sf	Dimensional Assessment of Personality Pathology - Short Form	Personality	136	33	SR	No	[27]
IES-R	Impact of Event Scale – Revised	Traumatic Events	22	5	SR	Yes	[28,29]
MASQ-D30	Mood & Anxiety Symptom Questionnaire -30	Mood and Anxiety	30	5	SR	Yes	[30]
MINI Plus 5.0.0. *	Mini International Neuropsychiatric Interview Plus 5.0.0.	General Pathology	-	30	Obs	Yes	[31]
SQ-48	Symptom Questionnaire -48 Items	General Pathology	55	4	SR	Yes	[32]
vCPRS *	Abbreviated Comprehensive Psychopathological Rating Scale	General Pathology	25	10	Obs	Yes	[33]
WSQ	Web Screening Questionnaire for common mental disorders	General Pathology	15	5	SR	Yes	{Donker, 2009 223 /id}

Table 2.1. continued.

Depressive disorder							
IDS-SR	Inventory of Depressive Symptoms	Depressive Disorder	34	3	SR	Yes	[35]
BDI-II	Beck Depression Inventory version II	Depression, Dysthymia & bipolar Disorder	21	5	SR	No	[36]
Anxiety disorder							
AGO	Agoraphobia Scale	Panic Disorder	20	5	SR	Yes	[37]
PADUA/PI-r	PADUA Inventory revised	Obsessive Compulsive Disorder	41	6	SR	Yes	[38]
PAI	Panic Appraisal Inventory	Panic Disorder	45	10	SR	Yes	[39]
PSWQ	Penn State Worry Questionnaire	Generalized Anxiety Disorder	16	3	SR	Yes	[40]
SPS	Social Phobia Scale	Social Phobia	20	5	SR	Yes	[41,42]
SIAS	Social Interaction and Anxiety Scale	Social Phobia	20	5	SR	Yes	[41,42]
WDQ	Worry Domains Questionnaire	Generalized Anxiety Disorder	30	3	SR	Yes	[43,43]
Somatoform disorder							[27]
BICI	Body Image Concern Inventory	Body dysmorphic disorder	19	4	SR	Yes	[44]
CIS20r	Checklist Individual Strength	Chronic Fatigue Syndrome	20	5	SR	Yes	[45]
WI	Whitely Index	Hypochondriasis	14	3	SR	Yes	[46]

* The MINI Plus 5.0.0 and vCPRS are used for diagnoses; no reference values were established
A list of all ROM instruments, including references of Dutch translations, is available at http://www.lumc.nl/psychiatry/ROM-instruments.
SR; self-report, Obs; observer-rated

Statistical analyses

Reference values will be calculated for all instruments, including subscales. Both for patients and for the reference group reference values will be determined for all subjects combined, as well as for 4 groups: young males (aged 18-40 yr), older males (aged 41-65 yr), young females (aged 18-40 yr), and older females (aged 41-65 yr). Means and SDs, 5th, 25th, 50th, 75th and 95th percentiles, and receiver operating characteristics (ROC) analyses (i.e., the cut off score with the optimal sensitivity and specificity, and area under curve values) will be computed. Reference limits are often defined by two standard deviations (SDs) below and above the mean if distributions are Gaussian. Since most distributions of total scores on the scales tested in the healthy reference group are expected to be strongly (positively) skewed, percentiles are more appropriate [47-49], with the lower interval bounded only by the 95th percentile being a common reference group [50]. However, trade-offs exist between the sensitivity and specificity, with a higher cut-off value (i.e., higher percentile boundary) having a relatively high specificity but low sensitivity, and vice versa (Figure 2.2; left panel). ROC analyses will provide additional cut offs reflecting discriminatory power [51]. Figure 2.2 (right panel) shows psychopathology expressed as the number of MINI diagnoses of MAS disorders in the ROM reference group and the ROM patient group.

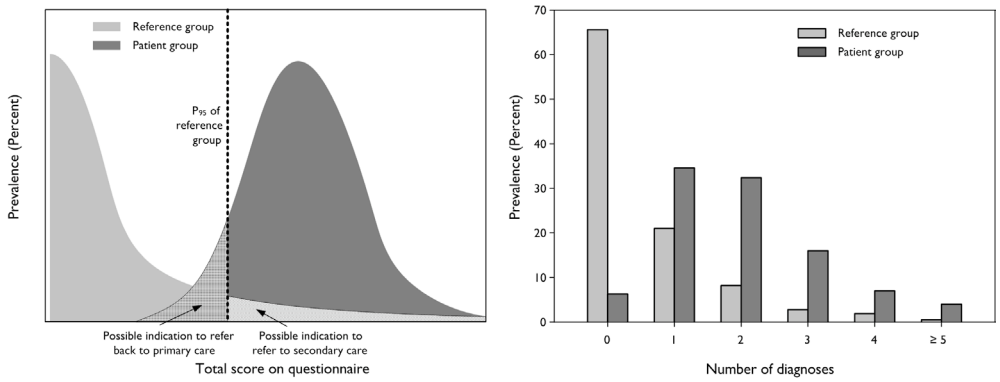


Figure 2.2. Left panel: the expected distribution of many of the 22 mood, anxiety and somatoform (MAS) disorder-assessment instruments in the ROM reference and patient groups; Right panel: the actual distribution of the number of MAS disorders in the ROM reference and patient groups. In the ROM reference group, above the 95th percentile (P_{95}; i.e., reference value) the probability is high for a person to meet the terms of psychopathology.

RESULTS

Figure 2.1 shows recruitment of the ROM reference group and the ROM patient group. A total of 1302 persons were interviewed and their data analysed. The duration of the interview was shorter (range 1.5-2.0 h) in participants without psychopathology and longer (range 2.5-4.0 h) in participants with psychopathology. Although the interview was extensive, all participants finished the full assessment. Additional telephone calls after the initial mailing proved to have a motivating effect on the subsequent response rates. Patients from the first GP only received the invitation by mail (no telephone call) and showed a response of 16.3%. We tried to contact patients recruited from all other GPs by telephone. The response to the initial mail, before the telephone call by the research team, was 15.9% (768 of 4840). The response to the telephone call and the mail was 45.3% of those the research team managed to contact (1613 of 3557). A total of 67 responders were not included because of a surplus in some of the age groups, or due to logistical reasons at the end of the study. Therefore, the response of persons contacted was 37.3% (1302 of 3490). However, when taking into account the large group of 1283 persons that could not be contacted by letter or telephone, the response of persons mailed was 26.9% (1302 of 4840). A total of 148 persons were excluded: 36 who consented following the mail (treated in a secondary psychiatric care centre, or insufficient mastery of the Dutch language), 101 after a telephone call (for similar reasons), and 11 during or after the interview (for similar reasons, as well as severe dyslexia or cognitive impairment).

Table 2.2 presents the sociodemographic characteristics of the ROM reference group (n=1294) and the ROM patient group (n=7840), 543 of whom did not complete the Demographic Inventory. Gender and age distributions in both samples were similar, and the mean age in both samples was 2 years higher for men than for women. Compared to the ROM patient group, the ROM reference group less often lived in a rural area, was less often divorced, separated or widowed, was less often unemployed or disabled, and had a higher educational level.

Table 2.2. Sociodemographic characteristics of the ROM reference group (n=1294) and the ROM patient group (n=7297).

	ROM reference group	ROM patient group
Gender		
Male	484 (37.4%)	2700 (37.0%)
Female	810 (62.6%)	4597 (63.0%)
Age (mean, SD) in years	40.2 (12.5)	37.9 (12.3)
18-25	194 (15.0%)	1508 (20.7%)
26-40	479 (37.0%)	2715 (37.2%)
41-55	448 (34.6%)	2370 (32.5%)
56-65	173 (13.4%)	704 (9.6%)
Urbanization level		
Urban	806 (62.3%)	3955 (54.2%)
Rural	488 (37.7%)	3342 (45.8%)
Marital status		
Married/cohabitating	890 (68.8%)	3721 (50.9%)
Divorced/separated/widow	78 (6.0%)	989 (13.6%)
Single	326 (25.2%)	2587 (35.5%)
Housing situation		
Living alone	200 (15.7)	1693 (23.2%)
Living with partner	902 (69.7)	3762 (51.6%)
Living with family	192 (14.8)	1842 (25.2%)
Educational status		
Lower	295 (22.8)	3133 (42.9%)
Higher	999 (77.2)	4164 (57.1%)
Employment status		
Employed part-time	508 (39.3%)	1737 (23.9%)
Employed full-time	554 (42.8%)	1702 (23.3%)
Unemployed/retired	197 (15.2%)	2118 (27.1%)
Work-related disability	35 (2.7%)	1874 (25.7%)
Ethnic background		
Dutch	1160 (89.6%)	5981 (80.0%)
Other ethnicity	134 (10.4%)	1316 (18.0%)

The aim for the ROM reference group was that it be a 'normal' group but allowed for prevalent psychopathology that could be treated in the GP practices and, therefore, showed some (co-)morbidity of psychiatric illness but to a much lesser extent than the ROM patient group (Figure 2.2). According to the MINI-Plus, 9.4% of the ROM reference group met criteria for one or more MAS disorders compared to 74.5% in the ROM patient group. A single MAS diagnosis was present in 7.8% participants and in 47.9% ROM patients. In the ROM reference group, anxiety disorders were most prevalent followed by somatoform disorders. In the ROM patient group, major depression was the most prevalent disorder followed by anxiety disorders. Thus, the ROM reference group showed lower comorbidity than the ROM patient group, and reflected psychiatric morbidity within the general population (Table 2.3, Figure 2.2).

Table 2.3. Mood, anxiety and somatoform (co-)morbidity in the ROM reference group (n=1302) and the ROM patient group (n=7840).

	ROM reference group		ROM patient group	
	Frequency	Percent	Frequency	Percent
MINI diagnoses (%)				
None	1193	**90.6**	1998	**25.5**
Anxiety	54	**4.1**	1568	**20.0**
Mood	7	**0.5**	1682	**21.5**
Somatoform	42	**3.2**	500	**6.4**
Anxiety & Mood	7	**0.5**	1377	**17.6**
Anxiety & Somatoform	9	**0.7**	209	**2.7**
Mood & Somatoform	1	**0.1**	275	**3.5**
Anxiety & Mood & Somatoform	2	**0.2**	231	**2.9**
Total Anxiety	72	**5.5**	3385	**43.2**
Total Mood	17	**1.3**	3565	**45.5**
Total Somatoform	54	**4.2**	1215	**15.5**
Total	1302	**100.0**	7840	**100.0**

Anxiety disorders comprise panic disorder with or without agoraphobia, agoraphobia without history of panic disorder, specific phobia, social phobia, obsessive compulsive disorder, post-traumatic stress disorder, and generalized anxiety disorder, and anxiety disorders NOS. Mood disorders comprise major depressive disorders, bipolar disorder, dysthymia, Somatoform disorders comprise somatization disorder, undifferentiated somatoform disorder, pain disorder (chronic), hypochondriasis, bodydysmorphic disorder, and conversion disorder.

DISCUSSION

This cross-sectional study in a randomly selected sample from a Dutch general population (aged 18-65 years) aimed to provide reference values for ROM instruments (and to serve as a control group for the biobank) for patients with MAS disorders. It is the first reference group study for mental health assessments of this size carried out in the Netherlands. The large sample size and extensive assessment of psychopathology provide data which, by comparison with data from ROM patients, is expected to yield reliable reference values for ROM instruments (across a wide age range) that are not yet available. Genetic and HPA axis data enable further biological research into MAS disorders.

Comparison of the demographics of the ROM reference and patient groups showed a similar gender and age distribution, as expected given the sampling frame. There was a slightly (unintentional) different urbanization level. However, the effects of urbanicity on psychopathology are generally of limited significance in international [52] and Dutch (NEMESIS) [4] comorbidity studies. Moreover, differences between rural and urban areas are declining in the Netherlands. Compared to the ROM patient group, the ROM reference group showed higher levels of education and less unemployment or work-related disability. Accordingly, both comorbidity studies [4,52]reported the highest morbidity rates for those with the lowest levels of education, and the lowest morbidity rates for those with the highest levels of education. Mental disorders were reported to be least prevalent amongst people in paid employment. Overall morbidity and comorbidity were strongly associated with occupational disability and unemployment.

As expected, morbidity of any current MAS disorder in the reference group was much lower than in the ROM patient group. Anxiety disorders were equally prevalent in the ROM reference group compared to a study in the general practice population (n=1778) in the Netherlands (5.5%) [53]. Mood disorders were less prevalent in the reference group (1.3%) than in the general practice population (4.1%) as well as compared to prevalence rates in various European countries, ranging between 4.6% and 7.4% [54]. The current prevalence rate for somatoform disorders was 4.2% in our ROM reference group, compared to 16.1% in a general practice population [53]. This discrepancy can probably be ascribed to differences in the recruitement procedure, as the latter study included consultation seeking patients whereas we included a random sample of the general practice population. Also, in our study most interviews took place in hospital versus home interviews in the study of De Waal et al. Another explanation could be differences in the ascertainment of depressive and somatoform disorders (MINI Plus 5.0.0. in our study versus the Scan diagnostic interview in the study of De Waal et al.). Moreover, selection and non-response bias may have occurred in our study, as depressed people are often less inclined to participate because of fatigue or loss of energy. Comorbidity rates of psychopathology in the reference group were similar to those reported in the Dutch comorbidity study [4] and very low compared to the ROM patient group.

Several issues need to be considered when analyzing reference values for psychiatric assessment scales from healthy populations. Reference values need to be accurate and reproducible. First, in samples derived from the general population many of the total scores do not have a bell-shaped Gaussian distribution, but rather an asymmetrical, right-sided, skewed distribution. When log-Gaussian curves are also not normally distributed, means with (1.96 times) SDs cannot be used to yield the central 95% of the reference population of subjects. Rather, percentile values (e.g., 97.5th, 95th or 90th) can be used, as this non-parametric method makes no specific assumption regarding the distribution from which the data are obtained. Nevertheless, extreme values can still have a profound effect in defining reference values and, therefore, sample sizes (in subgroups) of at least 120 are needed (for 90% confidence intervals) to reduce the amount of uncertainty [19,55]. Second, outliers can be removed before the analysis, using outlier detection methodology. For example, if the difference between the extreme and the next most extreme value exceeds 1/3rd of the range, the extreme value can be deleted (i.e., the Dixon test method) [19]; this may yield better reference values. However, an attempt should first be made to determine whether these extremes are errors in the assessment procedure. Third, there may be a profound influence from healthy and nonhealthy (psychiatrically ill individuals) individuals on the estimation of reference values. About 10% narrower reference intervals will be derived from samples that excluded nonhealthy subjects [56] but could make the reference range unreasonably narrow. Therefore, we chose to study a 'control' group rather than a 'healthy' group. Overall, there are many trade-offs between the different parametric, transformed parametric, and nonparametric methodologies.

Reference values for psychiatric instruments are essential for patient care. In this ROM reference group, data were collected enabling the calculation of reference values for 22 ROM instruments that often lack these values, because recruiting valid groups of reference subjects is costly and time intensive. These reference values are of major clinical importance because they can help to weigh the severity of symptoms and provide criteria that signify the transition from illness to health, and potential treatment termination. They can also be used by primary care physicians for referral to secondary care, and vice versa. Additionally, reference material to facilitate research on genetic characteristics (DNA) and the HPA axis stress system was collected.

Our study has specific strengths. First, to yield reliable and stable reference values the group has to be of sufficient size and representative for the patient group of interest. Tests for decisions at the individual level such as therapy indication or monitoring require a sample size of at least 250 subjects per reference group standardized for age and gender [57,58]. The size of the group and four subgroups surpassed this number and the previously described size of the 120 recommended participants [19,55], even when partitioning the test subjects by gender and age groups. Second, the diagnostic interview was structured leading to better identification of diagnostic comorbidity than unstructured interviews [59]. Next to self report data, observational data were collected using the MINI-Plus. This approach provided

comprehensive clinical information according to international standards (DSM-IV). Third, standardization of the interviews was assured, as both observation scales and self-report questionnaires were administered via a web-based computer program, implying a fixed order in administration of instruments with no instruments skipped or data missing, and no errors due to manually entering data. Fourth, recruitment through GPs allowed for a good description of the sample characteristics. Furthermore, contacting possible participants by telephone presumably increased the response rate. Finally, an on-going quality control and calibration among interviewers ensured that a high quality was maintained during data collection.

The present study also has some limitations. First, because recruitment of the ROM reference and patient group took place in the Dutch region of Leiden, reference values may not be directly internationally generalizable. Moreover, because ethnic participants formed a minority, generalizability of reference values to other countries and ethinicities is limited. Second, children and elderly were not included, thus requiring their own reference group studies. Third, non-response was significant, involving a possible, unknown bias. Finally, information about the characteristics of those who did not participate is lacking. It is unclear whether non-responders differed in a systematic way from the participating subjects.

In conclusion, we succeeded in collecting extensive data from 1302 persons from the general population, enabling the calculation of reference values for 22 ROM instruments. The results of the reference values are expected to become available within the next two years and will be useful for current and future diagnostic and research purposes in patients with MAS disorders.

Reference List

1. De Beurs E, Den Hollander-Gijsman ME, Van Rood YR, Van der Wee NJ, Giltay EJ, Van Noorden MS, Van der Lem R, Van Fenema EM, Zitman FG. (2011) Routine outcome monitoring in the Netherlands: practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clin Psychol Psychother,* 18, 1-12.

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

3. Miller SD, Duncan BL, Sorrell R, Brown GS. (2005) The partners for change outcome management system. *J Clin Psychol*, 61 (2), 199-208.

4. Bijl RV, Ravelli A, Van Zessen G. (1998) Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc Psychiatry Psychiatr Epidemiol,* 33 (12), 587-595.

5. Van Noorden MS, Giltay EJ, Den Hollander-Gijsman ME, Van der Wee NJ, Van Veen T, Zitman FG. (2010) Gender differences in clinical characteristics in a naturalistic sample of depressive outpatients: the Leiden Routine Outcome Monitoring Study. *J Affect Disord*, 125 (1-3), 116-123.

6. Linnet K. (2000) Nonparametric estimation of reference intervals by simple and bootstrap-based procedures. *Clin Chem*, 46 (6 Pt 1), 867-869.

7. Overbeek LI, Kapusta L, Peer PG, De Korte CL, Thijssen JM, Daniels O. (2006) New reference values for echocardiographic dimensions of healthy Dutch children. *Eur J Echocardiogr*, 7 (2), 113-121.

8. Carlier IVE, Meuldijk D, Van Vliet IM, Van Fenema EM, Van der Wee NJ, Zitman FG. (2012) Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *J Eval Clin Pract*, 18 (1), 104-110..

9. Van Noorden MS, Minkenberg SE, Giltay EJ, Den Hollander-Gijsman ME, Van Rood YR, Van der Wee NJ, Zitman FG. (2011) Pre-adult versus adult onset major depressive disorder in a naturalistic patient sample: the Leiden Routine Outcome Monitoring Study. *Psychol Med,* 41 (7), 1407-1417.

10. Sullivan PF, Neale MC, Kendler KS. (2000) Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry*, 157 (10), 1552-1562.

11. Uhl GR, Grow RW. (2004) The burden of complex genetics in brain disorders. *Arch Gen Psychiatry,* 61 (3), 223-229.

12. Carroll BJ, Cassidy F, Naftolowitz D, Tatham NE, Wilson WH, Iranmanesh A, Liu PY, Veldhuis JD. (2007) Pathophysiology of hypercortisolism in depression. *Acta Psychiatr Scand Suppl* (433), 90-103.

13. Pariante CM, Lightman SL. (2008) The HPA axis in major depression: classical theories and new developments. *Trends in Neurosciences,* 31 (9), 464-468.

14. Vreeburg SA, Hoogendijk WJ, DeRijk RH, Verhagen JC, Van Dyck R, Smit JH, Zitman FG, Penninx BW. (2009) Major depressive disorder and hypothalamic-pituitary-adrenal axis activity: results from a large cohort study. *Arch Gen Psychiatry,* 66 (6), 617-626.

15. De Kloet ER, Joels M, Holsboer F. (2005) Stress and the brain: from adaptation to disease. Nat Rev Neurosci, 6 (6), 463-475.

16. Vreeburg SA, Zitman FG, van PJ, DeRijk RH, Verhagen JC, van DR, Hoogendijk WJ, Smit JH, Penninx BW. (2010) Salivary cortisol levels in persons with and without different anxiety disorders. *Psychosom Med*, 72 (4), 340-347.

17. Kendall PC, Marrs-Garcia A, Nath SR, Sheldrick RC. (1999) Normative comparisons for the evaluation of clinical significance. *J Consult Clin Psychol*, 67 (3), 285-299.

18. Poortvliet MC, Lamkadden M, Deville W. Niet op naam ingeschreven (NONI) bij de huisarts. Inventarisatie en gevolgen voor de ziekenfondsverzekerden. Utrecht: NIVEL; 2005.

19. Horowitz GL, Altaie S, Boyd JC, Ceriotti F, Garg U, Horn P, et al. (2008) Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline - Third Edition CLSI Document C28-A3. Wayne, PA, Clinical and Laboratory Standards Institute.

20. Nannya Y, Taura K, Kurokawa M, Chiba S, Ogawa S. (2007) Evaluation of genome-wide power of genetic association studies based on empirical data from the HapMap project. *Hum Mol Genet*, 16 (20), 2494-2505.

21. Penninx BW, Beekman AT, Smit JH, et al. (2008) The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int J Methods Psychiatr Res*, 17 (3), 121-140.

22. Bernstein DP, Stein JA, Newcomb MD, et al. (2003) Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse Negl*, 27 (2), 169-190.

23. Endicott J, Spitzer RL, Fleiss JL, Cohen J. (1976) The global assessment scale. A procedure for measuring overall severity of psychiatric disturbance. *Arch Gen Psychiatry,* 33 (6), 766-771.

24. Scheier MF, Carver CS, Bridges MW. (1994) Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *J Pers Soc Psychol*, 67 (6), 1063-1078.

25. Ware JE, Snow KK, Kosinski M, Gandek B. (1993) SF-36 Health Survey Manual and Interpretation Guide. Boston, New England Medical Center, The Health Institute.

26. Derogatis LR. (1975) The Brief Symptom Inventory. Baltimore, MD., Clinical Psychometric Research.

27. Livesley WJ, Jackson DN. (2002) Manual for the dimensional assessment of personality pathology - basic questionnaire (DAPP-BQ). Port Huron, Sigma Press.

28. Creamer M, Bell R, Failla S. (2003) Psychometric properties of the Impact of Event Scale - Revised. *Behav Res Ther*, 41 (12), 1489-1496.

29. Horowitz M, Wilner N, Alvarez W. (1979) Impact of Event Scale: a measure of subjective stress. *Psychosom Med*, 41 (3), 209-218.

30. Clark LA, Watson D. (1991) Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *J Abnorm Psychol*, 100 (3), 316-336.

31. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC. (1998) The Mini-International Neuropsychiatric Interview

M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry,* 59 Suppl 20, 22-33.

32. Carlier I, Giltay E, Vergeer P. (2012) Development and validation of the 48-item Symptom Questionnaire (SQ-48) in patients with depressive, anxiety and somatoform disorders. *Psychiatry Res*, 200 (2-3), 904-910.

33. Asberg M, Montgomery SA, Perris C, Schalling D, Sedvall G. (1978) A comprehensive psychopathological rating scale. *Acta Psychiatr Scand Suppl* (271), 5-27.

34. Donker T, Van Straaten A, Marks I, Cuijpers P. A brief web screening questionnaire for common mental disorders: development and validation. submitted. In press 2009.

35. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. (1996) The Inventory of Depressive Symptomatology (IDS): psychometric properties. *Psychol Med*, 26 (3), 477-486.

36. Beck AT, Steer RA. (1984) Internal consistencies of the original and revised Beck Depression Inventory. *J Clin Psychol*, 40 (6), 1365-1367.

37. Ost LG. (1990) The Agoraphobia Scale: an evaluation of its reliability and validity. *Behav Res Ther*, 28 (4), 323-329.

38. Sanavio E. (1988) Obsessions and compulsions: the Padua Inventory. *Behav Res Ther,* 26 (2), 169-177.

39. Telch MJ, Brouillard M, Telch CF, Agras WS, Taylor CB. (1989) Role of cognitive appraisal in panic-related avoidance. *Behav Res Ther*, 27 (4), 373-383.

40. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. (1990) Development and validation of the Penn State Worry Questionnaire. *Behav Res Ther*, 28 (6), 487-495.

41. Brown EJ, Turovsky J, Heimberg RG, Juster HR, Brown TA, Barlow DH. (1997) Validation of the social interaction anxiety scale and the social phobia scale across the anxiety disorders. Psychological Assessment, 9 (1), 21-27.

42. Mattick RP, Clarke JC. (1998) Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behav Res Ther*, 36 (4), 455-470.

43. Tallis F, Eysenck M, Mathews A. (1992) A Questionnaire for the Measurement of Nonpathological Worry. *Personality and Individual Differences*, 13 (2), 161-168.

44. Littleton HL, Axsom D, Pury CL. (2005) Development of the body image concern inventory. *Behav Res Ther*, 43 (2), 229-241.

45. Vercoulen JH, Swanink CM, Fennis JF, Galama JM, Van der Meer JW, Bleijenberg G. (1994) Dimensional assessment of chronic fatigue syndrome. *J Psychosom Res*, 38 (5), 383-392.

46. Pilowsky I. (1967) Dimensions of hypochondriasis. *Br J Psychiatry*, 113 (494), 89-93.

47. Bourdon KH, Goodman R, Rae DS, Simpson G, Koretz DS. (2005) The Strengths and Difficulties Questionnaire: U.S. normative data and psychometric properties. *J Am Acad Child Adolesc Psychiatry*, 44 (6), 557-564.

48. Noerholm V, Groenvold M, Watt T, Bjorner JB, Rasmussen NA, Bech P. (2004) Quality of life in the Danish general population - normative data and validity of WHOQOL-BREF using Rasch and item response theory models. *Quality of Life Research*, 13 (2), 531-540.

49. Reedtz C, Bertelsen B, Lurie J, Handegard BH, Clifford G, Morch WT. (2008) Eyberg Child Behavior Inventory (ECBI): Norwegian norms to identify conduct problems in children. *Scand J Psychol*, 49 (1), 31-38.

50. Solberg HE. (2008) Establishment and use of reference values. Burtis CA, Ashwood ER, Bruns DE, editors. Fundamentals of clinical chemistry. 6[14], 229-238. St. Louis, Missouri, Saunders Elsevier.

51. Barnabei L, Marazia S, De CR. (2007) Receiver operating characteristic (ROC) curves and the definition of threshold levels to diagnose coronary artery disease on electrocardiographic stress testing. Part I: The use of ROC curves in diagnostic medicine and electrocardiographic markers of ischaemia. *J Cardiovasc Med (Hagerstown)*, 8 (11), 873-881.

52. Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler KS. (1994) Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Arch Gen Psychiatry*, 51 (1), 8-19.

53. De Waal MW, Arnold IA, Eekhof JA, Van Hemert AM. (2004) Somatoform disorders in general practice: prevalence, functional impairment and comorbidity with anxiety and depressive disorders. *Br J Psychiatry*, 184, 470-476.

54. Smith AL, Weissman MM. (1992) Epidemiology. Paykel ES, editor. Handbook of Affective Disorders. 111-129. New York, Guildford.

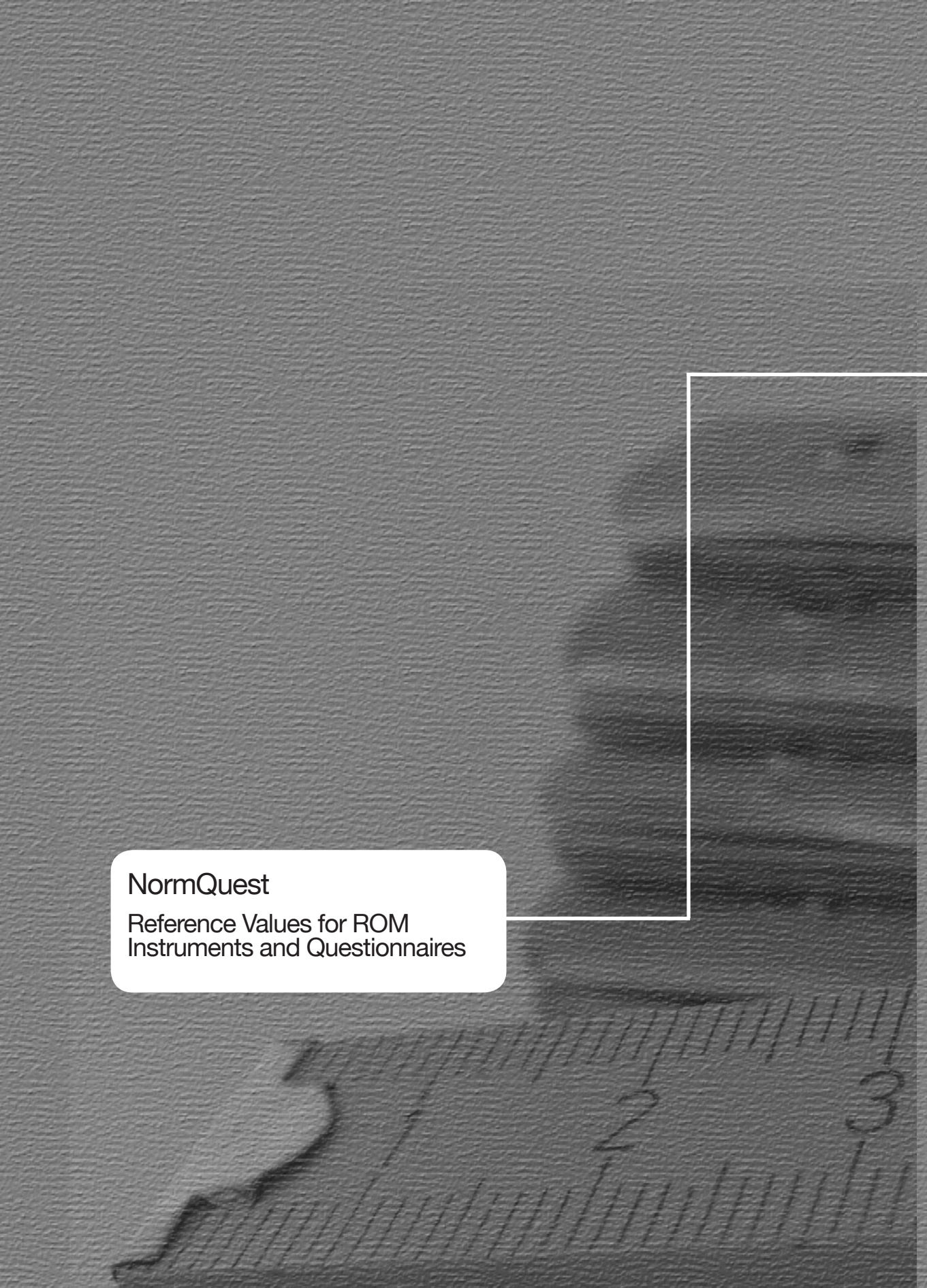
55. Reed AH, Henry RJ, Mason WB. (1971) Influence of statistical method used on the resulting estimate of normal range. *Clin Chem*, 17 (4), 275-284.

56. Horn PS, Feng L, Li Y, Pesce AJ. (2001) Effect of outliers and nonhealthy individuals on reference interval estimation. *Clin Chem*, 47 (12), 2137-2145.

57. Angoff WH. (1971) Scales, norms, and equivalent scores. Thorndike RL, editor. Educational Measurement. 508-600. Washington, DC, American Council on Education.

58. Bechger T, Hemker B, Maris G. (2009) Over het gebruik van continue normering. Arnhem, NL, CITO.

59. Zimmerman M, Mattia JI. (1999) Psychiatric diagnosis in clinical practice: is comorbidity being missed? *Compr Psychiatry*, 40 (3), 182-191.



NormQuest

Reference Values for ROM Instruments and Questionnaires

Chapter 3

Reference values for generic instruments used in Routine Outcome Monitoring

Yvonne W.M. Schulte-van Maaren
Ingrid V.E. Carlier
Frans G. Zitman
Albert M. van Hemert
Margot W.M de Waal
Martijn S. van Noorden
Erik J. Giltay

BMC Psychiatry (2012), 12:203

ABSTRACT

Background: The Brief Symptom Inventory (BSI), Mood & Anxiety Symptom Questionnaire -30 (MASQ-D30), Short Form Health Survey 36 (SF-36), and Dimensional Assessment of Personality Pathology-Short Form (DAPP-SF) are generic instruments that can be used in Routine Outcome Monitoring (ROM) of patients with common mental disorders. We aimed to generate reference values usually encountered in 'healthy' and 'psychiatrically ill' populations to facilitate correct interpretation of ROM results.

Method: We included the following specific reference populations: 1294 subjects from the general population (ROM reference group) recruited through general practitioners, and 5269 psychiatric outpatients diagnosed with mood, anxiety, or somatoform (MAS) disorders (ROM patient group). The outermost 5% of observations were used to define limits for one-sided reference intervals (95th percentiles for BSI, MASQ-D30 and DAPP-SF, and 5th percentiles for SF-36 subscales). Internal consistency and Receiver Operating Characteristics (ROC) analyses were performed.

Results: Mean age for the ROM reference group was 40.3 years (SD=12.6) and 37.7 years (SD=12.0) for the ROM patient group. The proportion of females was 62.8% and 64.6%, respectively. The mean for cut-off values of healthy individuals was 0.82 for the BSI subscales, 23 for the three MASQ-D30 subscales, 45 for the SF-36 subscales, and 3.1 for the DAPP-SF subscales. Discriminative power of the BSI, MASQ-D30 and SF-36 was good, but it was poor for the DAPP-SF. For all instruments, the internal consistency of the subscales ranged from adequate to excellent.

Discussion and conclusion: Reference values for the clinical interpretation were provided for the BSI, MASQ-D30, SF-36, and DAPP-SF. Clinical information aided by ROM data may represent the best means to appraise the clinical state of psychiatric outpatients.

INTRODUCTION

Routine outcome monitoring (ROM) was developed to enhance the effectiveness of Routine Outcome Monitoring (ROM) is a method for the continuous monitoring of patients' symptomatic and functional status. It provides the clinician with systematic information on type and severity of psychiatric complaints before, during, and after treatment. The web-based ROM assessment battery, which is used in the Leiden ROM Study, comprises both generic and disorder-specific measurement instruments. Generic instruments can be used to assess a broad range of psychopathological symptoms, maladaptive personality traits, and quality of life in any patient irrespective of their psychiatric disorder(s) [1]. In contrast, disorder-specific instruments are administered only to those patients who meet the criteria for a particular disorder.

Responsible clinical decision making (e.g., regarding the effectiveness and possible termination of treatment or referral from primary care to specialized mental health care and vice versa), based on ROM assessment, depends on the correct interpretation of the measures. Correct interpretation is only possible if patients' ROM data can be compared to reliable reference values (from a reference population).

Reference values [2] are often established in healthy populations [3]. Health, a relative condition lacking a universal definition, should nevertheless be clearly defined, a priori, via inclusion and exclusion criteria [4-6]. In non-realistic 'supernormal' (i.e., too healthy) reference groups [7] unreasonable narrow reference intervals can be expected. Horn and colleagues (2001) studied the effect of including physician-determined non-healthy individuals in a reference sample. Physician-defined healthy groups with and without non-healthy individuals were compared. Even in healthy samples, outliers may exist. There are marked effects to be expected of non-healthy individuals in the computation of reference values. As non-healthy individuals likely increase the chance of outliers, the width of reference intervals may increase by about 10% [8]. Thus, if non-healthy individuals are included in the reference group, then some subjects would be categorized as having responded to treatment. This would not have happened if only healthy individuals were included. Outlier removal would be an alternative methodology applied in the generation of reference values. Since extreme values can have a profound effect in establishing reference values, sample sizes of at least 120 (after partitioning in relevant subclasses) are needed to reduce the amount of uncertainty and error [9]. Common reference values are means and standard deviations (SDs), which can help to determine whether an individual or a group scores below or above the average of the 'healthy' or the 'psychiatrically ill' subjects. Also, percentile scores are often used as reference values. These non-parametric values do not rely on Gaussian data distributions [3,9]. The lower interval, bounded by the 95th percentile, commonly serves as the reference group [3]. When both reference and patient group data are available, Receiver Operating Characteristics (ROC) analyses can provide additional cut-offs, reflecting the

trade-off between sensitivity (measure of positivity; the proportion of actual positives correctly identified as such) and specificity (measure of negativity; the proportion of negatives which are legitimately ruled out) [10].

Some frequently used generic self-report ROM instruments include the Brief Symptom Inventory (BSI) [11,12], the Mood & Anxiety Symptom Questionnaire -30 (MASQ-D30) [13,14], the Short Form Health Survey 36 (SF-36) [15,16], and the Dimensional Assessment of Personality Pathology - Short Form (DAPP-SF) [17,18]. In this generic set of instruments the DAPP-SF is intended not so much for Axis II diagnoses of psychopathology according to the DSM-IV but for the assessment of (dysfunctional) personality traits.
Previous studies mainly reported means and SDs for the general population for the BSI [11,19] and SF-36 [15,20-22], and for the general population and psychiatric patients for the DAPP-SF [18,23], while for the MASQ-D30 no such reference values have been published. Except for the BSI [11], no clinically relevant cut-off scores between 'healthy' and 'psychiatrically ill' have been reported. In most of the studies the population-based reference groups were relatively small, ranging from 200 [11] to 719 [19] for the BSI, and between 51 [24] and 478 [18,23] for the DAPP-SF, leading to somewhat imprecise reference values [4,8]. Reference values subcategorized according to gender and age were reported for the SF-36 [20,21,25] but they are not available for the BSI, MASQ-D30 or DAPP-SF.

We aimed to establish reference values, means and SDs, percentile scores, and cut-off points, for a comprehensive set of generic ROM instruments that can be offered to every patient referred for (but not necessarily diagnosed with) mood, anxiety, or somatoform (MAS) disorders. These comprise the vast majority of psychiatric patients, notwithstanding those with addiction disorders. In this set, the severity of general psychopathology, (dysfunctional) personality traits, and subjective mental and physical well-being are covered respectively by the BSI, the MASQ-D30, the DAPP-SF, and the SF-36. We tested an apparently healthy population of 1294 subjects who were recruited through general practitioners, and examined similar data from a 'psychiatrically ill' population of 5269 outpatients diagnosed with MAS disorders. A novel aspect of the current study is that we could include samples of sufficient size for both the healthy reference and the well-defined psychiatric outpatient group.

METHODS

Participants

The group of participants comprised a reference sample from the general population (ROM reference group) and a ROM sample of psychiatric outpatients (ROM patient group), as previously described in detail [26].

The ROM reference group consisted of 1294 participants aged 18 to 65 years (62.8% females; mean age=40.3 years; SD=12.6) from the 'Leiden Routine Outcome Monitoring Study'. The study design, objectives, and methods have been described elsewhere [26,27]. Participants were randomly selected from registration systems of eight general practitioners (GPs) in the province South-Holland, the Netherlands. In the Netherlands, 99.9% of the general population is registered with a GP [28]. Therefore, non-consulting GP patients are a very good representation of the Dutch general population. The ROM reference group was stratified for gender, age, and urbanization-level (62.3% urban), to make the group demographically comparable to the ROM patient group. Invitations for this study were sent to 4840 persons; 1283 could not be contacted and 67 were not included because of time constraints. Of the remaining 3490 potential participants, 1302 were assessed and 1294 generated complete datasets, resulting in a response rate of 37.1%.

The ROM patient group consisted of 5269 psychiatric outpatients, aged 18 to 65 years (64.6% females; mean age=37.7, SD=12.0). They were diagnosed with and treated for one or more MAS disorders in the Leiden University Medical Center (LUMC) Department of Psychiatry or in the Rivierduinen Psychiatric Institute, the regional provider of specialized mental health care.

Procedures

Procedures for the web-based ROM program of the LUMC Department of Psychiatry are described elsewhere [27,29]. The participants in the ROM reference group were assessed in a similar way to the ROM patient group. Subjects from the ROM reference group completed the self-report instruments BSI, MASQ-D30, and SF-36, and due to time constraints, a random sample of 50% completed the DAPP-SF [26]. The BSI, MASQ-D30, and SF-36 were completed by all 5269 subjects from the ROM patient group, while 234 (4.6%) did not complete the DAPP-SF, again due to time constraints. To facilitate diagnoses of psychopathology according to the DSM-IV, the procedure for the two groups included a standardized diagnostic interview (i.e., the Mini-International Neuropsychiatric Interview plus (MINI-Plus 5.0.0.) [30,31]). The Medical Ethical Committee of the LUMC approved the general study protocol regarding ROM, in which ROM was organized as part of the treatment process for patients. It involved a comprehensive protocol (titled "Psychiatric Academic

Registration Leiden database") which safeguarded the anonymity of patients and participants and ensured proper handling of the ROM data. All patients gave permission for the use of their ROM data for scientific purposes (written informed consent for this study was not required). In addition, participants of the ROM reference group (non-patients) signed informed consent for the purpose of this study.

Instruments

The BSI, a short version of the Symptom Checklist (SCL-90) [19], measures psychopathological symptoms. The BSI consists of 53 items divided into 9 subscales: Somatization (SOM), Obsessive-Compulsive (O-C), Interpersonal Sensitivity (I-S), Depression (DEP), Anxiety (ANX), Hostility (HOS), Phobic Anxiety (PHOB), Paranoid Ideation (PAR), and Psychoticism (PSY). Item scores range from 0 ("not-at-all") to 4 ("extremely"). The subscale and total scores are calculated as an average of the relevant items, with higher scores indicating more severe psychopathology.

The MASQ-D30 measures the dimensions of Clark and Watson's tripartite model, covering both shared and distinct symptoms of depression and anxiety [13,14]. The MASQ-D30 consists of 30 items, divided into three subscales: Negative Affect (NA), associated with both depression and anxiety; lack of Positive Affect (PA), associated with depressive moods; and Somatic Arousal (SA), associated with anxiety. The items are rated on a 5-point Likert scale, with scores ranging from 1 ("not at all") to 5 ("extremely"). Subscale scores are calculated as the sum of the relevant items, ranging from 10 to 50, with higher scores indicating more severe psychopathology.

The SF-36, derived from the Rand Medical Outcome Study (MOS) [15,16], measures functional health status and well-being. It can be used as a population-based assessment of quality of life. The SF-36 consists of 36 items divided into eight subscales: Physical Functioning, Role limitations due to Physical health problems (Role-Physical), Bodily Pain, Social Functioning, General Mental Health (Mental Health), Role limitations due to Emotional problems (Role-Emotional), Vitality, General Health Perceptions (General Health) and a question about perceived change of health during the last year (Health Transition). Subscale scores are calculated as the sum of the relevant items, ranging from 0 to 100, with higher scores indicating better functioning.

The DAPP-SF, the short form of the Dimensional Assessment of Personality Pathology – Basic Questionnaire (DAPP-BQ) [17,18], measures personality pathology. It consists of 136 items divided into 18 subscales: Submissiveness, Cognitive Distortion, Identity Problems, Affective Lability, Stimulus Seeking, Compulsivity, Restricted Expression, Callousness, Oppositionality, Intimacy Problems, Rejection, Anxiousness, Conduct Problems, Suspiciousness, Social Avoidance, Narcissism, Insecure Attachment, and Self-harm. Item scores range between 1 ("very unlike me") and 5 ("very like me"). Subscale scores are calculated as an average of the relevant items, ranging from 1 to 5, with higher scores indicating more maladaptive personality traits.

The Dutch version of the Mini-International Neuropsychiatric Interview plus (MINIplus 5.0.0.) [30,32] was used to establish the presence of Axis I diagnoses according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). This standardized diagnostic interview comprises 23 modules for mood, anxiety, psychotic, somatoform, and eating disorders.

Statistical analyses

Means, standard deviations (SDs), and percentile scores were calculated for the two samples separately, while ROC analyses were performed in the combined groups. In both samples, subjects with 1 or more missing values per subscale were excluded. This allowed us to conduct a robust evaluation of the use of the instruments. The occurrence of missing values is not completely random, and it depends on unobserved predictors. Therefore we decided to use an almost complete-case analysis, as bias due to missing values was likely to be small due to the small percentage (i.e., 0.01%) of cases that needed to be excluded. A descriptive analysis of sociodemographic and psychopathological variables was performed, using percentages in the case of categorical variables and means and SDs for the continuous variables. Internal consistency was assessed using Cronbach's alpha, with >0.70 indicating adequate internal consistency. ROC analyses provided cut-off scores, indicating an optimal discrimination threshold between 'healthy' (reference population) and 'psychiatrically ill' (psychiatric outpatients). The cut-off was chosen at the value representing equal sensitivity and specificity, since this is the point that yields the best compromise between specificity and sensitivity, with the lowest number of false results (false positive plus false negative). The areas under the ROC curve (AUCs) were calculated to indicate the discriminatory power of the instrument (sub) scales, where AUCs over 0.75 were considered clinically useful with 0.85 showing moderate discriminatory power and 0.95 very high discriminatory power [33]. Furthermore, means and SDs were calculated, together with $5^{th,}$ 25^{th}, 50^{th}, 75^{th} and 95^{th} percentile scores. When instruments merely assess the level of dysfunctionality, and the discriminative power to detect the level of 'health' or normal functionality is limited (i.e., no persons can be earmarked as 'abnormally healthy or good functioning'), the lowest 2.5% is irrelevant. Therefore, the top 5% (or lower 5% in case of SF-36 subscales) was chosen as representing 'abnormal'. Reference values were also presented for 4 subgroups: young women (aged 18-40 years), older women (aged 41-65 years), young men (aged 18-40 years), and older men (aged 41-65 years). SPSS for Windows version 17.0 (SPSS Inc., Chicago, IL, USA) was used for data analysis. To test our decision not to exclude those individuals in the ROM reference group with a current psychiatric diagnosis, we performed a sensitivity analysis.

RESULTS

Sociodemographic and psychiatric characteristics of the samples

The sociodemographic and psychiatric characteristics of the ROM reference group and the ROM patient group are shown in Table 3.1.

Table 3.1: Sociodemographic and psychiatric characteristics of the ROM reference (n=1294) patient (n=5269) groups.

	ROM reference group	ROM patient group
Gender (%)		
Male	481 (37.2)	1864 (35.4)
Female	813 (62.8)	3405 (64.6)
Age (mean, SD) in years	40.3 (12.6)	37.7 (12.0)
Male	41.3 (12.6)	39.1 (11.9)
Female	39.8 (12.6)	36.9 (12.0)
Marital status (%)		
Married/cohabitating	890 (68.8)	25.19 (47.8)*
Divorced/separated/widow	77 (6.0)	688 (13.1)*
Single	327 (25.2)	1730 (32.8)*
Housing situation (%)		
Living alone	201 (15.5)	1128 (21.4)*
Living with partner	903 (69.8)	2568 (48.7)*
Living with family	190 (14.7)	1241 (23.6)*
Educational status (%)* **		
Lower	295 (22.8)	2112 (40.1)*
Higher	999 (77.2)	2824 (53.6)*
Employment status (%)		
Employed part-time	512 (39.6)	1141 (21.7)*
Employed full-time	552 (42.7)	1105 (21.0)*
Unemployed/retired	194 (15.0)	1337 (25.4)*
Work-related disability (%)	36 (2.7)	1354 (25.7)*
Ethnic background (%)		
Dutch	1163 (89.9)	4335 (82.3)
Other ethnicity	131 (10.1)	934 (17.7)

Table 3.1 continued

	ROM reference group	ROM patient group
MINI diagnoses (%)		
Currently None	1153 (89.1)	0**
Anxiety disorder	53 (4.1)	1449 (27.5)
Mood disorder	7 (0.5)	1573 (29.9)
Somatoform disorder	41 (3.2)	403 (7.6)
Anxiety & Mood disorders	7 (0.5)	1257 (23.9)
Anxiety & Somatoform disorders	9 (0.7)	172 (3.3)
Mood & Somatoform disorders	1 (0.1)	228 (4.3)
Anxiety & Mood & Somatoform	2 (0.2)	187 (3.5)
Total Anxiety disorder	71 (5.5)	3065 (58.2)
Total Mood disorder	17 (1.3)	3245 (61.6)
Total Somatoform disorder	53 (4.1)	990 (18.8)

*No data from 332 (6.3%) patients
**Selection criterion
*** Lower educational status: general basic education or lower vocational education; higher educational status: middle or higher vocational education, college or university.

Mean age (40.3 years versus 37.7 years, p<.001) and gender distribution (62.8% females versus 64.6% females, p=.80) were comparable for the ROM reference group and the ROM patient group, as expected due to the stratification. The ROM reference group showed higher levels of education (77.2% versus 53.6% higher education), were more often married (68.8% versus 47.8%), and were less often living alone (15.5% versus 21.4%) relative to the ROM patient group. Unemployment and work-related disability were less prevalent in the ROM reference group (17.7% versus 51.1%). In keeping with our decision to exclude patients without a MINI diagnosis, all subjects from the ROM patient group had at least one DSM-IV disorder. In the ROM reference group, on the other hand, 10.9% had a DSM-IV disorder.

REFERENCE VALUES

Percentiles, means and SDs

Table 3.3 presents the percentile scores and mean values of the BSI, SF-36, and MASQ-D30 subscales for the ROM reference group and the ROM patient group. For the ROM reference group, the distribution of each total score and subscale score was positively skewed, showing apparent health. This was also demonstrated by the substantial percentage of participants having the lowest possible scores (highest for the SF-36). For apparently healthy individuals, the mean of cut-off (P_{95}) values was 0.82 for the BSI subscales, 23 for the three MASQ

dimensions, 45 for the SF-36 subscales, and 3.1 for the DAPP-SF subscales. By contrast, the mean of P_5 values for the SF-36 subscales was 45.

Table 3.2. Internal consistency and cut-off scores in combined ROM reference (n=1294) and patient (n=5269) groups for four generic Routine Outcome Monitoring instruments.

	Number of items	Cronbach's Alpha	ROC cut-off	AUC	Sensitivity / specificity
Brief Symptom Inventory (BSI)					
Somatization (SOM)	7	0.86	0.23	0.87	0.80
Obsessive-Compulsive (O-C)	6	0.88	0.69	0.91	0.84
Interpersonal Sensitivity (I-S)	4	0.83	0.54	0.88	0.81
Depression (DEP)	6	0.91	0.50	0.93	0.87
Anxiety (ANX)	6	0.89	0.50	0.92	0.85
Hostility (HOS)	5	0.86	0.30	0.82	0.75
Phobic Anxiety (PHOB)	5	0.83	0.25	0.90	0.84
Paranoid Ideation (PAR)	5	0.84	0.37	0.83	0.76
Psychoticism (PSY)	5	0.77	0.37	0.92	0.85
BSI total score*	53	0.97	0.48	0.96	0.90
MASQ-D30					
General distress (GD)	10	0.84	19.0	0.96	0.90
Anhedonic depression (AD)	10	0.92	23.0	0.88	0.80
Anxious arousal (AA)	10	0.74	18.0	0.99	0.96
Short Form 36 (SF36)*					
Physical Functioning	10	0.92	93.5	0.76	0.68
Role-Physical	4	0.88	82.5	0.82	0.78
Bodily Pain	2	0.87	83.7	0.72	0.68
Social Functioning	2	0.85	72.9	0.92	0.79
Mental Health	5	0.90	63.0	0.95	0.89
Role-Emotional	3	0.83	79.6	0.88	0.88
Vitality	4	0.84	52.5	0.92	0.85
General Health	5	0.84	67.5	0.82	0.76

Table 3.2. continued.

	Number of items	Cronbach's Alpha	ROC cut-off	AUC	Sensitivity / specificity
DAPP-SF:					
Submissiveness	8	0.87	2.40	0.76	0.71
Cognitive Distortion	6	0.84	1.55	0.83	0.76
Identity Problems	6	0.87	2.08	0.90	0.83
Affective Lability	8	0.86	2.56	0.85	0.77
Stimulus Seeking	8	0.81	1.94	0.55	0.54
Compulsivity	8	0.84	2.69	0.60	0.57
Restricted Expression	8	0.82	2.75	0.78	0.71
Callousness	10	0.79	1.65	0.53	0.51
Oppositionality	9	0.87	2.22	0.79	0.73
Intimacy Problems	9	0.79	2.18	0.60	0.57
Rejection	8	0.83	2.36	0.56	0.55
Anxiousness	6	0.84	2.64	0.85	0.78
Conduct Problems	7	0.73	1.14	0.57	0.56
Suspiciousness	7	0.90	1.40	0.78	0.72
Social Avoidance	6	0.88	2.20	0.80	0.73
Narcissism	8	0.82	2.20	0.56	0.55
Insecure Attachment	6	0.89	2.10	0.80	0.74
Self-Harm	6	0.89	1.08	0.75	0.57

*Higher score corresponds with better functioning
AUC: Area under the curve; MASQ-D30: Mood & Anxiety Symptom Questionnaire 30-item short adaptation; DAPP-SF : Dimensional Assessment of Personality Pathology – short form; The optimal cut-off derived by the ROC analysis is defined by equal sensitivity and specificity
*The BSI total score comprises 4 additional items next to the subscale items.

Table 3.3. Percentile scores and mean values for generic Routine Outcome Monitoring instruments in the ROM reference (n=1294) and patient (n=5269) groups.

	ROM reference group						ROM patient group					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Brief Symptom Inventory (BSI)			n=1294							n=5269		
Somatization (SOM)	0.00	0.00	0.00	0.29	0.71	0.17 ± 0.28	0.00	0.43	0.86	1.43	2.71	1.03 ± 0.83
Obsessive-Compulsive (O-C)	0.00	0.00	0.17	0.50	1.17	0.35 ± 0.42	0.33	1.00	1.67	2.33	3.33	1.67 ± 0.95
Interpersonal Sensitivity (I-S)	0.00	0.00	0.00	0.50	1.00	0.29 ± 0.42	0.00	0.75	1.50	2.25	3.50	1.56 ± 1.04
Depression (DEP)	0.00	0.00	0.00	0.33	0.83	0.20 ± 0.34	0.17	0.83	1.67	2.50	3.50	1.68 ± 1.01
Anxiety (ANX)	0.00	0.00	0.17	0.33	0.83	0.22 ± 0.34	0.17	0.83	1.33	2.17	3.33	1.49 ± 0.94
Hostility (HOS)	0.00	0.00	0.20	0.20	0.80	0.20 ± 0.29	0.00	0.20	0.80	1.40	2.80	0.94 ± 0.86
Phobic Anxiety (PHOB)	0.00	0.00	0.00	0.20	0.60	0.11 ± 0.23	0.00	0.40	1.00	1.60	3.00	1.15 ± 0.93
Paranoid Ideation (PAR)	0.00	0.00	0.00	0.40	0.80	0.23 ± 0.35	0.00	0.40	1.00	1.80	3.00	1.15 ± 0.94
Psychoticism (PSY)	0.00	0.00	0.00	0.20	0.80	0.14 ± 0.28	0.20	0.60	1.20	1.80	2.80	1.23 ± 0.81
BSI total score	0.00	0.06	0.13	0.28	0.68	0.21 ± 0.25	0.34	0.79	1.23	1.75	2.66	1.33 ± 0.71
MASQ-D30			n=1294						n=5269			
General distress (GD)	10	11	12	15	23	13.8 ± 4.4	17	23	28	33	40	28.1 ± 6.9
Anhedonic depression (AD)	10	14	17	22	29	18.4 ± 5.8	17	24	31	37	44	30.7 ± 8.3
Anxious arousal (AA)	10	10	11	13	17	11.9 ± 3.0	18	26	31	37	43	31.3 ± 7.5
Short Form 36 (SF36)*			n=1294							n=5269		
Physical Functioning	65	90	100	100	100	92.6 ± 14.2	25	60	80	95	100	74.8 ± 23.7
Role-Physical	13	100	100	100	100	87.0 ± 27.2	0	0	25	75	100	37.2 ± 39.7
Bodily Pain	54	78	90	100	100	86.4 ± 17.6	20	45	67	90	100	65.9 ± 27.5
Social Functioning	63	88	100	100	100	89.9 ± 15.6	0	25	50	63	88	44.8 ± 26.1
Mental health	56	72	80	88	96	79.7 ± 12.3	12	28	40	52	76	41.5 ± 18.2
Role-Emotional	33	100	100	100	100	90.4 ± 24.8	0	0	0	33	100	28.2 ± 36.2
Vitality	40	60	70	80	90	68.6 ± 15.3	5	20	35	45	65	34.3 ± 17.8
General Health	45	65	80	90	100	76.2 ± 16.3	20	35	50	65	90	51. 6 ± 21.0

Table 3.3. continued.

	ROM reference group						ROM patient group					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
DAPP-SF			n=635						n=5035			
Submissiveness	1.13	1.50	2.00	2.50	3.50	2.10 ± 0.75	1.25	2.25	3.00	3.63	4.38	2.94 ± 0.94
Cognitive Distortion	1.00	1.00	1.17	1.50	2.33	1.36 ± 0.51	1.00	1.50	2.33	3.00	4.17	2.36 ± 0.96
Identity Problems	1.00	1.00	1.33	1.83	2.70	1.54 ± 0.59	1.33	2.33	3.17	3.83	4.67	3.12 ± 1.02
Affective Lability	1.00	1.38	1.88	2.50	3.50	2.01 ± 0.76	1.63	2.63	3.38	3.88	4.63	3.24 ± 0.88
Stimulus Seeking	1.10	1.38	1.88	2.38	3.38	1.99 ± 0.72	1.00	1.50	2.00	2.63	3.75	2.13 ± 0.81
Compulsivity	1.38	2.00	2.50	3.13	4.00	2.58 ± 0.77	1.38	2.13	2.88	3.63	4.50	2.89 ± 0.94
Restricted Expression	1.25	1.75	2.25	2.88	3.63	2.33 ± 0.75	1.75	2.63	3.25	3.88	4.63	3.23 ± 0.86
Callousness	1.00	1.30	1.60	2.00	2.60	1.69 ± 0.50	1.00	1.30	1.70	2.10	2.90	1.77 ± 0.60
Oppositionality	1.00	1.40	1.80	2.30	3.20	1.91 ± 0.65	1.40	2.20	2.80	3.50	4.30	2.83 ± 0.89
Intimacy Problems	1.13	1.63	2.13	2.50	3.38	2.14 ± 0.67	1.13	1.75	2.38	2.88	4.00	2.42 ± 0.85
Rejection	1.38	1.88	2.50	3.00	3.75	2.47 ± 0.76	1.13	1.63	2.25	2.88	3.75	2.31 ± 0.82
Anxiousness	1.00	1.33	1.83	2.50	3.50	2.03 ± 0.81	1.67	2.67	3.50	4.00	4.83	3.37 ± 0.94
Conduct Problems	1.00	1.00	1.13	1.38	2.13	1.26 ± 0.37	1.00	1.00	1.25	1.63	2.63	1.43 ± 0.57
Suspiciousness	1.00	1.00	1.13	1.50	2.15	1.32 ± 0.46	1.00	1.38	2.00	2.88	4.00	2.18 ± 0.99
Social Avoidance	1.00	1.17	1.67	2.17	3.33	1.82 ± 0.73	1.17	2.17	3.00	3.83	4.67	2.98 ± 1.07
Narcissism	1.00	1.63	2.13	2.63	3.50	2.18 ± 0.76	1.10	1.75	2.25	2.88	3.88	2.36 ± 0.83
Insecure Attachment	1.00	1.17	1.50	2.17	3.33	1.74 ± 0.77	1.00	2.00	2.83	3.83	4.83	2.91 ± 1.13
Self-Harm	1.00	1.00	1.00	1.00	1.50	1.07 ± 0.27	1.00	1.00	1.33	2.33	3.67	1.76 ± 0.96

*Higher score corresponds with better functioning
MASQ-D30 denotes Mood & Anxiety Symptom Questionnaire 30-item short adaptation; DAPP-SF denotes Dimensional Assessment of Personality Pathology – short form; P denotes percentile; SD denotes standard deviation
To calculate sum scores for the DAPP-SF subscales, multiply the mean scores by the number of items per subscale

The BSI subscale scores ranged between 0 and 4. The P_{95} reference scores for the BSI subscales ranged between 0.60 for Phobic Anxiety (PHOB) and 1.17 for Obsessive-Compulsive (O-C) 1.17; for the BSI total score it was 0.68. For six of the nine subscales, the median value (P_{50}) was equal to the minimum possible score of 0.
The MASQ-D30 subscale scores ranged between 10 and 50. The P_{95} reference scores for the three MASQ-D30 subscales were: General Distress (GD) - 23; Anhedonic Depression (AD) - 29; and Anxious Arousal (AA) - 17.

The SF-36 subscale scores ranged between 0 and 100, with higher scores indicating better health. Therefore the P_5 indicates the cut-off for a low level of functioning. The P_5 reference scores for the SF-36 subscales ranged between 65 for Physical functioning and 33 for Emotional problems, with the exception of the P_5 value for Physical health problems, which was 13. The scales that measure well-being as well as health-related limitations (General Health,Vitality,Mental health) showed lower average values, as expected [33]. The other five health-related disability scales had the highest mean subscale scores. For four of the eight subscales, the median value (P_{50}) was equal to the maximum possible score of 100. The DAPP-SF subscale scores ranged between 1 and 5. The range of P_{95} reference scores for the 18 subscales was between 1.50 for Self-Harm and 4.00 for Compulsivity.

Analyses of gender and age indicated that advancing age was associated with more symptoms of psychopathology for both sexes (see Supplementary Tables 3.1 through 3.4). There was a tendency for healthy women to show higher cut-off scores on the BSI and the MASQ-D30 relative to healthy men, while the two sexes showed a different pattern of cut-off scores on the DAPP-SF. Men, and especially young men, reported better health as reflected in higher scores on several subscales of the SF-36.

In a sensitivity analysis, we excluded all 122 (9.5%) subjects in the ROM reference group who had a MINI-diagnosis. Among the remaining 1161 subjects, we found that the median scores on the BSI total score, MASQ-D30 subscales, SF-36 subscales, and DAPP-SF subscales changed on average 2% (interquartile range 1 to 6%). The median P_{95} scores (P_5 score for the SF36) changed on average 5% (interquartile range 0 to 18%).

Receiver operating characteristic (ROC) curves

The results of the ROC analyses are presented in Table 3.2.

BSI: The cut-off point of the BSI total score, which discriminated the ROM reference group from the ROM patient group, was 0.48, with a sensitivity and specificity of 90%. Therefore, for subjects without psychopathology, 10% with a total score of 0.48 or higher would be classified wrongly as a patient with psychopathology. By the same token, the 10% of subjects from the ROM patient group with a total score of 0.48 or lower would be classified wrongly as a psychiatrically ‘healthy’ subject. The AUC values showed that all BSI subscales performed well in making a distinction between patients and non-patients. The discriminating performance of the total score was excellent (AUC=0.96).

The best performing subscale was DEP, followed by ANX and PSY. The HOS and PAR subscales showed the least distinctiveness but might perform better in specific subpopulations of patients. Figure 3.1 presents the discriminative power of the BSI total score.

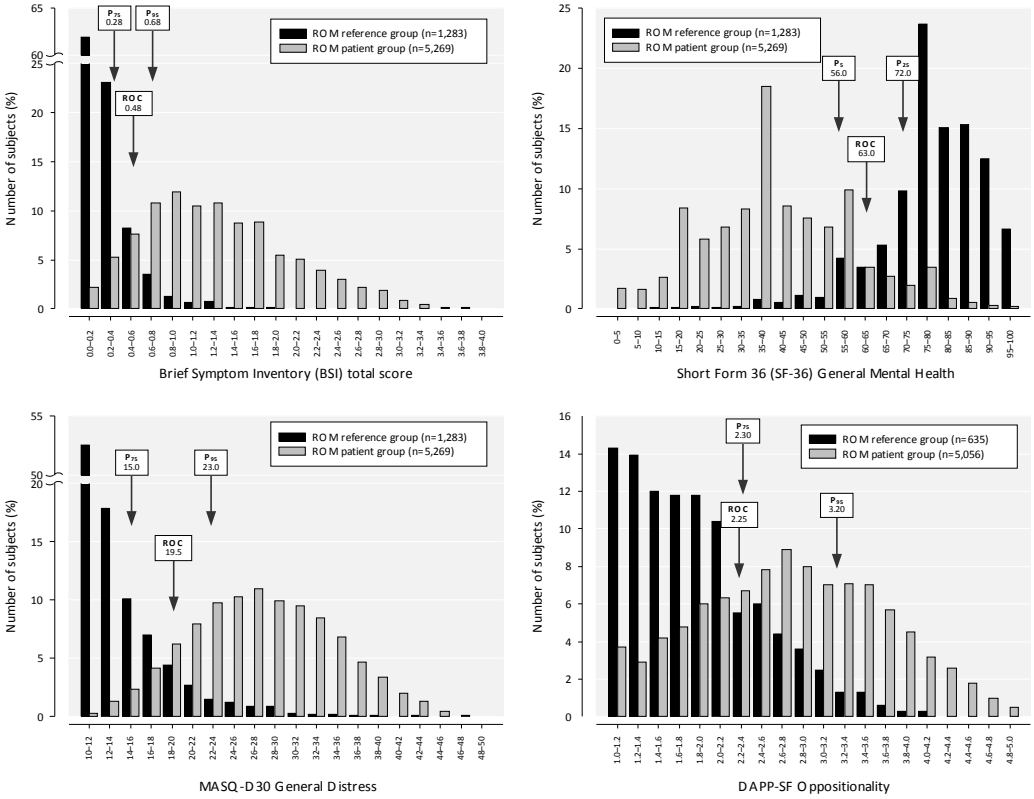


Figure 3.1: Distribution of the scores of Brief Symptom Inventory (BSI) total scale, and the subscales of Short Form-36 (SF-36) General Mental Health, Mood and Anxiety Symptom Questionnaire 30 (MASQ-D30) General Distress and Dimensional Assessment of Personality Pathology - Short Form (DAPP-SF) Oppositionality. Three types of cut-off points are depicted: the 75th percentile score (P_{75}), the 95th percentile score (P_{95}) and the Receiver Operating Characteristics (ROC) cut-off point defined by equal sensitivity and specificity. Note: in the SF-36 a higher score corresponds with better functioning

MASQ-D30: The cut-off score of 19 on the General Distress (GD) dimension, which discriminated the ROM reference group from the ROM patient group, had a sensitivity and specificity of 90%. For the cut-off of 23 on the Anhedonic Depression dimension, the sensitivity and specificity were only 80%. The cut-off score of 18 on the Anxious Arousal dimension, discriminating health from disease, had a sensitivity and specificity of 96%. The AUC values showed that all three scales performed well in discriminating between outpatients and non-patients. The most discriminating subscale was Anxious Arousal (AUC=0.99), followed by General Distress (AUC=0.96) and Anhedonic Depression (AUC=0.88). See Figure 3.1 for the discriminative power of the General Distress score.

SF-36: The cut-off point of the Mental Health score, which discriminated the ROM reference group from the ROM patient group, was 63, with a sensitivity and specificity of 89%. The AUC values showed that all SF-36 subscales performed well in making a distinction between patients and non-patients. The discriminating performance of Mental Health was excellent (AUC=0.95). The next best discriminating subscales were Social Functioning (AUC=0.92) and Vitality (AUC=0.92). The Bodily Pain and Physical Functioning scales showed the least distinctiveness, but they were still adequate, and are therefore still clinically useful. The discriminative power of General Mental Health is presented in Figure 3.1.

DAPP-SF: The cut-off point of the Identity Problems score, which discriminated the ROM reference group from the ROM patient group, was 2.08, with a sensitivity and specificity of 83%. The cut-off point of the Oppositionality score was 2.22 with a sensitivity and specificity of 73%. The discriminating performance of the DAPP-SF was moderate. The AUC values showed that 11 subscales performed well in distinguishing between patients and non-patients. The best performing subscale was Identity Problems (AUC=0.90), followed by Affective Lability (AUC=0.90) and Anxiousness (AUC=0.90). Seven subscales showed no clinically useful discriminatory power, with AUC values ranging from 0.53 to 0.60. All scales might perform better in the specific subpopulation of patients with personality disorders. As an example, the distributions of Oppositionality in the ROM reference group and the ROM patient group are presented in Figure 3.1. (This subscale was selected because it showed substantial interperson variability.)

DISCUSSION

We report reference values (95th percentiles) for the generic instruments BSI, MASQ-D30, SF-36 and DAPP-SF in large samples from 'healthy' and 'psychiatrically sick' populations. The internal consistency of the total score and subscale scores of the four generic instruments was consistently high. In the two samples, the expected differences in mean scores were confirmed, validating the clinical application of the ROC cut-off values or the 95th percentile scores (or 5th percentile for the SF-36). A clear gender difference in reference values was observed, with women showing higher values than men. It is remarkable that "healthy" men and women differed, and that the gender-specific distributions of the generic scales overlapped but did not coincide. Our data suggested that the degree of overlap between the sexes was not negligible, and that sex-specific reference values would increase the precision of the assessment of the clinical state of psychiatric outpatients. Advancing age was associated with more symptoms of Axis I psychopathology. Consequently, to be regarded as recovered, a young man would need to have lower scores on generic scales than would an older woman.

ROC analyses showed good discriminative power for the BSI, MASQ-D30, and SF-36 but not for the DAPP-SF subscales. The former three instruments address Axis-I psychopathology or distress, whereas the DAPP-SF measures Axis-II personality traits that are rather stable and less affected by psychopathology and treatment. The higher AUC values represent the more state-like than trait-like characteristics of the BSI, MASQ-D30, and SF-36, compared to the DAPP-SF.

The high internal consistency of the BSI, MASQ-D30, SF-36, and DAPP-SF are in accordance with previous studies [11,14,18,19,23,34]. Subscale means for the ROM reference group were somewhat lower than reported in previous studies of general population samples for the BSI [11,19]. In addition, they were slightly higher than in most [15,34-37] but not all [38] SF-36 studies and lower than in a DAPP-SF study [18]. Regarding the ROM patient group, means for the BSI, SF-36, and DAPP-SF approximated previously reported values in most clinical populations [11,15,19,23]. Previously, reference values subcategorized by gender and age have only been reported for the SF-36 [20,21,25]. Given that the assessment results for our ROM instruments generally had skewed distributions with a long tail toward the extreme values (i.e., lower in the case of the SF-36), we preferred percentile scores rather than means and SDs, in contrast to previous studies. For the BSI, ROC cut-off scores approximated cut-off scores with optimal sensitivity, as reported by De Beurs and Zitman (2006). Further, P^{95} reference scores approximated De Beurs and Zitman 's cut-off scores with optimal specificity [11]. Reference values derived from the ROM reference and patient groups have different functions. Reference values from the ROM reference and patient groups are important for screening a patient who is considered to have more than mild abnormalities. A precisely defined reference value will allow for the detection of subjects with psychopathology who could benefit from therapy or from referral from primary care to specialized mental health care (and vice versa). For screening purposes, we recommend the

use of cut-off scores with a high sensitivity, to be sure that a minimal number of patients with psychopathology get through undetected, although this would result in higher false positives. So, for the purpose of screening, ROC-based cut-offs, 75th percentile scores from the ROM reference group, or 5th percentile scores from the ROM patient group may be appropriate; for the SF-36 this would be represented by the 25th and 95th percentiles, respectively [26]. However, if the consequences of missing the disease are relatively minor, and if the costs of therapy providing for subjects who are wrongfully diagnosed are substantial, a somewhat higher specificity with lower sensitivity may be used [39]. The reference values established in the present study can be used to determine whether a patient's level of symptoms falls within the normal range of values after treatment (e.g., whether a treated patient is no longer any different from normal controls with respect to the level of depressive symptoms). These reference values are to be used to determine treatment goals.

Normality can be defined statistically or medically. The statistical model is based on the distribution of scores from the general population (including all individuals) and on deviation from the mean. The middle range of scores of the normal distribution is considered as normal (within 2 SD of the mean), and extreme high or low scores are considered deviant. The medical model considers psychopathology and normality (i.e. absence of psychopathology) in absolute terms. It excludes individuals with a disorder from a reference group [40]. In our study we chose the statistical approach and therefore included all non-consulting individuals, both with and without (sub clinical) symptoms. So, there are different viewpoints as to whether the general population should consist of non-treated subjects or whether it should be more restricted (i.e., only including subjects without psychiatric diagnoses). We have chosen for the former definition, because we tested generic instruments which are not confined to a single DSM-IV diagnosis. If we had excluded 122 (9.5%) subjects with a MINI-diagnosis from the main analysis, we think that the reference values would have been too strict. Nevertheless, we have already shown above that the reference values were not affected to any large extent by our inclusive methodology.

The present study has several strengths. The ROM reference group was sufficiently large, clearly defined, and similar to the ROM patient group with respect to age, gender, and level of urbanization. These non-consulting GP patients were highly representative of the general population, given the extremely high GP registration percentage. This was further illustrated by the fact that sufficient psychiatric symptoms were reported by approximately 10% of the population-based reference group to the point of warranting a DSM-IV diagnosis, which is in line with a Dutch (NEMESIS) comorbidity study [41]. Stratification of the ROM reference group into more homogeneous gender- and age-subsets resulted in a better differentiation of reference values. Assessment and analytical procedures were standardized and of high quality, similar to the ones used for the ROM patients. Limitations of our study that should be mentioned include the high non-response (63.2%) in the ROM reference group, which may have resulted in bias due to selection. Some populations (i.e., younger males with full-time employment) may have been underrepresented. We believe that this may

have resulted in a slight under-representation of the healthiest subjects, overly conservative estimates of the discriminative power of the instruments, slightly low percentile scores, and slightly high cut-off points for the transition from healthy to psychiatrically sick. At the same time, analyses of data from the ROM reference group without the 10.9% of subjects with a MINI diagnosis did not substantially alter our findings, suggesting that our reference values were fairly robust. As no information was available for non-responders and excluded individuals, they could not be compared with the ROM reference group for demographic variables. Furthermore, ethnic and cultural differences were not considered. Therefore, our reference values for the Dutch general population may not directly apply to other ethnic or cultural groups. Likewise, reference values for children and the elderly remain to be assessed. Another issue concerns the use of the DAPP-SF for the assessment of dysfunctional personality traits. It has been suggested that the limited validity of self-report instruments for assessing personality pathology is particularly relevant in clinical populations [42], especially among depressed [43] and psychotic patients [44]. Finally, it is important to recognize the limitations of population-based reference values. They should not be interpreted too rigidly.

CONCLUSION

This large-scale population-based study provides reference values for the BSI, MASQ-D30, SF-36, and DAPP-SF. These reference values are essential for use in clinical psychiatry care. The scales are commonly incorporated in the comprehensive set of generic ROM instruments and they can be administered with every patient with psychiatric disorders for the purpose of routine screening, referral, and treatment. This set of four scales thoroughly covers general psychopathology, mood- and anxiety disorders (which represent 80% of psychiatric disorders), personality disorders, and quality of life. ROM reference values inform therapists and patients on the severity of the complaints at intake, and the waxing and waning of symptoms over the course of treatment. Furthermore, they enable research of the effectiveness of treatments in everyday clinical practice and managers can use them for benchmarking.

Reference List

1. De Beurs E, Den Hollander-Gijsman ME, Van Rood YR, Van der Wee NJ, Giltay EJ, Van Noorden MS, Van der Lem R, Van Fenema EM, Zitman FG. (2011) Routine outcome monitoring in the Netherlands: practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clin Psychol Psychother,* 18, 1-12.

2. Solberg HE. (1989) Reference values. *Adv Clin Chem*, 27, 1-79.

3. Solberg HE. (2008) Establishment and use of reference values. Burtis CA, Ashwood ER, Bruns DE, editors. Fundamentals of clinical chemistry. 6[14], 229-238. St. Louis, Missouri, Saunders Elsevier.

4. Geffre A, Friedrichs K, Harr K, Concordet D, Trumel C, Braun JP. (2009) Reference values: a review. *Vet Clin Pathol*, 38 (3), 288-298.

5. Katayev A, Balciza C, Seccombe DW. (2010) Establishing reference intervals for clinical laboratory test results: is there a better way? Am *J Clin Pathol*, 133 (2), 180-186.

6. Sasse EA, Doumas BT, Miller WG, D'Orazio P, Eckfeldt JH, Evans SA, et al. (2000) How to define and determine reference intervals in the clinical laboratory; approved guideline-Second edition. NCCLS document C28-A2 . 20[13], 1-38. Wayne, PA, NCCLS.

7. Kendall PC, Marrs-Garcia A, Nath SR, Sheldrick RC. (1999) Normative comparisons for the evaluation of clinical significance. *J Consult Clin Psychol*, 67 (3), 285-299.

8. Horn PS, Feng L, Li Y, Pesce AJ. (2001) Effect of outliers and nonhealthy individuals on reference interval estimation. *Clin Chem*, 47 (12), 2137-2145.

9. Reed AH, Henry RJ, Mason WB. (1971) Influence of statistical method used on the resulting estimate of normal range. *Clin Chem*, 17 (4), 275-284.

10. Fan J, Upadhye S, Worster A. (2006) Understanding receiver operating characteristic (ROC) curves. *CJEM*, 8 (1), 19-20.

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

12. Derogatis LR. (1975) The Brief Symptom Inventory. Baltimore, MD., Clinical Psychometric Research.

13. Clark LA, Watson D. (1991) Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *J Abnorm Psychol*, 100 (3), 316-336.

14. Wardenaar KJ, Van Veen T, Giltay EJ, De Beurs E, Penninx BW, Zitman FG. (2010) Development and validation of a 30-item short adaptation of the Mood and Anxiety Symptoms Questionnaire (MASQ). *Psychiatry Res*, 179 (1), 101-106.

15. Aaronson NK, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, Te Velde A, Verrips E. (1998) Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol*, 51 (11), 1055-1068.

16. Ware JE, Snow KK, Kosinski M, Gandek B. (1993) SF-36 Health Survey Manual and Interpretation Guide. Boston, New England Medical Center, The Health Institute.

17. Livesley WJ, Jackson DN. (2002) Manual for the dimensional assessment of personality pathology - basic questionnaire (DAPP-BQ). Port Huron, Sigma Press.

18. Van Kampen D, De Beurs E, Andrea H. (2008) A short form of the Dimensional Assessment of Personality Pathology-Basic Questionnaire (DAPP-BQ): the DAPP-SF. *Psychiatry Res*, 160 (1), 115-128.

19. Derogatis LR, Melisaratos N. (1983) The Brief Symptom Inventory: an introductory report. *Psychol Med*, 13 (3), 595-605.

20. Jenkinson C, Coulter A, Wright L. (1993) Short form 36 (SF36) health survey questionnaire: normative data for adults of working age. *BMJ,* 306 (6890), 1437-1440.

21. Loge JH, Kaasa S. (1998) Short form 36 (SF-36) health survey: normative data from the general Norwegian population. *Scand J Soc Med*, 26 (4), 250-258.

22. Ware JE, Gandek B. (1998) Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. *J Clin Epidemiol*, 51 (11), 903-912.

23. De Beurs E, Rinne T, Van Kampen D, Verheul R, Andrea H. (2009) Reliability and validity of the Dutch Dimensional Assessment of Personality Pathology-Short Form (DAPP-SF), a shortened version of the DAPP-Basic Questionnaire. *J Pers Disord*, 23 (3), 308-326.

24. Tiemensma J, Biermasz NR, Middelkoop HA, Van der Mast RC, Romijn JA, Pereira AM. (2010) Increased prevalence of psychopathology and maladaptive personality traits after long-term cure of Cushing's disease. *J Clin Endocrinol Metab*, 95 (10), E129-E141.

25. Bowling A, Bond M, Jenkinson C, Lamping DL. (1999) Short Form 36 (SF-36) Health Survey questionnaire: which normative data should be used? Comparisons between the norms provided by the Omnibus Survey in Britain, the Health Survey for England and the Oxford Healthy Life Survey. J Public Health Med, 21 (3), 255-270.

26. Schulte-van Maaren YWM, Carlier IVE, Giltay EJ, Van Noorden MS, De Waal MW, Van der Wee NJ, Zitman FG. (2012) Reference values for mental health assessment instruments: objectives and methods of the Leiden Routine Outcome Monitoring Study. J Eval Clin Pract.

27. Van Noorden MS, Giltay EJ, Den Hollander-Gijsman ME, Van der Wee NJ, Van Veen T, Zitman FG. (2010) Gender differences in clinical characteristics in a naturalistic sample of depressive outpatients: the Leiden Routine Outcome Monitoring Study. J Affect Disord, 125 (1-3), 116-123.

28. Poortvliet MC, Lamkadden M, Deville W. Niet op naam ingeschreven (NONI) bij de huisarts. Inventarisatie en gevolgen voor de ziekenfondsverzekerden. Utrecht: NIVEL; 2005.

29. Carlier IVE, Meuldijk D, Van Vliet IM, Van Fenema EM, Van der Wee NJ, Zitman FG. (2012) Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. J Eval Clin Pract, 18 (1), 104-110.

30. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC. (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry, 59 Suppl 20, 22-33

31. Van Vliet IM, De Beurs E. (2007) The MINI-International Neuropsychiatric Interview. A brief structured diagnostic psychiatric interview for DSM-IV en ICD-10 psychiatric disorders. *Tijdschr Psychiatr*, 49 (6), 393-397.

32. Van Vliet IM, Leroy H, Van Megen HJGM. (2000) M.I.N.I. Internationaal Neuropsychiatrisch Interview. Nederlandse Versie 5.0.0.

33. Barnabei L, Marazia S, De CR. (2007) Receiver operating characteristic (ROC) curves and the definition of threshold levels to diagnose coronary artery disease on electrocardiographic stress testing. Part I: The use of ROC curves in diagnostic medicine and electrocardiographic markers of ischaemia. *J Cardiovasc Med (Hagerstown)*, 8 (11), 873-881.

34. Gandek B, Ware JE. (1998) Methods for validating and norming translations of health status questionnaires: the IQOLA Project approach. International Quality of Life Assessment. *J Clin Epidemiol*, 51 (11), 953-959.

35. Apolone G, Mosconi P. (1998) The Italian SF-36 Health Survey: translation, validation and norming. *J Clin Epidemiol*, 51 (11), 1025-1036.

36. Hammerlid E, Taft C. (2001) Health-related quality of life in long-term head and neck cancer survivors: a comparison with general population norms. *Br J Cancer,* 84 (2), 149-156.

37. Razavi D, Gandek B. (1998) Testing Dutch and French translations of the SF-36 Health Survey among Belgian angina patients. *J Clin Epidemiol*, 51 (11), 975-981

38. Demiral Y, Ergor G, Unal B, Semin S, Akvardar Y, Kivircik B, Alptekin K. (2006) Normative data and discriminative properties of short form 36 (SF-36) in Turkish urban population. *BMC Public Health*, 6, 247.

39. Marazia S, Barnabei L, De Caterina R. (2008) Receiver operating characteristic (ROC) curves and the definition of threshold levels to diagnose coronary artery disease on electrocardiographic stress testing. Part II: the use of ROC curves in the choice of electrocardiographic stress test markers of ischaemia. *J Cardiovasc Med (Hagerstown),* 9 (1), 22-31.

40. Zimmerman M, Chelminski I, Posternak M. (2004) A review of studies of the Montgomery-Asberg Depression Rating Scale in controls: implications for the definition of remission in treatment studies of depression. *Int Clin Psychopharmacol*, 19 (1), 1-7.

41. Bijl RV, Ravelli A, Van Zessen G. (1998) Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc Psychiatry Psychiatr Epidemiol*, 33 (12), 587-595.

42. Tromp NB, Koot HM. (2010) Self- and parent report of adolescent personality pathology: informant agreement and relations to dysfunction. *J Pers Disord*, 24 (2), 151-170.

43. Hirschfeld RM, Klerman GL, Clayton PJ, Keller MB, McDonald-Scott P, Larkin BH. (1983) Assessing personality: effects of the depressive state on trait measurement. *Am J Psychiatry*, 140 (6), 695-699.

44. Zimmerman M, Coryell WH. (1990) Diagnosing personality disorders in the community. A comparison of self-report and interview measures. *Arch Gen Psychiatry*, 47 (6), 527-531.

Supplementary Table 3.1: Percentile scores and mean values in the ROM reference (n=1294) and patient (n=5269) groups for the subscales and total score of the Brief Symptom Inventory (BSI).

	ROM reference group (n=1294)						ROM patient group (n=5269)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Somatization (SOM)									n=5269			
All participants	0.00	0.00	0.00	0.29	0.71	0.17 ± 0.28	0.00	0.43	0.86	1.43	2.71	1.03 ± 0.83
- Women aged 18-40 yr	0.00	0.00	0.14	0.29	0.86	0.20 ± 0.31	0.00	0.43	0.86	1.57	2.71	1.06 ± 0.85
- Women aged 41-65 yr	0.00	0.00	0.14	0.29	0.86	0.20 ± 0.29	0.14	0.43	0.86	1.57	2.71	1.09 ± 0.81
- Men aged 18-40 yr	0.00	0.00	0.00	0.14	0.44	0.11 ± 0.19	0.00	0.29	0.71	1.43	2.57	0.94 ± 0.79
- Men aged 41-65 yr	0.00	0.00	0.00	0.14	0.57	0.11 ± 0.23	0.00	0.29	0.71	1.43	2.68	0.96 ± 0.82
Obsessive-Compulsive (O-C)												
All participants	0.00	0.00	0.17	0.50	1.17	0.35 ± 0.42	0.33	1.00	1.67	2.33	3.33	1.67 ± 0.95
- Women aged 18-40 yr	0.00	0.00	0.33	0.50	1.17	0.38 ± 0.44	0.33	1.00	1.67	2.33	3.33	1.68 ± 0.95
- Women aged 41-65 yr	0.00	0.00	0.17	0.50	1.25	0.37 ± 0.45	0.27	0.83	1.67	2.33	3.17	1.64 ± 0.92
- Men aged 18-40 yr	0.00	0.00	0.17	0.50	1.17	0.32 ± 0.37	0.17	0.83	1.50	2.33	3.33	1.64 ± 0.95
- Men aged 41-65 yr	0.00	0.00	0.17	0.50	1.00	0.29 ± 0.37	0.22	1.00	1.67	2.50	3.50	1.73 ± 0.98
Interpersonal Sensitivity (I-S)												
All participants	0.00	0.00	0.00	0.50	1.00	0.29 ± 0.42	0.00	0.75	1.50	2.25	3.50	1.56 ± 1.04
- Women aged 18-40 yr	0.00	0.00	0.25	0.50	1.25	0.35 ± 0.46	0.25	0.75	1.63	2.50	3.75	1.74 ± 1.09
- Women aged 41-65 yr	0.00	0.00	0.25	0.50	1.25	0.31 ± 0.44	0.00	0.69	1.25	2.00	3.50	1.45 ± 1.01
- Men aged 41-65 yr	0.00	0.00	0.00	0.25	0.75	0.17 ± 0.34	0.00	0.50	1.25	2.00	3.25	1.35 ± 0.97
Depression (DEP)												
All participants	0.00	0.00	0.00	0.33	0.83	0.20 ± 0.34	0.17	0.83	1.67	2.50	3.50	1.68 ± 1.01

Supplementary Table 3.1: continued.

- Women aged 18-40 yr	0.00	0.00	0.00	0.33	0.83	0.22 ± 0.34	0.17	0.83	1.67	2.50	3.50	1.71 ± 1.05
- Women aged 41-65 yr	0.00	0.00	0.00	0.33	0.83	0.22 ± 0.37	0.17	0.83	1.50	2.50	3.33	1.66 ± 1.01
- Men aged 18-40 yr	0.00	0.00	0.00	0.17	0.68	0.17 ± 0.30	0.17	0.83	1.67	2.33	3.33	1.66 ± 0.96
- Men aged 41-65 yr	0.00	0.00	0.00	0.17	0.67	0.14 ± 0.31	0.17	0.83	1.50	2.42	3.50	1.68 ± 0.99
Anxiety (ANX)												
All participants	0.00	0.00	0.17	0.33	0.83	0.22 ± 0.34	0.17	0.83	1.33	2.17	3.33	1.49 ± 0.94
- Women aged 18-40 yr	0.00	0.00	0.17	0.33	1.00	0.27 ± 0.39	0.17	0.83	1.33	2.17	3.33	1.53 ± 0.95
- Women aged 41-65 yr	0.00	0.00	0.17	0.33	0.92	0.23 ± 0.36	0.17	0.83	1.33	2.17	3.33	1.49 ± 0.94
- Men aged 18-40 yr	0.00	0.00	0.17	0.33	0.67	0.19 ± 0.25	0.17	0.67	1.33	2.00	3.17	1.42 ± 0.92
- Men aged 41-65 yr	0.00	0.00	0.00	0.17	0.67	0.17 ± 0.29	0.17	0.67	1.33	2.00	3.33	1.49 ± 0.94
Hostility (HOS)												
All participants	0.00	0.00	0.20	0.20	0.80	0.20 ± 0.29	0.00	0.20	0.80	1.40	2.80	0.94 ± 0.86
- Women aged 18-40 yr	0.00	0.00	0.20	0.20	1.00	0.25 ± 0.36	0.00	0.40	0.80	1.60	3.00	1.07 ± 0.90
- Women aged 41-65 yr	0.00	0.00	0.20	0.20	0.60	0.18 ± 0.24	0.00	0.20	0.60	1.00	2.20	0.73 ± 0.73
- Men aged 18-40 yr	0.00	0.00	0.20	0.20	0.60	0.17 ± 0.24	0.00	0.40	0.80	1.40	2.80	0.97 ± 0.86
- Men aged 41-65 yr	0.00	0.00	0.20	0.20	0.60	0.17 ± 0.23	0.00	0.20	0.60	1.20	2.60	0.90 ± 0.84
Phobic Anxiety (PHOB)												
All participants	0.00	0.00	0.00	0.20	0.60	0.11 ± 0.23	0.00	0.40	1.00	1.60	3.00	1.15 ± 0.93
- Women aged 18-40 yr	0.00	0.00	0.00	0.20	0.60	0.11 ± 0.23	0.00	0.40	1.00	1.80	3.20	1.19 ± 0.96
- Women aged 41-65 yr	0.00	0.00	0.00	0.20	0.80	0.13 ± 0.26	0.00	0.40	0.80	1.60	3.00	1.12 ± 0.94
- Men aged 18-40 yr	0.00	0.00	0.00	0.00	0.40	0.07 ± 0.16	0.00	0.40	1.00	1.65	3.00	1.15 ± 0.91
- Men aged 41-65 yr	0.00	0.00	0.00	0.00	0.60	0.10 ± 0.23	0.00	0.40	0.80	1.60	2.80	1.09 ± 0.8

Supplementary Table 3.1: continued.

Paranoid Ideation (PAR)												
All participants	0.00	0.00	0.00	0.40	0.80	0.23 ± 0.35	0.00	0.40	1.00	1.80	3.00	1.15 ± 0.94
- Women aged 18-40 yr	0.00	0.00	0.20	0.40	0.82	0.25 ± 0.38	0.00	0.40	1.00	1.80	3.00	1.20 ± 0.96
- Women aged 41-65 yr	0.00	0.00	0.20	0.40	0.80	0.23 ± 0.33	0.00	0.40	0.80	1.60	2.80	1.07 ± 0.90
- Men aged 18-40 yr	0.00	0.00	0.20	0.40	0.80	0.22 ± 0.31	0.00	0.40	1.00	1.80	3.00	1.14 ± 0.93
- Men aged 41-65 yr	0.00	0.00	0.00	0.20	0.80	0.19 ± 0.34	0.00	0.40	1.00	1.80	3.00	1.15 ± 0.94
Psychoticism (PSY)												
All participants	0.00	0.00	0.00	0.20	0.80	0.14 ± 0.28	0.20	0.60	1.20	1.80	2.80	1.23 ± 0.81
- Women aged 18-40 yr	0.00	0.00	0.00	0.20	0.80	0.16 ± 0.29	0.20	0.60	1.20	1.80	2.80	1.29 ± 0.85
- Women aged 41-65 yr	0.00	0.00	0.00	0.20	0.80	0.15 ± 0.28	0.00	0.60	1.00	1.60	2.60	1.13 ± 0.79
- Men aged 18-40 yr	0.00	0.00	0.00	0.20	0.80	0.14 ± 0.28	0.20	0.60	1.20	1.80	2.80	1.26 ± 0.79
- Men aged 41-65 yr	0.00	0.00	0.00	0.20	0.60	0.11 ± 0.25	0.20	0.60	1.00	1.60	2.60	1.19 ± 0.78
BSI total score												
All participants	0.00	0.06	0.13	0.28	0.68	0.21 ± 0.25	0.34	0.79	1.23	1.75	2.66	1.33 ± 0.71
- Women aged 18-40 yr	0.00	0.06	0.17	0.34	0.72	0.25 ± 0.27	0.34	0.83	1.30	1.85	2.74	1.38 ± 0.73
- Women aged 41-65 yr	0.00	0.06	0.15	0.30	0.78	0.23 ± 0.25	0.30	0.75	1.19	1.72	2.55	1.28 ± 0.69
- Men aged 18-40 yr	0.00	0.06	0.13	0.21	0.55	0.18 ± 0.20	0.34	0.79	1.21	1.72	2.58	1.29 ± 0.68
- Men aged 41-65 yr	0.00	0.02	0.09	0.23	0.51	0.16 ± 0.22	0.31	0.75	1.21	1.68	2.70	1.29 ± 0.72

Supplementary Table 3.2: Percentile scores and mean values in the ROM reference (n=1294) and patient (n=5269) groups for the subscales and total score of the Mood & Anxiety Symptom Questionnaire-30 (MASQ-D30).

	ROM reference group (n=1294)						ROM patient group (n=5269)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
General distress (GD)									n=5269			
All participants	10	11	12	15	23	13.8 ± 4.4	17	23	28	33	40	28.1 ± 6.9
- Women aged 18-40 yr	10	11	13	17	25	14.7 ± 5.0	17	24	29	34	40	28.6 ± 7.0
- Women aged 41-65 yr	10	11	12	15	23	13.7 ± 4.4	17	23	28	33	40	28.1 ± 6.9
- Men aged 18-40 yr	10	11	12	15	21	13.2 ± 3.7	16	23	27	32	39	27.3 ± 6.6
- Men aged 41-65 yr	10	10	12	14	20	12.8 ± 3.8	17	23	27	32	40	27.7 ± 6.8
Anhedonic depression (AD)												
All participants	10	14	17	22	29	18.4 ± 5.8	**17**	**24**	**31**	**37**	**44**	**30.7 ± 8.3**
- Women aged 18-40 yr	11	14	17	22	28	18.3 ± 5.5	17	24	30	37	44	30.4 ± 8.5
- Women aged 41-65 yr	11	15	18	23	32	19.4 ± 6.2	18	24	31	37	44	30.9 ± 8.4
- Men aged 18-40 yr	9	13	16	20	28	17.0 ± 5.4	18	25	31	37	44	30.9 ± 8.0
- Men aged 41-65 yr	11	14	17	22	29	18.6 ± 5.6	17	25	31	37	44	31.1 ± 8.2
Anxious arousal (AA)												
All participants	10	10	11	13	17	11.9 ± 3.0	18	26	31	37	43	31.3 ± 7.5
- Women aged 18-40 yr	10	10	11	13	18	12.3 ± 3.5	18	26	31	37	43	31.2 ± 7.6
- Women aged 41-65 yr	10	10	11	13	19	12.2 ± 3.1	18	26	31	37	44	31.1 ± 7.5
- Men aged 18-40 yr	10	10	10	12	15	11.3 ± 2.0	19	26	32	36	43	31.3 ± 7.3
- Men aged 41-65 yr	10	10	10	12	16	11.3 ± 2.2	18	27	32	37	44	31.6 ± 7.5
- Men aged 41-65 yr	0.00	0.00	0.00	0.25	0.75	0.17 ± 0.34	0.00	0.50	1.25	2.00	3.25	1.35 ± 0.97

Supplementary Table 3.3: Percentile scores and mean values in the ROM reference (n=1294) and patient (n=5269) groups for the subscales and total score of the Short Form 36 (SF-36).

	ROM reference group (n=1294)						ROM patient group (n=5269)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Physical Functioning									n=5269			
All participants	65	90	100	100	100	92.6 ± 14.2	25	60	80	95	100	74.8 ± 23.7
- Women aged 18-40 yr	70	95	100	100	100	93.8 ± 12.4	30	60	85	95	100	76.3 ± 23.1
- Women aged 41-65 yr	53	85	95	100	100	89.4 ± 17.0	18	50	73	90	100	67.8 ± 25.2
- Men aged 18-40 yr	80	95	100	100	100	96.9 ± 7.8	40	70	90	100	100	81.3 ± 20.5
- Men aged 41-65 yr	55	90	95	100	100	91.1 ± 15.9	25	55	80	90	100	73.1 ± 24.0
Role-physical												
All participants	13	100	100	100	100	87.0 ± 27.2	**0**	**0**	**25**	**75**	**100**	**37.2 ± 39.7**
- Women aged 18-40 yr	25	75	100	100	100	87.0 ± 26.4	0	0	25	75	100	38.8 ± 40.1
- Women aged 41-65 yr	0	75	100	100	100	84.4 ± 31.0	0	0	0	50	100	30.9 ± 38.7
- Men aged 18-40 yr	25	100	100	100	100	90.9 ± 22.6	0	0	25	75	100	44.2 ± 40.2
- Men aged 41-65 yr	25	70	100	100	100	87.0 ± 26.3	0	0	25	75	100	33.8 ± 38.0
Bodily Pain												
All participants	54	78	90	100	100	86.4 ± 17.6	20	45	67	90	100	65.9 ± 27.5
- Women aged 18-40 yr	54	78	90	100	100	84.2 ± 18.2	20	45	67	90	100	65.7 ± 26.9
- Women aged 41-65 yr	45	78	90	100	100	84.8 ± 19.2	10	45	57	90	100	61.3 ± 28.4
- Men aged 18-40 yr	67	90	100	100	100	91.4 ± 13.6	22	57	78	100	100	72.0 ± 26.0
- Men aged 41-65 yr	57	80	90	100	100	87.9 ± 16.2	20	45	67	90	100	65.7 ± 28.2
Social Functioning												
All participants	63	88	100	100	100	89.9 ± 15.6	0	25	50	63	88	44.8 ± 26.1
- Women aged 18-40 yr	50	75	100	100	100	88.2 ± 16.8	0	25	50	63	88	44.5 ± 26.2
- Women aged 41-65 yr	56	88	100	100	100	88.7 ± 16.1	0	25	38	63	88	42.4 ± 25.9

Supplementary Table 3.3: continued

- Men aged 18-40 yr	63	88	100	100	100	93.8 ± 12.3	0	25	50	63	88	48.3 ± 26.1
- Men aged 41-65 yr	63	88	100	100	100	90.9 ± 14.9	0	25	50	63	88	44.9 ± 25.6
Mental health												
All participants	56	72	80	88	96	79.7 ± 12.3	12	28	40	52	76	41.5 ± 18.2
- Women aged 18-40 yr	56	72	80	88	96	78.9 ± 12.2	12	28	40	52	76	42.0 ± 18.2
- Women aged 41-65 yr	52	72	80	88	94	78.2 ± 13.3	12	28	40	52	76	41.1 ± 18.8
- Men aged 18-40 yr	64	76	84	89	96	81.8 ± 10.8	16	32	40	52	72	41.9 ± 17.2
- Men aged 41-65 yr	60	76	84	88	96	81.5 ± 11.6	12	28	40	52	76	40.5 ± 18.5
Role-emotional												
All participants	33	100	100	100	100	90.4 ± 24.8	0	0	0	33	100	28.2 ± 36.2
- Women aged 18-40 yr	0	100	100	100	100	88.5 ± 26.9	0	0	0	50	100	28.4 ± 36.4
- Women aged 41-65 yr	0	100	100	100	100	88.4 ± 28.5	0	0	0	33	100	26.0 ± 36.3
- Men aged 18-40 yr	67	100	100	100	100	94.3 ± 16.5	0	0	33	67	100	30.3 ± 35.8
- Men aged 41-65 yr	33	100	100	100	100	92.9 ± 20.7	0	0	0	33	100	28.7 ± 36.1
Vitality												
All participants	40	60	70	80	90	68.6 ± 15.3	5	20	35	45	65	34.3 ± 17.8
- Women aged 18-40 yr	40	55	70	75	90	66.2 ± 15.0	5	20	35	45	65	34.1 ± 17.6
- Women aged 41-65 yr	35	60	70	80	90	68.7 ± 16.3	5	20	30	45	65	32.6 ± 18.1
- Men aged 18-40 yr	45	60	70	80	90	70.1 ± 14.0	10	25	35	50	70	36.9 ± 17.3
- Men aged 41-65 yr	50	60	70	85	95	71.3 ± 15.1	0	20	35	45	65	33.9 ± 17.8
General Health												
All participants	45	65	80	90	100	76.2 ± 16.3	20	35	50	65	90	51.6 ± 21.0
- Women aged 18-40 yr	45	68	80	90	100	77.0 ± 16.2	18	35	50	69	90	51.6 ± 21.4
- Women aged 41-65 yr	37	65	80	90	100	74.9 ± 17.9	20	35	50	65	85	51.2 ± 20.4
- Men aged 18-40 yr	55	70	80	90	100	78.8 ± 13.4	20	40	50	70	90	53.3 ± 20.8
- Men aged 41-65 yr	40	65	75	85	100	74.1 ± 16.0	15	35	50	65	85	49.9 ± 20.9

Supplementary Table 3.4: Percentile scores and mean values in the ROM reference (n=635) and patient (n=5035) groups for the subscales and total score of the Dimensional Assessment of Personality Pathology – short form (DAPP-SF).

	ROM reference group (n=1294)						ROM patient group (n=5269)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Submissiveness										n=5269		
All participants	1.13	1.50	2.00	2.50	3.50	2.10 ± 0.75	1.25	2.25	3.00	3.63	4.38	2.94 ± 0.94
- Women aged 18-40 yr	1.13	1.50	2.13	2.75	3.75	2.20 ± 0.82	1.38	2.50	3.13	3.75	4.50	3.09 ± 0.92
- Women aged 41-65 yr	1.00	1.63	2.00	2.63	3.38	2.10 ± 0.71	1.25	2.25	2.88	3.63	4.38	2.91 ± 0.96
- Men aged 18-40 yr	1.13	1.63	2.00	2.44	3.26	2.10 ± 0.69	1.38	2.25	2.88	3.50	4.38	2.84 ± 0.90
- Men aged 41-65 yr	1.00	1.25	1.75	2.25	3.38	1.89 ± 0.71	1.25	2.00	2.75	3.38	4.25	2.71 ± 0.93
Cognitive Distortion												
All participants	1.00	1.00	1.17	1.50	2.33	1.36 ± 0.51	1.00	1.50	2.33	3.00	4.17	2.36 ± 0.96
- Women aged 18-40 yr	1.00	1.00	1.17	1.50	2.33	1.38 ± 0.56	**1.00**	**1.67**	**2.33**	**3.17**	**4.17**	**2.40 ± 0.97**
- Women aged 41-65 yr	1.00	1.00	1.17	1.50	2.33	1.35 ± 0.45	1.00	1.50	2.17	2.83	4.00	2.23 ± 0.93
- Men aged 18-40 yr	1.00	1.00	1.17	1.50	2.68	1.36 ± 0.57	1.00	1.67	2.33	3.17	4.17	2.44 ± 0.96
- Men aged 41-65 yr	1.00	1.00	1.17	1.50	2.17	1.31 ± 0.41	1.00	1.50	2.33	3.00	4.00	2.34 ± 0.95
Identity Problems												
All participants	1.00	1.00	1.33	1.83	2.70	1.54 ± 0.59	1.33	2.33	3.17	3.83	4.67	3.12 ± 1.02
- Women aged 18-40 yr	1.00	1.17	1.50	1.83	2.68	1.61 ± 0.58	1.33	2.50	3.33	4.00	4.67	3.20 ± 1.00
- Women aged 41-65 yr	1.00	1.00	1.33	1.83	2.83	1.54 ± 0.61	1.17	2.17	3.17	3.83	4.67	3.00 ± 1.04
- Men aged 18-40 yr	1.00	1.00	1.33	1.83	3.02	1.55 ± 0.61	1.33	2.50	3.33	4.00	4.67	3.18 ± 1.00
- Men aged 41-65 yr	1.00	1.00	1.17	1.67	2.50	1.40 ± 0.52	1.17	2.33	3.17	3.83	4.50	3.02 ± 1.02
Affective Lability												
All participants	1.00	1.38	1.88	2.50	3.50	2.01 ± 0.76	1.63	2.63	3.38	3.88	4.63	3.24 ± 0.88
- Women aged 18-40 yr	1.00	1.56	2.00	2.75	3.64	2.16 ± 0.79	1.88	2.88	3.50	4.00	4.63	3.42 ± 0.85

Supplementary Table 3.4: continued.

- Women aged 41-65 yr	1.00	1.50	2.00	2.63	3.63	2.11 ± 0.79	1.50	2.50	3.25	3.88	4.50	3.16 ± 0.90
- Men aged 18-40 yr	1.00	1.25	1.63	2.13	3.25	1.78 ± 0.64	1.63	2.50	3.25	3.75	4.50	3.14 ± 0.87
- Men aged 41-65 yr	1.00	1.25	1.63	2.00	3.38	1.78 ± 0.69	1.57	2.47	3.13	3.63	4.38	3.06 ± 0.85
Stimulus Seeking												
All participants	1.10	1.38	1.88	2.38	3.38	1.99 ± 0.72	1.00	1.50	2.00	2.63	3.75	2.13 ± 0.81
- Women aged 18-40 yr	1.00	1.50	1.88	2.25	3.64	2.00 ± 0.75	1.00	1.50	2.00	2.50	3.63	2.10 ± 0.80
- Women aged 41-65 yr	1.00	1.25	1.63	2.13	2.75	1.72 ± 0.56	1.00	1.38	1.75	2.25	3.13	1.85 ± 0.65
- Men aged 18-40 yr	1.11	1.88	2.38	2.88	3.50	2.37 ± 0.71	1.13	1.88	2.50	3.13	4.04	2.51 ± 0.90
- Men aged 41-65 yr	1.13	1.38	2.00	2.50	3.38	2.02 ± 0.70	1.00	1.63	2.06	2.63	3.63	2.15 ± 0.74
Compulsivity												
All participants	1.38	2.00	2.50	3.13	4.00	2.58 ± 0.77	1.38	2.13	2.88	3.63	4.50	2.89 ± 0.94
- Women aged 18-40 yr	1.38	2.00	2.63	3.25	4.01	2.63 ± 0.83	1.43	2.25	2.88	3.63	4.63	2.94 ± 0.95
- Women aged 41-65 yr	1.25	2.00	2.56	3.16	4.00	2.59 ± 0.81	1.38	2.13	2.88	3.63	4.63	2.87 ± 0.98
- Men aged 18-40 yr	1.49	2.00	2.38	2.88	3.38	2.37 ± 0.59	1.25	2.13	2.75	3.38	4.29	2.75 ± 0.90
- Men aged 41-65 yr	1.50	2.13	2.63	3.13	4.00	2.66 ± 0.72	1.50	2.25	3.00	3.63	4.50	2.95 ± 0.91
Restricted Expression												
All participants	1.25	1.75	2.25	2.88	3.63	2.33 ± 0.75	1.75	2.63	3.25	3.88	4.63	3.23 ± 0.86
- Women aged 18-40 yr	1.13	1.63	2.13	2.88	3.64	2.26 ± 0.76	1.75	2.63	3.25	3.75	4.63	3.19 ± 0.85
- Women aged 41-65 yr	1.25	1.75	2.25	2.88	3.63	2.33 ± 0.75	1.63	2.50	3.25	3.75	4.50	3.14 ± 0.90
- Men aged 18-40 yr	1.24	1.75	2.25	2.88	3.64	2.34 ± 0.74	1.88	2.75	3.38	4.00	4.75	3.37 ± 0.85
- Men aged 41-65 yr	1.25	1.88	2.50	3.00	3.88	2.46 ± 0.74	1.88	2.75	3.38	3.88	4.50	3.30 ± 0.82
Callousness												
All participants	1.00	1.30	1.60	2.00	2.60	1.69 ± 0.50	1.00	1.30	1.70	2.10	2.90	1.77 ± 0.60
- Women aged 18-40 yr	1.00	1.30	1.60	2.00	2.70	1.67 ± 0.49	1.00	1.30	1.60	2.10	2.77	1.74 ± 0.56
- Women aged 41-65 yr	1.00	1.10	1.40	1.80	2.30	1.50 ± 0.43	1.00	1.10	1.40	1.80	2.50	1.51 ± 0.48
- Men aged 18-40 yr	1.19	1.70	2.00	2.20	2.71	1.97 ± 0.49	1.10	1.60	2.00	2.50	3.30	2.10 ± 0.66

Supplementary Table 3.4: continued.

- Men aged 41-65 yr	1.00	1.40	1.70	2.10	2.80	1.76 ± 0.51	1.00	1.40	1.80	2.20	2.90	1.86 ± 0.58
Oppositionality												
All participants	1.00	1.40	1.80	2.30	3.20	1.91 ± 0.65	1.40	2.20	2.80	3.50	4.30	2.83 ± 0.89
- Women aged 18-40 yr	1.10	1.40	1.80	2.30	3.20	1.96 ± 0.68	1.40	2.20	2.90	3.50	4.30	2.86 ± 0.87
- Women aged 41-65 yr	1.00	1.40	1.70	2.20	2.99	1.82 ± 0.59	1.20	1.90	2.60	3.30	4.20	2.62 ± 0.90
- Men aged 18-40 yr	1.09	1.50	1.90	2.55	3.31	2.05 ± 0.67	1.50	2.30	3.00	3.60	4.50	3.01 ± 0.89
- Men aged 41-65 yr	1.00	1.40	1.70	2.20	3.10	1.85 ± 0.62	1.40	2.20	2.80	3.50	4.30	2.82 ± 0.87
Intimacy Problems												
All participants	1.13	1.63	2.13	2.50	3.38	2.14 ± 0.67	1.13	1.75	2.38	2.88	4.00	2.42 ± 0.85
- Women aged 18-40 yr	1.24	1.63	2.00	2.50	3.26	2.10 ± 0.66	1.25	1.88	2.38	2.88	4.00	2.44 ± 0.83
- Women aged 41-65 yr	1.25	1.88	2.25	2.66	3.75	2.35 ± 0.71	1.25	2.00	2.50	3.25	4.25	2.60 ± 0.89
- Men aged 18-40 yr	1.13	1.63	2.00	2.38	3.25	2.03 ± 0.61	1.13	1.63	2.13	2.75	3.68	2.25 ± 0.78
- Men aged 41-65 yr	1.13	1.50	2.00	2.38	3.25	1.99 ± 0.61	1.13	1.69	2.13	2.75	3.88	2.29 ± 0.82
Rejection												
All participants	1.38	1.88	2.50	3.00	3.75	2.47 ± 0.76	1.13	1.63	2.25	2.88	3.75	2.31 ± 0.82
- Women aged 18-40 yr	1.13	1.75	2.38	3.06	3.88	2.46 ± 0.81	1.13	1.63	2.13	2.75	3.70	2.25 ± 0.79
- Women aged 41-65 yr	1.13	1.63	2.25	2.88	3.48	2.28 ± 0.73	1.00	1.50	2.00	2.50	3.38	2.04 ± 0.74
- Men aged 18-40 yr	1.61	2.25	2.75	3.13	3.89	2.71 ± 0.67	1.25	2.00	2.63	3.13	4.00	2.58 ± 0.83
- Men aged 41-65 yr	1.50	2.13	2.50	3.00	3.75	2.56 ± 0.71	1.25	1.88	2.50	3.13	3.88	2.53 ± 0.82
Anxiousness												
All participants	1.00	1.33	1.83	2.50	3.50	2.03 ± 0.81	1.67	2.67	3.50	4.00	4.83	3.37 ± 0.94
- Women aged 18-40 yr	1.00	1.50	2.17	2.83	3.83	2.22 ± 0.86	1.83	3.00	3.67	4.17	4.83	3.52 ± 0.90
- Women aged 41-65 yr	1.00	1.33	2.00	2.67	3.33	2.06 ± 0.78	1.50	2.50	3.33	4.00	4.67	3.26 ± 0.98
- Men aged 18-40 yr	1.00	1.33	1.67	2.25	3.50	1.89 ± 0.75	1.67	2.75	3.50	4.00	4.67	3.35 ± 0.91

Supplementary Table 3.4: continued.

Conduct Problems												
All participants	1.00	1.00	1.13	1.38	2.13	1.26 ± 0.37	1.00	1.00	1.25	1.63	2.63	1.43 ± 0.57
- Women aged 18-40 yr	1.00	1.00	1.00	1.25	1.88	1.18 ± 0.27	1.00	1.00	1.13	1.50	2.38	1.33 ± 0.48
- Women aged 41-65 yr	1.00	1.00	1.00	1.13	1.63	1.13 ± 0.23	1.00	1.00	1.00	1.25	1.88	1.20 ± 0.33
- Men aged 18-40 yr	1.00	1.13	1.38	1.75	2.50	1.53 ± 0.49	1.00	1.25	1.63	2.25	3.25	1.80 ± 0.73
- Men aged 41-65 yr	1.00	1.00	1.13	1.50	2.38	1.33 ± 0.42	1.00	1.13	1.38	1.88	2.75	1.57 ± 0.60
Suspiciousness												
All participants	1.00	1.00	1.13	1.50	2.15	1.32 ± 0.46	1.00	1.38	2.00	2.88	4.00	2.18 ± 0.99
- Women aged 18-40 yr	1.00	1.00	1.13	1.50	2.14	1.31 ± 0.45	1.00	1.38	2.13	3.00	4.13	2.27 ± 1.02
- Women aged 41-65 yr	1.00	1.00	1.13	1.25	2.11	1.25 ± 0.45	1.00	1.13	1.63	2.50	3.75	1.92 ± 0.91
- Men aged 18-40 yr	1.00	1.00	1.25	1.63	2.40	1.43 ± 0.50	1.00	1.50	2.25	3.00	4.13	2.33 ± 0.97
- Men aged 41-65 yr	1.00	1.00	1.13	1.50	2.13	1.32 ± 0.44	1.00	1.25	2.00	2.88	4.00	2.18 ± 0.99
Social Avoidance												
All participants	1.00	1.17	1.67	2.17	3.33	1.82 ± 0.73	1.17	2.17	3.00	3.83	4.67	2.98 ± 1.07
- Women aged 18-40 yr	1.00	1.33	1.83	2.33	3.50	1.88 ± 0.73	1.17	2.33	3.17	4.00	4.67	3.11 ± 1.06
- Women aged 41-65 yr	1.00	1.17	1.67	2.17	3.33	1.80 ± 0.73	1.00	1.83	2.83	3.67	4.50	2.78 ± 1.07
- Men aged 18-40 yr	1.00	1.17	1.67	2.17	3.50	1.84 ± 0.77	1.17	2.33	3.17	3.83	4.67	3.05 ± 1.06
- Men aged 41-65 yr	1.00	1.17	1.67	2.00	3.17	1.71 ± 0.70	1.17	2.00	2.83	3.67	4.50	2.84 ± 1.04
Narcissism												
All participants	1.00	1.63	2.13	2.63	3.50	2.18 ± 0.76	1.10	1.75	2.25	2.88	3.88	2.36 ± 0.83
- Women aged 18-40 yr	1.00	1.75	2.25	3.00	3.63	2.33 ± 0.79	1.25	1.88	2.50	3.00	3.88	2.47 ± 0.80
- Women aged 41-65 yr	1.00	1.38	1.88	2.38	3.13	1.92 ± 0.66	1.00	1.38	1.88	2.50	3.38	1.98 ± 0.73
- Men aged 18-40 yr	1.25	1.88	2.50	2.88	3.89	2.43 ± 0.76	1.25	2.00	2.63	3.25	4.13	2.63 ± 0.86
- Men aged 41-65 yr	1.13	1.50	2.00	2.50	3.50	2.07 ± 0.73	1.00	1.63	2.25	2.88	3.88	2.31 ± 0.82

Supplementary Table 3.4: continued.

Insecure Attachment												
All participants	1.00	1.17	1.50	2.17	3.33	1.74 ± 0.77	1.00	2.00	2.83	3.83	4.83	2.91 ± 1.13
- Women aged 18-40 yr	1.00	1.17	1.67	2.33	3.33	1.85 ± 0.79	1.17	2.17	3.17	4.00	4.83	3.08 ± 1.11
- Women aged 41-65 yr	1.00	1.17	1.50	2.17	3.64	1.76 ± 0.80	1.00	2.00	2.83	3.83	4.83	2.86 ± 1.17
- Men aged 18-40 yr	1.00	1.17	1.33	1.83	3.17	1.57 ± 0.65	1.00	1.83	2.67	3.50	4.67	2.68 ± 1.07
- Men aged 41-65 yr	1.00	1.17	1.50	2.00	3.50	1.70 ± 0.76	1.00	1.83	2.67	3.67	4.83	2.81 ± 1.13
Self-Harm												
All participants	1.00	1.00	1.00	1.00	1.50	1.07 ± 0.27	1.00	1.00	1.33	2.33	3.67	1.76 ± 0.96
- Women aged 18-40 yr	1.00	1.00	1.00	1.00	1.50	1.07 ± 0.27	1.00	1.00	1.33	2.33	4.00	1.78 ± 1.01
- Women aged 41-65 yr	1.00	1.00	1.00	1.00	1.67	1.09 ± 0.31	1.00	1.00	1.17	2.17	3.67	1.69 ± 0.92
- Men aged 18-40 yr	1.00	1.00	1.00	1.00	1.33	1.06 ± 0.26	1.00	1.00	1.50	2.33	3.67	1.79 ± 0.92
- Men aged 41-65 yr	1.00	1.00	1.00	1.00	1.50	1.05 ± 0.23	1.00	1.00	1.33	2.33	3.67	1.76 ± 0.93

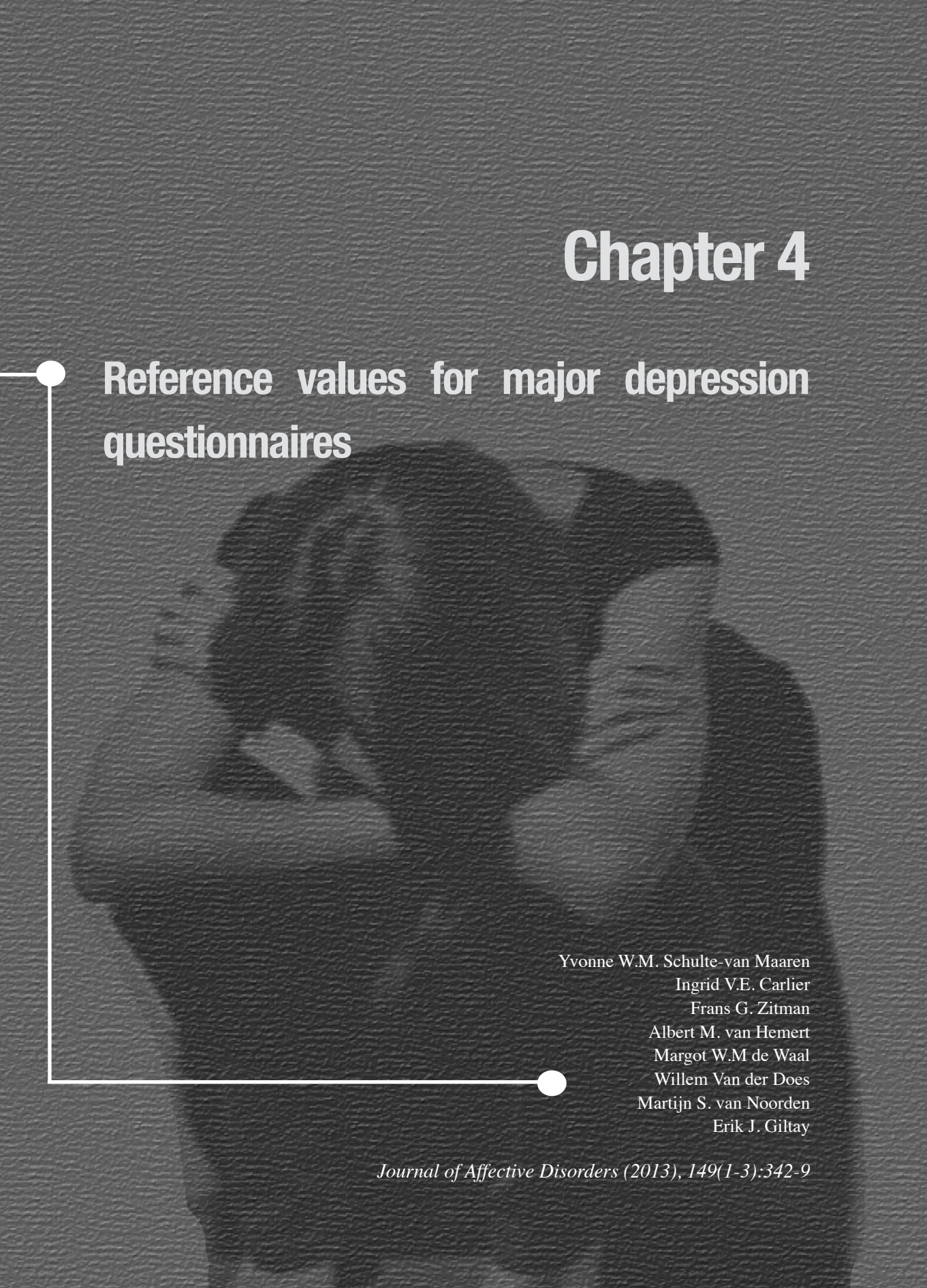
To calculate sum scores for the DAPP-SF subscales, multiply the mean scores by the number of items per subscale.

NormQuest

Reference Values for ROM Instruments and Questionnaires

Chapter 4

Reference values for major depression questionnaires



Yvonne W.M. Schulte-van Maaren
Ingrid V.E. Carlier
Frans G. Zitman
Albert M. van Hemert
Margot W.M de Waal
Willem Van der Does
Martijn S. van Noorden
Erik J. Giltay

Journal of Affective Disorders (2013), 149(1-3):342-9

ABSTRACT

Background: The Beck Depression Inventory-II (BDI-II), the Inventory of Depressive Symptoms (Self-Report) (IDS-SR), and the Montgomery-Åsberg Depression Rating Scale (MADRS) are questionnaires that assess symptom severity in patients with a depressive disorder. They are often incorporated in Routine Outcome Monitoring (ROM). We aimed to generate reference values for both 'healthy' and 'clinically depressed' populations to promote correct interpretation of ROM results.

Method: We included 1295 subjects from the general population (ROM reference-group) recruited through general practitioners, and 4627 psychiatric outpatients diagnosed with major depressive disorder (MDD) or dysthymic disorder (DD) (ROM patient-group). The outermost 5% of observations were used to define limits for one-sided reference intervals (95^{th} percentiles; P_{95}). Receiver Operating Characteristics (ROC) analyses were used to yield alternative cut-off values. Internal consistency reliability of the instruments was assessed.

Results: There was no significant difference between groups with respect to age and gender ratio. Mean age for the ROM reference-group was 40.3 years (SD=12.6) and for the ROM patient-group it was 39.3 years (SD=12.3). The proportion of females was 62.8% and 61.0% respectively. Cut-off values (P_{95}) were significantly different for women and men. Respectively, the cut-off values were 15 and 12 for the BDI-II, 23 and 18 for the IDS-SR, and 13 and 9 for the MADRS. ROC analyses yielded very similar reference values. The discriminative power of the BDI-II, IDS-SR, and MADRS scores was very high. Moreover, internal consistency was excellent for the total scores of all instruments. Internal consistency was satisfactory for all subscales with the exception of the IDS-SR subscale Atypical Characteristics.

Limitations: Non-response of 63% and limited generalizability (children, elderly, ethnic minorities).

Conclusion: For the BDI-II, IDS-SR, and MADRS a comprehensive set of reference values were provided. Reference values in the general population were higher in women than in men, suggesting the need to use gender-specific cut-off values. Each instrument can be offered to patients with mood-, anxiety or somatoform disorders to facilitate responsible decision-making with respect to continuing, changing or terminating therapy.

INTRODUCTION

Routine Outcome Monitoring (ROM) is the periodically repeated assessment of the condition of patients using diagnostic instruments and severity scales. It may thus aid in the evaluation of treatment outcome. Both generic and disorder-specific measurement instruments are used. Generic instruments are completed by all patients. They assess a broad range of psychopathological symptoms irrespective of the psychiatric disorder(s) experienced by patients. Disorder-specific instruments are administered to patients who meet the criteria for a particular disorder [1-3].

Reliable ratings from reference populations are essential for the correct interpretation of ROM results when making clinical decisions about continuing, altering, or terminating treatment [4]. Furthermore, reliable reference values can facilitate referral from specialized mental health care back to primary care. When establishing and interpreting reference values, several issues need to be considered. First, reference values [5] are often established in healthy populations [6] with health clearly defined by a priori inclusion and exclusion criteria [7-9]. As a consequence 'supernormal' (i.e., overly healthy) participants are sometimes selected [10], resulting in unreasonable reference intervals which are often 10% narrower [11]. Second, (sub)sample sizes of at least 120 are needed to reduce the amount of uncertainty and error caused by potential outliers [12]. Third, when data tend toward a non-Gaussian distribution, non-parametric percentile scores are more appropriate reference values than parametric mean values (and standard deviation (SD) of confidence interval (CI) values) [6,12]. For non-Gaussian distributions, weighted cut-off scores calculated by the Jacobson & Truax method [13] are equally unsuitable. In the case of non-Gaussian distributions, the 95th percentile (P_{95}) commonly serves as the reference value [6]. Finally, Receiver Operating Characteristics (ROC) analyses can provide cut-offs when both reference and patient data are available, reflecting the optimal trade-off between sensitivity and specificity [14].

The self-report Beck Depression Inventory-II (BDI-II; [15-17], the Inventory of Depressive Symptoms - Self-Report (IDS-SR) [18,19], and the observer-rated Montgomery-Åsberg Depression Rating Scale (MADRS; [20] are three frequently used ROM instruments that assess symptom severity of major depressive disorder (MDD). The BDI-II, unlike the IDS-SR and MADRS, can also be used as a diagnostic screening instrument for MDD [18]. Previous BDI-II studies reported cut-off and reference values for MDD outpatients [21,22] and inpatients [23,24]. A study in 376 undergraduates (17-29 years of age) and older adults (55-90 years of age) reported a mean total score of 8.6 (SD=7.7) [25]. IDS-SR reference values have been reported for depressed outpatients [26-29] and inpatients [30]. Based on 23 normal controls, a mean of 2.1 (SD=2.2) was reported. Based on 118 normal controls, a cut-off value of ≥ 18 was recommended [19]. on 118 normal controls, a cut-off value of ≥ 18 was recommended [19]. Many studies have reported means with SDs or cut-off values for the MADRS but these studies were conducted with outpatients with MDD [31-34], inpatients

with MDD [32,35,36], stroke patients [37], and old age pensioners [38]. In a review of studies of healthy controls (total n=569), Zimmerman et al. [39] reported means, SDs, and optimal cut-off scores. However, because of the strongly positively skewed distributions of all these total scores in healthy populations, the assumption of a normal distribution does not seem to be satisfied. Preferably, reference values would be based on a distribution-free percentile or ROC methodology.

The aim of this study was to establish reference values for the BDI-II, the IDS-SR, and the MADRS. Percentiles and ROC-based cut-off points were calculated, together with the more commonly reported means with SDs. A sample from the general population was recruited through general practitioners (GPs). These subjects were compared with a sample of outpatients diagnosed with MDD or DD (with or without other psychiatric disorders). Thus, we focused on a well-defined psychiatric patient group, we included a reference-group which was healthy but not necessarily symptom-free, and both samples were large in size.

METHODS

Participants

Our analyses of reference values were based on two study samples: a ROM reference sample from the general population and a ROM sample of psychiatric outpatients.

A total of 1295 participants (62.8% females) aged 18 to 65 years (M=40.3 years; SD=12.6) were included in the ROM reference-group as part of the 'Leiden Routine Outcome Monitoring Study' [2,3,40]. They were randomly selected from the registration systems of eight general practitioners (GPs) in the region of Leiden, the Netherlands. The response rate was 37.1%, as described elsewhere [2,3]. Inclusion criteria and exclusion criteria (e.g., treatment for psychiatric disorders and/or dependence on alcohol or drugs within six months prior to assessment) are described in detail elsewhere [2,3,40]. To make the group demographically comparable to the ROM patient-group the reference-group was matched for gender, age, and urbanization-level (62.3% urban).

The ROM patient-group consisted of a baseline sample of 4627 psychiatric outpatients, (61.0% females) aged between 18 and 65 years (M=39.3 years, SD=12.3). These outpatients were diagnosed with and treated for depressive disorders (MDD or dysthymic disorder, DD) in the Leiden University Medical Center (LUMC) Department of Psychiatry or the Rivierduinen specialized mental healthcare centres. Baseline assessment was part of the usual ROM procedure. About 80% of the referred patients with a tentative diagnosis of mood-, anxiety- and/or somatoform (MAS) disorder were assessed with ROM during the study period 2004-2009 [40].

Procedures and instruments

Procedures for the web-based ROM program of the LUMC Department of Psychiatry are described elsewhere [40,41]. All patients gave permission for the use of their ROM data for scientific purposes (written informed consent for this study was not required). In addition, participants of the ROM reference-group (non-patients) signed informed consent for the purpose of this study. For our study we used baseline data of ROM assessments. This included a standardized diagnostic interview (Dutch version of the Mini-International Neuropsychiatric Interview Plus, version 5.00-R: MINI-Plus; [42,43], the collection of sociodemographic and socioeconomic data, observer-rated scales, and self-report instruments. The BDI-II [16,17] was completed by 455 reference subjects and 4019 patients. The IDS-SR was completed by another group of 769 reference subjects and 474 patients. The MADRS was completed by the majority of the reference group (n=1291)and by all patients (n=4627). Halfway through the study we replaced the BDI-II with the IDS-SR. The IDS-SR is also a depression severity scale and is license-free. The MADRS is license-free as well.

The BDI-II, a revised version of the BDI [15], measures the severity of self-reported depression in adolescents and adults according to the criteria for diagnosing MDD as presented in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [44]. The BDI-II total score is derived by summing the scores for each of the 21 items. Each item is rated on a 4-point scale ranging from 0-3, with higher scores indicating higher levels of depression. The total score ranges from 0-63, where scores between 0-13 denote "minimal" depression, scores between 14-19 denote "mild" depression, scores between 20-28 denote "moderate" depression, and scores between 29-63 denote "severe" depression. In the international literature, two subscales [22,45,46], three subscales [17], and zero subscales [25] have been identified. The Cognitive and Somatic-Affective subscales are most commonly reported [22,23]. Respectively, they consist of 8 and 13 items, and subscale total scores range between 0-24 and between 0-39 [16]. The time-frame for the BDI-II is "the past two weeks, including today".

The IDS-SR self-report instrument [18,19] is designed to measure overall depressive symptom severity. The IDS-SR consists of 30 items, 23 of which cover the DSM-IV diagnostic criteria for MDD (including its atypical and melancholic subtypes). Seven items are not related to diagnostic criteria but to symptoms commonly associated with MDD (e.g. irritable mood, anxious mood). The items are rated on a 4-point Likert scale, with scores ranging from 0 (no symptomatology) to 3 (most severe). The total score ranges between 0-84, because only 28 of 30 items are scored (either decreased or increased weight are scored, and either decreased or increased appetite are scored). Scores between 0-13 denote "normal/no" depression, scores between 14-21 denote "possible/mild" depression, scores between 22-30 denote "moderate" depression, scores between 31-38 denote "severe" depression, and scores of 39 or higher denote "very severe" depression [47]. Frequently used subscales are the 10-item Atypical Characteristics subscale [48] and the 11-item Melancholic Characteristics subscale [49]. The time frame for the IDS-SR pertains to the previous 7 days, except in the

case of weight change which is rated for the previous 14 days.

The MADRS [20] is a clinician-rated instrument assessing the range and severity of depressive symptoms. The 10 items were designed to be particularly sensitive to treatment effects. The symptoms occur in the majority of cases although they do not cover all 9 DSM-IV MDD criteria. Rather, the items emphasize psychological symptoms such as apparent sadness and concentration problems [50]. Items are rated on a 7-point Likert scale anchored at 4 points (0: symptom is absent; 6: symptom is totally dominant) and summed to yield a total score between 0 and 60. Total scores between 0-8 denote "normal/no" depression, scores between 9-18 denote "possible/mild" depression, scores between 19-26 denote "moderate" depression, scores between 27-34 denote "severe" depression, and scores of 35 or higher denote "very severe" depression [35]. The time frame for the MADRS is for the previous seven days.

The MINI-Plus [42,43] was used to establish the presence of Axis I diagnosis according to the DSM-IV.

Statistical analyses

Analyses resulting in percentile scores and means (SDs) were conducted for the two groups separately, while ROC and internal consistency reliability analyses were conducted using data from both groups combined. In both groups, subjects who had 1 or more missing values per subscale were excluded. This permitted a robust evaluation of the use of the instruments [3]. Sociodemographic and psychopathological variables were descriptively analyzed (percentages in the case of categorical variables, means and SDs for the continuous variables). Cut-off scores indicating an optimal discrimination threshold between 'healthy' and 'diseased' were obtained by ROC analyses. Sensitivity and specificity were chosen to be equal, taking into account the trade-off between the two [14]. Although the scales are not diagnostic instruments, we assessed the discriminatory power of the instrument total scales and subscales, using the associated areas under the ROC curve (AUCs). AUCs over 0.75 were considered clinically useful, with values above 0.85 showing moderate discriminatory power and values above 0.95 showing very high discriminatory power [51]. The 5^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th} percentiles were calculated. In reference groups, the central 95% of the distribution is commonly used in the case of non- Gaussian distributions [5,7]. The remaining 5% is commonly categorized as 'abnormal' [52]. We regarded the top 5% of the reference-group (95^{th} percentiles, P_{95}) as 'abnormal' because the lowest 2.5% (i.e., functioning 'abnormally' good) is not identifiable in general population samples. That is, the BDI-II, IDS-SR, and MADRS merely assess the level of dysfunctionality and not the level of 'health' or normal functionality. Likewise, the bottom 5% of the patient-group (5^{th} percentiles, P_5) can be considered as indistinguishable from people in the normal range. Furthermore, means and SDs were calculated. Reference values were calculated for the entire reference-group and the entire patient-group, as well as for 4 strata in each group: young women (aged 18-40 years), older women (aged 41-65 years), young men (aged 18-40 years), and older men (aged 41-65

(aged 41-65 years). The internal consistency reliability of the instruments was tested using Cronbach's alpha for the total scale and for subscales when present. To test our decision not to exclude those individuals in the ROM reference-group with a current psychiatric diagnosis, we performed a sensitivity analysis. For all analyses, SPSS version 17.0 was used.

RESULTS

Statistical analyses

The sociodemographic and psychopathological characteristics of the ROM reference-group and ROM patient-group are shown in Table 4.1.

The ROM reference-group and the ROM patient-group were comparable with respect to age (M=40.3 years [SD=12.6] and M=39.3 [SD=12.3] respectively, p=0.14) and gender distribution (62.8% females and 61.0% females, respectively, p=0.25). Participants from the ROM reference-group were more often married than those from the ROM patient-group (68.7% versus 43.5%, p<0.001) and were less often living alone (15.5% versus 22.0%, p<0.001). The ROM reference-group showed higher levels of education relative to the ROM patient-group (77.2% higher education versus 49.5%, p<0.001).

Table 4.1. Sociodemographic and psychopathological characteristics of the ROM reference (n=1295) and ROM patient (n=4627) groups.

	ROM reference group (n= 1295)		ROM patient group (n=4627)		
Gender (%)					
Male	482	(37.2)	1779	(38.4)	(p=0.41)
Female	813	(62.8)	2848	(61.6)	(p=0.36)
Age (mean. SD) in years	40.3	(12.6)	39.3	(12.3)	(p=0.12)
Male	41.2	(12.6)	41.2	(12.0)	(p=0.98)
Female	39.7	(12.6)	38.1	(12.3)	(p=0.001)
Marital status (%)*				(p<0.001)	
Married/cohabitating	890	(68.7)	2027	(43.8)	
Divorced/separated/widow	78	(6.0)	689	(14.9)	
Single	327	(25.3)	1382	(39.9)	

Table 4.1: continued.

	ROM reference group (n= 1295)		ROM patient group (n=4627)	
Housing situation (%)*				(p<0.001)
Living alone	201	(15.5)	995	(21.5)
Living with partner	902	(69.7)	2067	(44.7)
Living with family	192	(14.8)	1036	(22.4)
Educational status (%)*[1]				(p<0.001)
Lower	295	(22.8)	1843	(39.8)
Higher	1000	(77.2)	2253	(48.7)
Employment status (%)*				(p<0.001)
Employed part-time	509	(39.3)	838	(18.1)
Employed full-time	554	(42.8)	803	(17.4)
Unemployed/retired	197	(15.2)	1189	(25.7)
Work-related disability (%)	35	(2.7)	1268	(27.4)
Ethnic background (%)* - **				(p<0.001)
Dutch	1150	(88.8)	3103	(67.1)
Other ethnicity	134	(10.3)	954	(20.6)
MINI diagnoses (%)				(p<0.001)
Currently None	1174	(90.7)	0***	
Mood disorder (single)	7	(0.5)	2159	(46.7)
Mood disorder (with comorbidity)	10	(0.8)	2468	(53.3)
Other psychiatric disorder	104	(8.1)	0***	

*No data from 70 (1.5%) - 570 (12.3%) patients; **No data from 11 reference subjects
***Selection criterion
[1] Lower education: general basic education only, or lower vocational education
Higher education: middle or higher vocational education, college or university

Furthermore, work-related disability and unemployment were less prevalent in the ROM reference-group (17.9% versus 54.3%, p<0.001). Fewer participants in the ROM reference-group were of ethnic origin (defined as oneself or both parents not being born in the Netherlands). In keeping with our decision to exclude patients without a diagnosis of MDD or DD, all subjects from the ROM patient-group had at least one DSM-IV disorder.

In the ROM reference-group, on the other hand, 10.9% had a DSM-IV disorder. In the ROM patient-group, a high proportion of subjects (53.3%) reported psychopathological co-morbidity (e.g. anxiety disorders and/or somatoform disorders).

REFERENCE VALUES

Percentiles, means and SDs

Table 4.2 presents the percentile scores and the mean scores for the BDI-II, IDS-SR, and MADRS for both the ROM reference-group and the patient-group. For the ROM reference-group, the distributions of total scores and subscale scores were positively skewed, indicating apparent health. This was also demonstrated by the substantial percentage of participants with the lowest possible scores. The Supplementary Tables 4.1-4.3 depict the percentile scores and the mean scores for men and women separately.

For the ROM reference-group, the cut-off (P_{95}) value was 13 for the BDI-II total score, 20 for the IDS-SR total score, and 11 for the MADRS total score. For the ROM patient-group the cut-off (P_5) value was 14 for the BDI-II total score, 18 for the IDS-SR total scale, and 11 for the MADRS total score. The mean BDI-II total score was 3.7 (SD=4.7) for the ROM reference-group, indicating that the majority was not depressed. For comparison, in the MDD patient-group the mean was 30.8 (SD=10.5), indicating severe depression in the majority of patients. The mean IDS-SR total score was 6.7 (SD=6.9) for the ROM reference-group, compared to 38.1 (SD=12.1) for the MDD patient-group. The mean MADRS score was 2.8 (SD=3.8) for the ROM reference-group, compared to 23.4 (SD=7.8) for the MDD patient-group.

For the self-report instruments (BDI-II and IDS-SR), analyses of gender and age indicated that advancing age was associated with more symptoms of psychopathology for both genders (see Supplementary Tables 4.3 through 4.5). There was a tendency for healthy women to show higher cut-off scores on all three MDD severity scales relative to healthy men.

In a sensitivity analysis, we excluded from the ROM reference-group all 126 subjects (9.7%) with a MINI-diagnosis. Among the remaining 1169 subjects, we found that the median of the changes on the BDI-II, the IDS-SR total score and subscale scores, and the MADRS was 11% (interquartile range 9 to 14%). The median of the changes of the P_{95} scores was 15% (interquartile range 11 to 20%).

Receiver operating characteristic (ROC) curves

Cut-off points, defined by equal sensitivity and specificity, were calculated with ROC analyses (see Table 4.3). See also Figure 4.1 for the discriminative power of each of the three MDD scales.

The cut-off point of the BDI-II, which discriminated health from disease (i.e.,, the ROM reference-group from the ROM patient-group), was 13.5, with a sensitivity and specificity of 96%. Therefore for subjects without psychopathology, 4% of those with a total

Table 4.2: Percentiles and mean values for Routine Outcome Monitoring mood disorder instruments in the ROM reference (n=1295) and patient (n=4627) groups.

	ROM reference group							ROM patient group						
	N	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	N	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Beck Depression Inventory-II (BDI-II)*														
Total score	455	0	0	2	5	13	3.74 ±4.74	4019	14	24	30	38	49	30.80 ±10.52
Cognitive[1]	455	0	0	0	1	5	0.98 ±1.67	4019	2	7	10	14	19	10.44 ±5.19
Somatic-Affective[1]	455	0	0	2	4	10	2.76 ±3.53	4019	9	16	20	25	31	20.36 ±6.64
Inventory of Depressive Symptomatology - Self-Report*														
Total score	769	0	2	5	9	20	6.74 ±6.88	474	18	30	38	46	58	38.05 ±12.07
Atypical characteristics[2]	196	1	3	4	6	11	4.71 ±3.01	208	8	11	14	17	21	14.12 ±4.13
Melancholic characteristics[3]	165	1	3	4	6	10	4.62 ±2.31	115	6	8	10	12	17	10.24 ±3.5
Montgomery Åsberg Depression Rating Scale**														
Total score	1291	0	0	2	4	11	2.79 ± 3.84	4627	11	18	23	28	36	23.44 ±7.75

* * BDI-II samples and IDS-SR samples were non-overlapping, while the MADRS sample was overlapping with BDI-II or IDS-SR samples.
[1] Steer et al., 1987; Cognitive subscale comprises items 2, 3, 5, 6, 7, 8, 9 and 14; Somatic-Affective subscale comprises items 1, 4, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20 and 21
[2] Novick et al., 2005: Atypical characteristics subscale comprises items 4, 8, 9, 10, 11, 15, 18, 24, 33 and 34
[3] van Reedt Dortland et al;., 2010: Melancholic characteristics subscale comprise items 3, 6, 9, 10, 11, 12, 17, 20, 25, 27 and 28

score of 13.5 or higher would be incorrectly classified as depressed. By the same token, 4% of the ROM patient-group who had a total score of 13.5 or lower would be incorrectly classified as non-depressed. The AUC value (0.99), indicating the discriminating performance, showed that the BDI-II performed excellently in making a distinction between patients and non-patients. Both subscales showed excellent discriminative power. The best performing subscale was the Somatic-Affective subscale, with AUC=0.99.

The cut-off point of the IDS-SR total score, which discriminated the ROM reference-group from the ROM patient-group, was 18.5 with a sensitivity and specificity of 94%. The AUC value was 0.98 for the total score, showing excellent discriminative power. The best performing subscale was Melancholic Characteristics (AUC=0.97).

The cut-off point of the MADRS total score, which discriminated the ROM reference-group from the ROM patient-group, was 10.5 with a sensitivity and specificity of 95%. The AUC value was 0.99, showing excellent discriminative power.

Internal consistency reliability

The internal consistency reliability of the instruments (for all subjects combined) is presented in Table 4.3. The total scales of all three instruments showed excellent internal consistency. Except for the IDS-SR Atypical Characteristics subscale (with a questionable alpha of 0.68), none of the subscales had Cronbach's alphas below the critical cut-off of 0.70, indicating adequate internal consistency.

Table 4.3: Percentiles and mean values for Routine Outcome Monitoring mood disorder instruments in the ROM reference (n=1295) and patient (n=4627) groups.

	Cronbach's Alpha	Number of items	N	ROC cut-off	AUC	Sensitivity/ specificity
Beck Depression Inventory-II (BDI-II),			4474*			
Total score	0.93	21		13.5	0.99	0.96
Cognitive[1]	0.87	8		3.5	0.97	0.91/.093
Somatic-Affective[1]	0.89	13		9.5	0.99	0.95
Inventory of Depressive Symptomatology-Self-Report (IDS-SR)			1243*			
Total score	0.94	32		18.5	0.98	0.94
Atypical characteristics[3]	0.68	10		7.5	0.92	0.84
Melancholic characteristics[º]	0.78	11		9.5	0.97	0.94/0.91
Montgomery Åsberg Depression Rating Scale (MADRS)			5918**			
Total score	0.90	10		10.5	0.99	0.95

BDI-II samples and IDS-SR samples were non-overlapping.
** MADRS samples were partly overlapping with BDI-II or IDS-SR samples.
[1] Steer et al., 1987; Cognitive subscale comprises items 2, 3, 5, 6, 7, 8, 9 and 14; Somatic-Affective subscale comprises items 1, 4, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20 and 21.
[3] Novick et al., 2005: Atypical characteristics subscale comprises items 4, 8, 9, 10, 11, 15, 18, 24, 33 and 34
[º] van Reedt Dortland et al;., 2010: Melancholic characteristics subscale comprise items 3, 6, 9, 10, 11, 12, 17, 20, 25, 27 and 28.

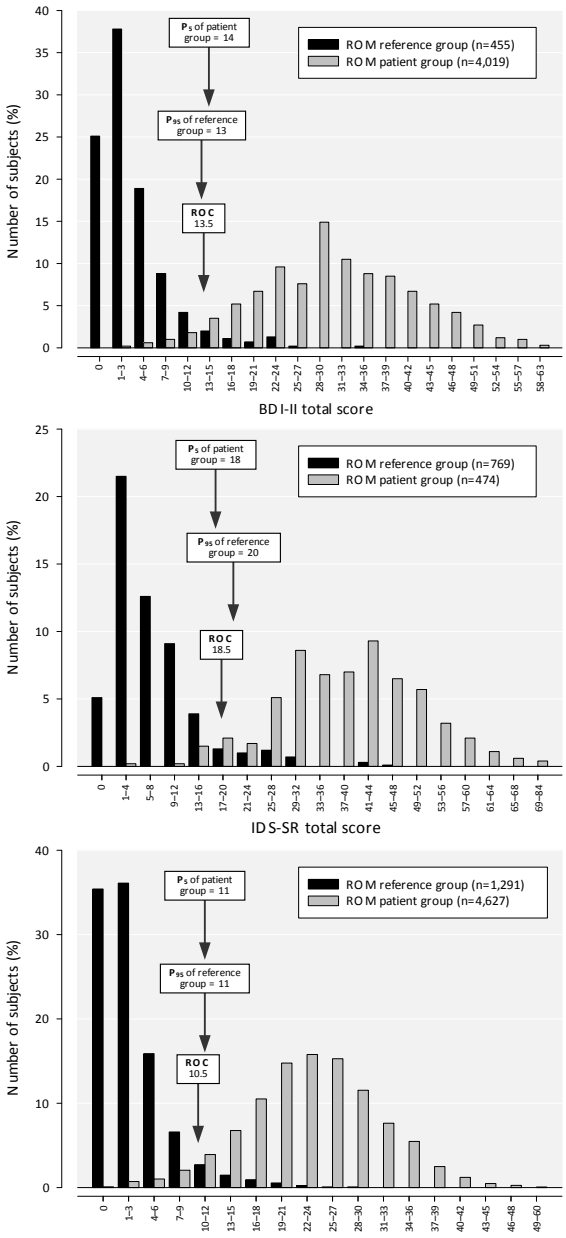


Figure 4.1: Distribution of the scores of Beck Depression Inventory-II (BDI-II) total scale, the Inventory of Depressive Symptoms (self-report) (IDS-SR) total scale and the Montgomery Äsberg Depression Rating Scale (MADRS). Three types of cut-off points are depicted: the 75th percentile score (P_{75}), the 95th percentile score (P_{95}) and the Receiver Operating Characteristics (ROC) cut-off point defined by equal sensitivity and specificity.

DISCUSSION AND CONCLUSION

We reported reference values for the generic instruments BDI-II, IDS-SR, and MADRS in large samples from a 'healthy' population (i.e., ROM reference-group) and a 'psychiatrically ill' population (i.e., ROM patient-group). P_{95} values of the ROM reference-group, ROC analysis based cut-off scores, and P_5 values of the ROM patient-group yielded almost equal values. A remarkable gender-specific pattern in reference values was observed, with women in the general population showing higher values than men. Our data suggest that gender-specific reference values will increase precision in the assessment of the clinical state of psychiatric outpatients. Advancing age was associated with more symptoms of psychopathology for the BDI-II and IDS-SR. Therefore, to be regarded as recovered, a young man would need to have lower scores on generic scales than would an older woman. The BDI-II and MADRS showed different results. This could be explained by the fact that the BDI-II a more symptom-specific instrument is, relative to the MADRS. However, the MADRS and the BDI-II provide internally consistent estimates of depression severity [53].

The mean BDI-II scores for the ROM reference-group (mean=3.8, SD=4.7) were lower than the mean BDI-II scores reported by Segal et al. (2008; mean=8.6, SD=7.7), suggesting that our reference-group was relatively healthy. The mean IDS-SR score for our ROM reference-group (mean=6.7, SD=6.9), however, was slightly higher than the mean IDS-SR scores reported by Rush et al. (1986; mean=2.1, SD=2.2). It should be noted, however, that their skewed distributions preclude an accurate comparison of these two estimates. The ROC cut-off value for the ROM reference-group (18.5) was similar to the value reported by Rush et al. (18.0) [19]. The mean MADRS score for the ROM reference-group (mean=2.8, SD=3.8) was slightly lower than the weighted mean MADRS score reported by Zimmerman et al. (M=4.0, SD=5.8) [39]. These differences among studies are relatively small and of minor clinical importance, and may be due to sociodemographic and socio-cultural differences. The larger size of our ROM reference-group has probably yielded rather precise estimates. The high internal consistencies of the BDI-II, IDS-SR, and MADRS are in accordance with previous studies [19,21,23-26,29,31].

The GP sample in our study is representative of the general population, given that almost everyone in the Netherlands is registered with a family doctor/GP. The reference values established in the present study can be used to determine whether a patient's level of symptoms falls within the normal range of values after treatment (i.e., whether a treated patient is no longer any different from normal controls with respect to the level of depressive symptoms). Normality can be defined statistically or medically. The statistical model is based on the distribution of scores from the general population (including all individuals) and on deviation from the mean. The medical model considers psychopathology and normality (i.e., absence of psychopathology) in absolute terms. It excludes individuals with a disorder from a reference-group. In our study we chose the statistical approach in which we included all non-treated individuals, both with and without (subclinical) symptoms. We have chosen

for this approach because we wanted to have reference values that were representative for the population that was not treated in secondary care. If we had excluded subjects with a MINI-diagnosis from the main analysis, the reference values would probably have been too strict. Our results showed that the reference values were not affected to any large extent by our inclusive methodology.

Country-specific normative data are important, because reference values are not necessarily the same in different translations and across different cultures [54,55]. Compared to English reference values, our values were slightly lower for the reference population, as were the previously-published Dutch reference values for the BDI-II (Beck et al. 2002). Until now, no Dutch reference values have been reported for the IDS-SR and MADRS. Our data showed a somewhat lower mean total score on the MADRS relative to that reported by Zimmerman et al. (2004a; M=4.0, SD=5.8) in a review of studies of the MADRS in healthy controls.

The following clinical implications arise from the results of the current study. The excellent performance of the instruments indicates that our reference values are suitable for different purposes: 1) decisions about treatment termination and referral back to primary care; and 2) identification of people who may benefit from therapy or from referral by primary care to specialized mental health care. Although the scales are not validated as diagnostic instruments, the ROC analyses suggested that the discriminative power of the instruments was excellent. Therefore , these cut-off values can aid in screening for MDD, although clinical judgment and validated diagnostic tools remain the gold standard (e.g., MINI [42,43]]; Composite International Diagnostic Interview [CIDI; [56]]; the Structured Clinical Interview for Diagnostic and Statistical Manual [SCID; [57]. Moreover, cut-off scores may be used to classify depression severity [16]. When making decisions about treatment termination or referral to primary care, specificity has to be high. The 95th percentile score of the ROM reference-group may result in few false positives. For referral from primary care to specialized mental health care, cut-off scores with a high sensitivity are more appropriate, and we recommend the use of ROC-based cut-offs or 5th percentile scores from the ROM patient-group.

It is noteworthy that the 95th percentile of the reference-group, the 5th percentile of the patient-group, and the ROC cut-off values overlapped considerably. They were also largely consistent with the internationally used cut-off values for the BDI-II, IDS-SR, and MADRS to distinguish individuals without depressive symptoms from those with mild symptoms (values of 14, 18, and 12, respectively; [16,19,20]. Furthermore, on average, women from the ROM reference-group scored higher on all three MDD severity scales relative to males. For the BDI-II, the respective P_{95} values for women and men were 15 and 12. For the IDS-SR the respective values were 23 and 18, and for the MADRS the respective values were 13 and 9. It may be too early to recommend gender-specific reference values, as more research (e.g., replication) is needed in reference populations. Nevertheless, it was striking that reference values from non-depressed populations showed clinically important

gender differences. Most previous studies did not stratify for gender [25,27,39] but in the one study in which stratification did occur [19], no gender difference was found. For the self-report instruments (BDI-II and IDS-SR), higher age was associated with higher P_{95} reference scores for both women and men in the ROM reference-group; this was not the case for the observer-rated MADRS. Therefore, age-specific reference values seem redundant.

The present study has several strengths. First, the assessment procedures for both groups were standardized and of high quality. The interviewers were specially trained research nurses and psychologists who were regularly supervised. Second, the ROM reference-group was large, it was clearly defined, and it resembled the patient-group in all relevant respects (age, gender, level of urbanization) other than those under investigation (level of psychopathology). Furthermore, the reference-group is probably quite representative of the general population, given the high GP registration rate in the Netherlands.

Limitations of the present study include the non-response in the ROM reference-group. At 63.2%, this was substantial, pointing to potential selection bias. Bias may have resulted in slightly higher cut-off and percentile scores. Furthermore, because the BDI-II was replaced by the IDS-SR during the study, sample sizes of ROM patients that completed the IDS-SR (n=474) and of ROM reference subjects that completed the BDI-II (n=455) were not as high as the other sample sizes. Additionally, the generalizability of the results is limited by the nature of our ROM reference-group; it comprised Dutch-speaking people aged between 18 and 65 years. Reference values may not automatically be applicable to other ethnic or cultural groups, or to children and the elderly.

In conclusion, this large-scale population-based study provides reference values and reliability coefficients for the BDI-II, IDS-SR, and MADRS. These reference values improve the usability of the instruments as ROM instruments for the assessment of severity of mood disorder symptoms. Either instrument can be administered to every patient with a depressive disorder to help make responsible decisions about continuing, changing, or terminating therapy. Additionally, these reference values are suitable for indentifying patients that have recovered enough to be referred back from specialized mental health care to primary care.

Reference List

1. De Beurs E, Den Hollander-Gijsman ME, Van Rood YR, Van der Wee NJ, Giltay EJ, Van Noorden MS, Van der Lem R, Van Fenema EM, Zitman FG. (2011) Routine outcome monitoring in the Netherlands: practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clin Psychol Psychother,* 18, 1-12.

2. Schulte-van Maaren YWM, Carlier IVE, Giltay EJ, Van Noorden MS, De Waal MW, Van der Wee NJ, Zitman FG. (2012) Reference values for mental health assessment instruments: objectives and methods of the Leiden Routine Outcome Monitoring Study. *J Eval Clin Pract*.

3. Schulte-van Maaren YWM, Carlier IVE, Zitman FG, Hemert AM, De Waal MW, Van Noorden MS, Giltay EJ. (2012) Reference values for generic instruments used in Routine Outcome Monitoring: the Leiden Routine Outcome Monitoring Study (in press). *BMC Psychiatry*.

4. Kazdin AE. (2008) Evidence-based treatment and practice: new opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *Am Psychol*, 63 (3), 146-159.

5. Solberg HE. (1989) Reference values. *Adv Clin Chem*, 27, 1-79.

6. Solberg HE. (2008) Establishment and use of reference values. Burtis CA, Ashwood ER, Bruns DE, editors. Fundamentals of clinical chemistry. 6[14], 229-238. St. Louis, Missouri, Saunders Elsevier.

7. Geffre A, Friedrichs K, Harr K, Concordet D, Trumel C, Braun JP. (2009) Reference values: a review. *Vet Clin Pathol,* 38 (3), 288-298.

8. Katayev A, Balciza C, Seccombe DW. (2010) Establishing reference intervals for clinical laboratory test results: is there a better way? Am *J Clin Pathol*, 133 (2), 180-186.

9. Sasse EA, Doumas BT, Miller WG, D'Orazio P, Eckfeldt JH, Evans SA, et al. (2000) How to define and determine reference intervals in the clinical laboratory; approved guideline-Second edition. NCCLS document C28-A2 . 20[13], 1-38. Wayne, PA, NCCLS.

10. Kendall PC, Marrs-Garcia A, Nath SR, Sheldrick RC. (1999) Normative comparisons for the evaluation of clinical significance. *J Consult Clin Psychol,* 67 (3), 285-299.

11. Horn PS, Feng L, Li Y, Pesce AJ. (2001) Effect of outliers and nonhealthy individuals on reference interval estimation. *Clin Chem*, 47 (12), 2137-2145.

12. Reed AH, Henry RJ, Mason WB. (1971) Influence of statistical method used on the resulting estimate of normal range. *Clin Chem*, 17 (4), 275-284.

13. Jacobson NS, Truax P. (1991) Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*, 59 (1), 12-19.

14. Fan J, Upadhye S, Worster A. (2006) Understanding receiver operating characteristic (ROC) curves. CJEM, 8 (1), 19-20.

15. Beck AT, Steer RA. (1987) Manual for the revised Beck Depression Inventory. San Antonio, TX, Psychological Corporation.

16. Beck AT, Steer RA, Brown GK. (1996) Manual for the Beck Depression Inventory-II. San Antonio,TX, Psychological Corporation.

17. Beck AT, Steer RA, Brown GK. (2002) Beck Depression Inventory-II-NL. Handleiding. De Nederlandse versie van de Beck Depression Inventory, 2nd edition. (A.J.W.van der Does, vert. en bew.). Lisse, NL, Swets Test Publishers.

18. Nolen WA, Dingemans PMAJ. (2004) Meetinstrumenten bij stemmingsstoornissen (Instruments for measuring mood disorders). *Tijdschrift voor psychiatrie*, 46, 681-686.

19. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. (1996) The Inventory of Depressive Symptomatology (IDS): psychometric properties. *Psychol Med*, 26 (3), 477-486.

20. Montgomery SA, Asberg M. (1979) New Depression Scale Designed to be Sensitive to Change. *British Journal of Psychiatry*, 134 (APR), 382-389.

21. Beck AT, Steer RA, Ball R. (1996) Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. *J Pers Assess*, 67 (3), 588-597.

22. Steer RA, Ball R, Ranieri WF, Beck AT. (1999) Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. *J Clin Psychol,* 55 (1), 117-128.

23. Kumar G, Steer RA, Teitelman KB, Villacis L. (2002) Effectiveness of Beck Depression Inventory-II subscales in screening for major depressive disorders in adolescent psychiatric inpatients. *Assessment*, 9 (2), 164-170.

24. Steer RA, Rissmiller DJ, Beck AT. (2000) Use of the Beck Depression Inventory-II with depressed geriatric inpatients. *Behav Res Ther*, 38 (3), 311-318.

25. Segal DL, Coolidge FL, Cahill BS, O'Riley AA. (2008) Psychometric properties of the Beck Depression Inventory II (BDI-II) among community-dwelling older adults. *Behav Modif,* 32 (1), 3-20.

26. Biggs MM, Shores-Wilson K, Rush AJ, Carmody TJ, Trivedi MH, Crismon ML, Toprac MG, Mason M. (2000) A comparison of alternative assessments of depressive symptom severity: a pilot study. *Psychiatry Res*, 96 (3), 269-279.

27. Rush AJ, Giles DE, Schlesser MA, Fulton CL, Weissenburger JE, Burns C. (1986) The Inventory for Depressive Symptomatology (IDS): preliminary findings. *Psychiatry Res*, 18 (1), 65-87.

28. Rush AJ, Giles DE, Schlesser MA, Orsulak PJ, Parker CR, Jr., Weissenburger JE, Crowley GT, Khatami M, Vasavada N. (1996) The dexamethasone suppression test in patients with mood disorders. *J Clin Psychiatry,* 57 (10), 470-484.

29. Trivedi MH, Rush AJ, Ibrahim HM, et al. (2004) The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation. *Psychol Med*, 34 (1), 73-82.

30. Corruble E, Legrand JM, Zvenigorowski H, Duret C, Guelfi JD. (1999) Concordance between self-report and clinician's assessment of depression. *J Psychiatr Res*, 33 (5), 457-465.

31. Bondolfi G, Jermann F, Rouget BW, Gex-Fabry M, McQuillan A, Dupont-Willemin A, Aubry JM, Nguyen C. (2010) Self- and clinician-rated Montgomery-Asberg Depression Rating Scale: evaluation in clinical practice. *J Affect Disord*, 121 (3), 268-272.

32. Hawley CJ, Gale TM, Sivakumaran T. (2002) Defining remission by cut off score on the MADRS: selecting the optimal value. *Journal of Affective Disorders*, 72 (2), 177-184.

33. Khan A, Khan SR, Shankles EB, Polissar NL. (2002) Relative sensitivity of the Montgomery-Asberg Depression Rating Scale, the Hamilton Depression rating scale and the Clinical Global Impressions rating scale in antidepressant clinical trials. *Int Clin Psychopharmacol,* 17 (6), 281-285.

34. Zimmerman M, Posternak MA, Chelminski I. (2004) Derivation of a definition of remission on the Montgomery-Asberg depression rating scale corresponding to the definition of remission on the Hamilton rating scale for depression. *J Psychiatr Res,* 38 (6), 577-582.

35. Müller MJ, Szegedi A, Wetzel H, Benkert O. (2000) Moderate and severe depression. Gradations for the Montgomery-Asberg Depression Rating Scale. *J Affect Disord,* 60 (2), 137-140.

36. Müller MJ, Himmerich H, Kienzle B, Szegedi A. (2003) Differentiating moderate and severe depression using the Montgomery-Asberg depression rating scale (MADRS). *J Affect Disord,* 77 (3), 255-260.

37. Sagen U, Vik TG, Moum T, Morland T, Finset A, Dammen T. (2009) Screening for anxiety and depression after stroke: comparison of the hospital anxiety and depression scale and the Montgomery and Asberg depression rating scale. *J Psychosom Res*, 67 (4), 325-332.

38. Engedal K, Kvaal K, Korsnes M, Barca ML, Borza T, Selbaek G, Aakhus E. (2012) The validity of the Montgomery-Asberg depression rating scale as a screening tool for depression in later life. *J Affect Disord*.

39. Zimmerman M, Chelminski I, Posternak M. (2004) A review of studies of the Montgomery-Asberg Depression Rating Scale in controls: implications for the definition of remission in treatment studies of depression. *Int Clin Psychopharmacol,* 19 (1), 1-7.

40. Van Noorden MS, Giltay EJ, Den Hollander-Gijsman ME, Van der Wee NJ, Van Veen T, Zitman FG. (2010) Gender differences in clinical characteristics in a naturalistic sample of depressive outpatients: the Leiden Routine Outcome Monitoring Study. *J Affect Disord*, 125 (1-3), 116-123.

41. Carlier IVE, Meuldijk D, Van Vliet IM, Van Fenema EM, Van der Wee NJ, Zitman FG. (2012) Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *J Eval Clin Pract*, 18 (1), 104-110.

42. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC. (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*, 59 Suppl 20, 22-33.

43. Van Vliet IM, Leroy H, Van Megen HJGM. (2000) M.I.N.I. Internationaal Neuropsychiatrisch Interview. Nederlandse Versie 5.0.0.

44. American Psychiatric Association. (2000) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision. Washington, DC, American Psychiatric Association.

45. Storch EA, Roberti JW, Roth DA. (2004) Factor structure, concurrent validity, and internal consistency of the Beck Depression Inventory-Second Edition in a sample of college students. *Depress Anxiety*, 19 (3), 187-189.

46. Whisman MA, Perez JE, Ramel W. (2000) Factor structure of the Beck Depression Inventory-Second Edition (BDI-II) in a student sample. *J Clin Psychol,* 56 (4), 545-551.

47. Rush AJ, Pincus HA, First MB. (2000) Handbook of psychiatric measures. Washington DC, American Psychiatric Association.

48. Novick JS, Stewart JW, Wisniewski SR, et al. (2005) Clinical and demographic features of atypical depression in outpatients with major depressive disorder: preliminary findings from STAR*D. *J Clin Psychiatry,* 66 (8), 1002-1011.

49. Van Reedt Dortland AKB, Giltay EJ, Van Veen T, Van Pelt J, Zitman FG, Penninx BW. (2010) Associations between serum lipids and major depressive disorder: results from the Netherlands Study of Depression and Anxiety (NESDA). *J Clin Psychiatry*, 71 (6), 729-736.

50. Heo M, Murphy CF, Meyers BS. (2007) Relationship between the Hamilton Depression Rating Scale and the Montgomery-Asberg Depression Rating Scale in depressed elderly: a meta-analysis. *Am J Geriatr Psychiatry*, 15 (10), 899-905.

51. Barnabei L, Marazia S, De CR. (2007) Receiver operating characteristic (ROC) curves and the definition of threshold levels to diagnose coronary artery disease on electrocardiographic stress testing. Part I: The use of ROC curves in diagnostic medicine and electrocardiographic markers of ischaemia. *J Cardiovasc Med (Hagerstown)*, 8 (11), 873-881.

52. Campbell WW, Robinson LR. (1993) Deriving reference values in electrodiagnostic medicine. *Muscle Nerve*, 16 (4), 424-428.

53. Uher R, Farmer A, Maier W, et al. (2008) Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychol Med*, 38 (2), 289-300.

54. Fernandez AL, Marcopulos BA. (2008) A comparison of normative data for the Trail Making Test from several countries: equivalence of norms and considerations for interpretation. *Scand J Psychol*, 49 (3), 239-246.

55. Geisinger KF. (1994) Cross-Cultural Normative Assessment: Translation and Adaptation Issues Influencing the Normative Interpretation of Assessment Instruments. *Psychological Assessment*, 6 (4), 304-312.

56. World Health Organization. (1990) The Composite International Diagnostic Interview (CIDI). Authorized Core Version 1.0. Geneva, WHO.

57. Spitzer RL, Williams JBW, Gibbon M, First MB. (1988) Structured Clinical Interview for DSM-III-R - Patient Version (SCID-P, 4/I/88). New York, Biometrics Research Department, New York State Psychiatric Institute.

Supplementary Material

Supplementary Table 4.1. Percentile scores and mean values in the ROM reference (n=455) and patient (n=4019) groups for **the Beck Depression Inventory-II (BDI-II).**

	ROM reference group (n=455)						ROM patient group (n=4019)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
BDI-II Total score												
All participants	0	0	2	5	13	3.74 ± 4.74	14	24	30	38	49	30.80 ± 10.52
- Women aged 18-40 yr	0	1	2	5	12	3.68 ± 4.15	15	25	32	40	50	32.21 ± 10.51
- Women aged 41-65 yr	0	1	3	6	18	4.62 ± 5.94	14	23	31	38	48	30.73 ± 10.66
- Men aged 18-40 yr	0	0	1	4	11	2.88 ± 4.07	14	23	29	36	48	29.85 ± 10.09
- Men aged 41-65 yr	0	1	2	5	13	3.26 ± 3.94	13	22	28	36	48	29.14 ± 10.40
BDI-II Cognitive												
All participants	0	0	0	1	5	0.98 ± 1.67	2	7	10	14	19	10.44 ± 5.19
- Women aged 18-40 yr	0	0	0	1	4	0.93 ± 1.44	3	7	11	15	20	11.30 ± 5.28
- Women aged 41-65 yr	0	0	0	1	5	1.07 ± 1.98	2	6	9	14	19	9.74 ± 5.35
- Men aged 18-40 yr	0	0	0	1	5	1.03 ± 1.78	3	7	10	14	19	10.66 ± 4.69
- Men aged 41-65 yr	0	0	0	1	4	0.87 ± 1.38	2	6	9	13	19	9.54 ± 4.96
BDI-II Somatic-Affective												
All participants	0	0	2	4	10	2.76 ± 3.53	9	16	20	25	31	20.36 ± 6.64
- Women aged 18-40 yr	0	0	2	4	9	2.75 ± 3.23	10	17	21	25	31	20.91 ± 6.50
- Women aged 41-65 yr	0	0	2	5	12	3.55 ± 4.41	10	16	21	26	32	20.99 ± 6.63
- Men aged 18-40 yr	0	0	1	3	7	1.86 ± 2.58	8	15	19	24	31	19.19 ± 6.77
- Men aged 41-65 yr	0	0	1	3	9	2.39 ± 2.99	9	15	19	24	31	19.60 ± 6.58

Supplementary Table 4.2: Percentile scores and mean values in the ROM reference (n=769) and patient (n=474) groups for the subscales and total score of the **Inventory of Depressive Symptomatology - Self-Report (IDS-SR).**

	ROM reference group							ROM patient group						
	N	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	N	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
IDS-SR Total Score														
- Women aged 18-40 yr	769	0	2	5	9	20	6.74 ± 6.88	474	18	30	38	46	58	38.05 ± 12.07
- Women aged 41-65 yr		0	3	6	10	22	7.73 ± 7.33		19	30	37	49	59	39.06 ± 11.93
- Men aged 18-40 yr		0	3	5	10	24	7.54 ± 7.56		20	32	40	47	58	39.34 ± 11.08
- Men aged 41-65 yr		0	2	3	7	17	4.97 ± 5.29		14	28	34	43	50	34.68 ± 10.55
		0	2	3	8	19	5.40 ± 5.70		17	27	38	47	61	37.81 ± 13.92
Atypical characteristics[1]														
All participants	165	1	3	4	6	10	4.62 ± 2.31	115	6	8	10	12	17	10.24 ± 3.5
- Women aged 18-40 yr		1	3	5	6	10	4.89 ± 2.67		6	9	10.5	13	20	11.29 ± 3.53
- Women aged 41-65 yr		3	4	5	6	9	4.78 ± 1.69		4	7	10	11	18	9.70 ± 3.32
- Men aged 18-40 yr		1	3	4	5	8	3.89 ± 1.9		2	8	9	12	14	9.52 ± 3.23
- Men aged 41-65 yr		1	3	4	7	10	4.63 ± 2.39		5	7	9	12	17	9.69 ± 3.58
Melancholic characteristics[2]	196	1	3	4	6	11	4.71 ± 3.01	208	8	11	14	17	21	14.12 ± 4.13
All participants		1	2	4	6	11	4.76 ± 3.19		8	11	13	17	22	14.01 ± 4.03
- Women aged 18-40 yr		2	4	5	7	13	5.62 ± 2.93		8	10	14	17	20	13.52 ± 3.89
- Women aged 41-65 yr		1	2	3	5	10	3.87 ± 2.51		5	11	14	16	19	13.46 ± 3.54
- Men aged 18-40 yr		1	3	3	6	11	4.37 ± 2.95		9	12	15	20	25	15.40 ± 4.75
- Men aged 41-65 yr		0.00	0.00	0.20	0.20	0.60	0.17 ± 0.23		0.00	0.20	0.60	1.20	2.60	0.90 ± 0.84

[1] Novick et al., 2005: Atypical characteristics subscale comprises items 4, 8, 9, 10, 11, 15, 18, 24, 33 and 34
[2] van Reedt Dortland et al;., 2010: Melancholic characteristics subscale comprise items 3, 6, 9, 10, 11, 12, 17, 20, 25, 27 and 28

Supplementary Table 4.3: Percentile scores and mean values in the ROM reference (n=1291) and patient (n=4627) groups for the **Montgomery Åsberg Depression Rating Scale (MADRS).**

	ROM reference group (n=455)						ROM patient group (n=4019)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
BDI-II Total score												
All participants	0	0	2	4	11	2.79 ± 3.84	11	18	23	28	36	23.44 ± 7.75
- Women aged 18-40 yr	0	0	2	4	12	3.29 ± 4.09	10	18	23	28	35	23.02 ± 7.40
- Women aged 41-65 yr	0	0	2	4	13	2.99 ± 4.29	11	19	24	29	37	24.13 ± 7.85
- Men aged 18-40 yr	0	0	1	3	9	2.27 ± 3.17	9	18	23	28	36	22.63 ± 7.89
- Men aged 41-65 yr	0	0	1	3	9	2.08 ± 2.97	12	18	24	30	37	24.05 ± 8.01

NormQuest

Reference Values for ROM Instruments and Questionnaires

Chapter 5

Reference values for anxiety questionnaires

Yvonne W.M. Schulte-van Maaren
Erik J. Giltay
Albert M. van Hemert
Frans G. Zitman
Margot W.M. de Waal
Ingrid V.E. Carlier

Journal of Affective Disorders (2013) 150 (3), 1008-1018

ABSTRACT

Background: The monitoring of patients with an anxiety disorder can benefit from Routine Outcome Monitoring (ROM). As anxiety disorders differ in phenomenology, several anxiety questionnaires are included in ROM: Brief Scale for Anxiety (BSA), PADUA Inventory Revised (PI-R), Panic Appraisal Inventory (PAI), Penn State Worry Questionnaire (PSWQ), Worry Domains Questionnaire (WDQ), Social Interaction, Anxiety Scale (SIAS), Social Phobia Scale (SPS), and the Impact of Event Scale-Revised (IES-R). We aimed to generate reference values for both 'healthy' and 'clinically anxious' populations for these anxiety questionnaires.

Methods: We included 1295 subjects from the general population (ROM reference-group) and 5066 psychiatric outpatients diagnosed with a specific anxiety disorder (ROM patient-group). The MINI was used as diagnostic device in both the ROM reference group and the ROM patient group. To define limits for one-sided reference intervals (95^{th} percentile; P_{95}) the outermost 5% of observations were used. Receiver Operating Characteristics (ROC) analyses were used to yield alternative cut-off values for the anxiety questionnaires.

Results: For the ROM reference-group the mean age was 40.3 years (SD=12.6), and for the ROM patient-group it was 36.5 years (SD=11.9). Females constituted 62.8% of the reference-group and 64.4% of the patient-group. P_{95} ROM reference group cut-off values for reference versus clinically anxious populations were 11 for the BSA, 43 for the PI-R, 37 for the PAI Anticipated Panic, 47 for the PAI Perceived Consequences, 65 for the PAI Perceived Self-efficacy, 66 for the PSWQ, 74 for the WDQ, 32 for the SIAS, 19 for the SPS, and 36 for IES-R. ROC analyses yielded slightly lower reference values. The discriminative power of all eight anxiety questionnaires was very high.

Limitations: Substantial non-response and limited generalizability.

Conclusions: For 8 anxiety questionnaires, the BSA, PI-R, PAI, PSWQ, WDQ, SIAS, SPS, and IES-R, a comprehensive set of reference values was provided. Reference values were generally higher in women than in men, implying the use of gender-specific cut-off values. Each instrument can be offered to every patient with MAS disorders to make responsible decisions about continuing, changing or terminating therapy.

INTRODUCTION

Anxiety disorders are characterized by pervasive, persistent, anxious affective states. The DSM-IV recognizes various specific types of anxiety disorders: panic disorder (PD); phobic disorders (i.e., agoraphobia (AD), social phobia (SoPD), and specific phobia (SpPD)); obsessive-compulsive disorder (OCD); acute stress disorder (ASD); posttraumatic stress disorder (PTSD); and generalized anxiety disorder (GAD). Anxiety disorders frequently occur as comorbid disorders. The current global prevalence of anxiety disorders is 7.3% (4.8–10.9%), ranging from 5.3% (3.5–8.1%) in African cultures to 10.4% (7.0–15.5%) in Euro/Anglo cultures [1]. Lifetime prevalence rates in the Netherlands are 19.6% for any anxiety disorder, 3.8% for PD, 0.9% for AD, 9.3% for SoPD, 0.9% for OCD, 7.4% for PTSD, and 4.5% for GAD [2-4].

Routine outcome monitoring (ROM) is the assessment of treatment outcome at regular intervals in order to monitor patients' progress during treatment. Alongside generic questionnaires completed by all patients, patients who meet the criteria for a particular disorder can be administered disorder-specific questionnaires [5,6] The correct interpretation of ROM results for making clinical decisions about continuing, altering, or terminating treatment requires reliable ratings from reference populations [7]. These ratings can be used to determine whether a patient's level of symptoms falls within the normal range of values following treatment (e.g., whether a treated patient is now no different from normal controls with respect to the severity of anxiety symptoms).

Important issues regarding reference values appear in the literature. First, when data tend toward a non-Gaussian distribution, non-parametric percentile scores provide more appropriate reference values compared to parametric means and standard deviations (SDs) [8,9], and to weighted cut-off values calculated by the Jacobson & Truax method [10]. In that case, the 95th percentile (P_{95}) of the reference-group and the 5th percentile (P_5) of the patient-group commonly serve as reference values [9]. Second, when both reference data and patient data are available, Receiver Operating Characteristics (ROC) analyses can be used to provide cut-offs. The optimal trade-off between sensitivity and specificity, the point of (near) equality, leads to the optimal number of false results (i.e., false positives plus false negatives) [11], depending on the prevalence of the disorder in the general population. It is of note that this applies to disorders that are not very rare. Third, reference values are often established in healthy populations [9]. Absolute health does not exist but is a relative statement. Health should nevertheless be clearly defined, a priori, via inclusion and exclusion criteria [12-14]. Kendall et al., [15] stated that excluding with MDD participants from the reference group if they exhibit elevated levels of the target psychopathology, might lead to creating a nonrepresentative, "supernormal" sample. When comparing the patient group with a supernormal reference group an overly stringent criterion with unreasonable narrow reference intervals would be the result [16]. The inclusion of all possible participants in the

reference group, including those who may currently be experiencing elevated levels of psychopathology is therefore preferable. The goal is to generate a sample that is representative of the general community population [15]. This is in line with a statistical definition of normality, as opposed to a medical definition, both proposed by Wakefield [17]. The statistical perspective of normality is based on the distribution of scores from the population, including all individuals who are not currently treated in secondary care, with extreme scores considered as deviant. The medical perspective excludes individuals with psychopathology from the reference group. A similar definition of disease was given by Cohen [18]: "quantitative deviations from the normal". Fourth, to reduce the amount of uncertainty and random error, (sub)sample sizes of at least 120 are needed [8].

Symptoms of anxiety are suitable for self-rating because anxious persons in general tend to have rather realistic perception and insight (relative to other psychopathological conditions) [19]. We focused on 8 anxiety questionnaires that are often implemented in ROM (Table 5.1). These questionnaires are the self-rated PADUA Inventory Revised (PI-R), Panic Appraisal Inventory (PAI), Penn State Worry Questionnaire (PSWQ), Worry Domains Questionnaire (WDQ), Social Interaction, Anxiety Scale (SIAS), Social Phobia Scale (SPS), and the Impact of Event Scale-Revised (IES-R). Finally, the Brief Scale for Anxiety (BSA) is an observer-rated scale

For healthy control groups, reference values (in the form of means and SDs) have been published for the following questionnaires: PI-R [20,21], PSWQ [22-30], WDQ [26-28,30], both SIAS and SPS [22,31,32], and IES-R [33]. To our knowledge, no reference values have been reported for the BSA and the PAI. For patient groups, means and SDs were published for the BSA [34,35], the PI-R [20,21,36], the PAI [37-39], the PSWQ [22,23,25,40], the WDQ [40], both the SIAS and SPS [22,31], and the IES-R [33,41-44]. However, because of the strong positively skewed distribution of total scores in healthy populations, such as our ROM reference-group, the assumption of a normal distribution is unlikely to be satisfied [8,9]. Reference values should preferably be based on a distribution-free percentile or ROC methodology.

In previous studies, cut-off values (i.e., clinical thresholds) were assessed for the PI-R [21], the PSWQ [23], and the IES-R [33] [45]. Gender differences were reported previously for the PSWQ and WDQ [25,26], the SIAS and the SPS [31,32], and the IES-R [43] healthy control groups. All of these studies reported higher mean values for women than for men. Characteristics of previous studies on reference values are summarized in Table 5.1. The aim of this study was to establish reference values for the BSA, PI-R, PAI, PSWQ, WDQ, SIAS, SPS, and IES-R. These reference values included percentile scores, ROC-based cut-off values, and the more commonly reported means and SDs. We compared a sample of 1295 subjects from the general population with a sample of 5066 outpatients suffering from anxiety disorders. A special contribution of the current study is that a healthy (but not necessarily symptom-free) reference-group was included, alongside a well-defined psychiatric patient-group and that both sample sizes were large.

METHODS

Participants

Our analyses of reference values were based on two study samples: a ROM reference-sample from the general population (i.e., the ROM reference-group) and a ROM sample of psychiatric outpatients diagnosed with at least one anxiety disorder (i.e., ROM patient-group).

A total of 1295 participants aged 18 to 65 years (mean age=40.3 years; SD=12.6; 62.8% females) were included in the ROM reference-group, as part of the 'Leiden Routine Outcome Monitoring Study' [6,46]. A representative general population sample was randomly selected from the registration systems of eight general practitioners (GPs) in the region of Leiden, the Netherlands. In the Netherlands, 99.9% of the general population is registered with a GP [47]. The aim was to recruit an apparently psychiatrically healthy reference-group (but not necessarily symptom-free). Therefore, persons who were receiving treatment for psychiatric disorders and/or alcohol or drugs dependency during the six months prior to assessment were excluded. Additional exclusion criteria were hearing impairment or limited cognitive or language abilities (i.e., aphasia, severe dyslexia or dementia; illiteracy or insufficient mastery of the Dutch language). To ensure that the group was demographically comparable to the ROM patient-group, the ROM reference-group was matched for gender, age and urbanization-level (62.3% urban). Participants in the ROM reference-group were assessed in a similar way to the ROM patient-group, except that those in the ROM reference-group completed every disorder-specific questionnaire. As noted previously, the response rate of the ROM reference-group recruitment was 37.1% [6,48], perhaps due to the extensive number of questionnaires which needed to be completed by participants. The BSA was completed by the majority of the ROM reference-group (n=1291), the self-report questionnaires were completed by 50% of the ROM reference-group (due to time-constraints).

The ROM patient-group consisted of a sample of 5066 psychiatric outpatients, aged between 18 and 65 years (mean age=39.3, SD=12.3; 61.0% females), who were diagnosed with and treated for anxiety disorders at the Leiden University Medical Center (LUMC) Department of Psychiatry or the Rivierduinen specialized mental healthcare centres. Baseline assessment was part of the usual ROM procedure. On average, 80% of the patients with a tentative diagnosis of mood-, anxiety- and/or somatoform (MAS) disorder were assessed with ROM in the study period [46]. The BSA was completed by the majority of the ROM patient-group (n=4368), the self-report questionnaires were completed by those who were diagnosed with the relevant anxiety disorder.

To diagnose psychopathology in a standardized manner according to the DSM-IV, a diagnostic interview with the Mini-International Neuropsychiatric Interview plus (MINI-Plus 5.0.0.) [49,50] was done in all participants.

Procedures and questionnaires

Procedures for the web-based ROM program of the LUMC Department of Psychiatry are

described in detail elsewhere [46,51]. For the current study, we used baseline ROM assessments that comprised a standardized diagnostic interview (Dutch version of the Mini-International Neuropsychiatric Interview Plus, version 5.00-R: MINI-Plus) [49,50], the gathering of sociodemographic and socioeconomic data, observer-rated scales, and self-report questionnaires. The assessments were performed by specially trained and constantly supervised research nurses in outpatient clinics of the LUMC and Rivierduinen. Table 5.1 presents the description of each questionnaire, including domains, subscales, ratings, and score-ranges, as well as the respective ROM sample sizes. Sample sizes were determined by participants that completed the particular questionnaire (and not by presence of a particular anxiety disorder). The MINI-Plus was used to establish the presence of Axis I diagnoses according to the DSM-IV.

The Medical Ethical Committee of the LUMC approved the general study protocol associated with ROM, in which ROM was administered as part of the routine treatment process for patients. It involved a comprehensive protocol (titled "Psychiatric Academic Registration Leiden database") which safeguarded the anonymity of patients and persons in the reference-group and ensured proper handling of the ROM data. At intake, patients were informed that the data would be used for research purposes, but only in anonymized form. If patients object to such use, their data were removed. The Medical Ethical Committee of the LUMC approved the regulations and agreed with this policy. In addition, persons in the ROM reference-group signed informed consent for the purpose of this study.

Table 5.1: Anxiety questionnaires used in Routine Outcome Monitoring

Questionnaire {Abbreviation}	Domain	number of items	Rating	Range for score	Our sample sizes Reference / Patient-group	Range for sample sizes in previous studies Reference/Patient-group	References
Brief Scale for Anxiety (vCPRS subscale) {BSA (vCPRS)}	General anxiety	10	0=symptom is absent; 6=symptom is totally dominant	0-60	1291 / 4368	- / 50-101	[47] [30] [29]
PADUA Inventory revised {PI-R}	Obsessive Compulsive Disorder		0=not at all; 4=very much		651 / 657	76-430 / 30-222	[48-50] [15] [16]
Impulses		7		0-28			
Washing		10		0-40			
Checking		7		0-28			
Rumination		11		0-44			
Precision		6		0-24			
Total		41		0-164			
Panic Appraisal Inventory {PAI}	Panic Disorder				630 / 1392	- / 35-47	[32,34] [33]
Anticipated panic		15	0=no chance of panic occurrence; 100=definite panic occurrence	0-100 (average score)			
Perceived consequences of panic:			0=not at all troubling; 10=extremely troubling				
Physical		5		0-50			
Social		5		0-50			
Loss of control		5		0-50			
Total		15		0-150			

Table 5.1: continued

Questionnaire {Abbreviation}	Domain	number of items	Rating	Range for score	Our sample sizes Reference / Patient-group	Range for sample sizes in previous studies Reference/Patient-group	References
Perceived self-efficacy in coping with panic		15	0=not confident at all; 100=completely confident	0-100 (average score)			
Penn State Worry Questionnaire {PSWQ}	Generalized Anxiety Disorder: Excessive and uncontrollable (pathological) worry	16	1=not at all typical of me; 5=very typical of me		651 / 893	32–1138 / 60–436	[20,23,35] [18,51] [24]
Worry Domains Questionnaire {WDQ}	Generalized Anxiety Disorder: Non-pathological worry		0=not at all; 4=extremely		649 / 887	136–432 / -	[23,25,52] [53]
Relationships		4		0-16			
Lack of confidence		5		0-20			
Aimless Future		8	0=no chance of panic occurrence; 100=definite panic occurrence	0-32			
Work incompetence		3	0=not at all troubling; 10=extremely troubling	0-12			
Financial		4		0-16			
Physical Health		6		0-24			
Total		30		0-120			

Table 5.1: continued

Questionnaire {Abbreviation}	Domain	number of items	Rating	Range for score	Our sample sizes Reference / Patient-group	Range for sample sizes in previous studies Reference/Patient-group	References
Social Interaction and Anxiety Scale {SIAS}	Social Phobia	20	0= not at all characteristic or true of me; 4- extremely characteristic or true of me	0-80	651 / 1231	21–482 / 13-165	[17,27]
Social Phobia Scale {SPS}	Social Phobia	20	0= not at all characteristic or true of me; 4- extremely characteristic or true of me	0-80	651 / 1237	21–482 / 13-165	[17,27]
Impact of Event Scale – Revised {IES-R}	Traumatic Events		0=not at all; 4=extremely		1272 / 390	154 / 120–4167	[39,54] [55];[28] [37]
Intrusions		8		0-32			
Avoidance		8		0-32			
Hyperarousal		6		0-24			
Total		22		0-88			
Mini International Neuropsychiatric Interview Plus 5.0.0. {MINI Plus 5.0.0}	General Pathology				1295/5066		[45,46]

Statistical analyses

Analyses were performed separately for the ROM reference-group and the patient-group, while ROC and internal consistency analyses were conducted using data from both groups combined. In both groups, participants who had more than one missing value per subscale were excluded. This allowed us to conduct a robust evaluation of the use of the anxiety questionnaires. Sociodemographic and psychopathological variables were analyzed using descriptive statistics (percentages in the case of categorical variables, means and SDs for the continuous variables). Cut-off values indicating an optimal discrimination threshold between 'healthy' and 'diseased' were obtained by ROC analyses. We chose to allow sensitivity and specificity to be equal, taking into account the trade-off between the two [11]. The discriminatory power of the questionnaire (sub) scales was assessed with the associated areas under the ROC curve (AUCs). AUC's over 0.75 were considered clinically useful, with 0.85 showing moderate discriminatory power and 0.95 very high power [52]. The 5^{th}, 25^{th}, 50^{th} (i.e. median), 75^{th}, and 95^{th} percentile scores were calculated. The central 95% of the distribution in reference-groups is commonly used in cases of non-Gaussian distributions [12,53]. The remaining 5% was categorized as 'abnormal' [54]. We chose to categorize the top 5% of the reference-group (95^{th} percentile scores, P_{95}) as 'abnormal' because the lowest 2.5% (functioning 'abnormally' good) cannot be identified in general population samples; the studied anxiety questionnaires merely assess the level of dysfunctionality and not the level of 'health' or normal functionality. Likewise, we regarded the bottom 5% of the patient-group (5^{th} percentile scores, P_5) as indistinguishable from people in the normal range. Furthermore, means and SDs were calculated. Reference values were calculated for all participants combined, as well as for men and women separately. To test our decision not to exclude those individuals in the ROM reference-group with a current psychiatric diagnosis, we performed a sensitivity analysis. The internal consistency of the questionnaires was evaluated using Cronbach's alpha for the total scores and the subscores (with >0.70 indicating adequate internal consistency) [55]. For all analyses, SPSS version 20.0 was used (SPSS Inc, Chicago, Illinois).

RESULTS

Sociodemographic and psychopathological characteristics

The sociodemographic and psychopathological characteristics of the ROM reference-group and patient-group are shown in Table 5.2.

Participants in the ROM reference-group and the ROM patient-group were comparable with respect to mean age and similar with respect to gender distribution. For the ROM reference-group the mean age was 40.3 years (SD=12.6), for the ROM patient-group it was 36.5 years (SD=11.9). Females constituted 62.8% of the reference-group and 64.4% of the patient-group. Those in the ROM reference-group were more often married relative to those in the ROM patient-group and they were less often living alone. Those in the ROM

reference-group were more often married relative to those in the ROM patient-group and they were less often living alone. Those in the ROM reference-group also showed higher levels of education relative to those in the ROM patient-group. Furthermore, work-related disability and unemployment were less prevalent in the ROM reference-group. Fewer participants in the ROM reference-group were of ethnic origin (defined as oneself not being born in the Netherlands or both parents not being born in the Netherlands). Of the ROM reference-group 9.3% had at least one anxiety disorder and 5.2% met criteria for a psychiatric disorder in addition to an anxiety disorder as diagnosed with the MINI-Plus. There was a high rate of psychopathological co-morbidity (i.e., psychopathology in addition to psychopathological anxiety) among participants in the ROM patient-group (55.6%).

REFERENCE VALUES

Percentile scores

Table 5.3 presents the reference values of the eight anxiety questionnaires for the ROM reference-group and the ROM patient-group. For the ROM reference-group, the distribution of each total score and sub score was positively skewed. Mental health was also demonstrated for the ROM reference-group by the substantial percentage of participants (5-25%) having the lowest possible scores (e.g., 5% for the BSA, PAI, SPS, and 25% for the IES-R).

Analyses of gender indicated that both healthy and women with anxiety disorders showed more symptoms of anxiety relative to the men, both in the ROM reference- and ROM patient-groups (see Supplementary Tables 1 through 6).

Table 5.2.: Sociodemographic and psychiatric characteristics of the ROM reference (n=1295) patient (n=5066) groups.

	ROM reference group (n= 1295)		ROM patient group (n=4627)	
Gender: - n (%)				
Male	482	(37.2)	1806	(35.7)
Female	813	(62.8)	3260	(64.4)
Age in years: - mean (± SD)	40.3	(12.6)	36.5	(11.9)
Male	41.2	(12.6)	37.8	(11.9)
Female	39.7	(12.6)	35.8	(11.8)
Marital status[1]: - n (%)				
Married/cohabitating	890	(68.7)	2206	(43.5)
Divorced/separated/widow	78	(6.0)	539	(10.6)
Single	327	(25.3)	1744	(34.4)
Housing situation[1]: - n (%)				
Living alone	201	(15.5)	982	(19.4)
Living with partner	902	(69.7)	2259	(44.6)
Living with family	192	(14.8)	1248	(24.6)
Educational status1,3: - n (%)				
Lower	295	(22.8)	1867	(36.9)
Higher	1000	(77.2)	2619	(51.7)
Employment status[1]: - n (%)				
Employed part-time	509	(39.3)	1033	(20.4)
Employed full-time	554	(42.8)	986	(19.5)
Unemployed/retired	197	(15.2)	1298	(25.6)
Work-related disability	35	(2.7)	1172	(23.1)
Ethnic background[1]: - n (%)				
Dutch	1150	(88.8)	3505	(69.2)
Other ethnicity	145	(11.2)	982	(19.4)
MINI diagnoses: - n (%)				
Currently None	1174	(90.7)	0[2]	
Anxiety disorder (single)	54	(4.2)	2246	(44.3)
Anxiety disorder (comorbidity)	18	(1.4)	2820	(55.6)
Other psychiatric disorder	49	(3.8)	0[2]	

SD denotes standard deviation
[1] Data not available for 128 (2.4%) to 640 (11.8%) of patients
[2] Selection criterion
[3] Lower education: primary or vocational school: Higher education: college or university

Table 5.3: Percentile scores and mean values for Routine Outcome Monitoring anxiety disorder questionnaires in the ROM reference (n=1295) and patient (n=5066) groups.

	ROM reference group						ROM patient group					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
			(n=1291)							(n=4368)		
Brief Scale for Anxiety (BSA)	0	1	3	6	11	3.91 ± 3.92	6	12	16	21	28	16.36 ± 6.78
PADUA Inventory Revised (PI-R)			(n=651)							(n=657)		
Impulses	0	0	0	1	4	0.84 ± 1.70	0	1	4	8	15	5.09 ± 5.18
Washing	0	0	1	3	11	2.27 ± 3.83	0	1	6	17	32	9.86 ± 10.72
Checking	0	1	3	6	12	4.07 ± 4.04	1	8	14	20	26	13.95 ± 7.60
Rumination	1	3	7	11	18	7.71 ± 5.69	10	19	24	29	38	23.87 ± 8.23
Precision	0	0	1	2	6	1.57 ± 2.20	0	3	6	11	18	7.38 ± 5.72
Total	2	7	13	22	43	16.46±13.30	20	40	58	78	106	60.15 ±26.21
Panic Appraisal Inventory (PAI)			(n=630)							(n=1392)		
Anticipated panic	0	1	7	17	37	10.82±12.16	14	32	47	62	82	47.42 ±20.32
Perceived consequences of Panic:												
-Physical	0	0	0	2	17	2.87 ± 6.62	0	7	18	31	44	19.52 ± 13.99
-Social	0	0	0	4	14	3.01 ± 5.72	0	6	14	26	40	16.68 ± 12.79
-Loss of Control	0	0	0	3	17	3.13 ± 5.86	2	9	17	27	40	18.43 ±11.89
-Total	0	0	2	11	47	9.01 ± 15.03	10	31	52	75	108	54.63 ±29.84

Table 5.3: continued.

	ROM reference group						ROM patient group					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Perceived self-efficacy in coping with panic	0	7	21	36	65	24.19±21.30	29	49	62	76	90	61.48 ±18.41
			(n=651)						(n=893)			
Penn State Worry Questionnaire (PSWQ)	22	30	36	47	66	39.52±13.19	**48**	**62**	**69**	**74**	**79**	**66.95 ± 9.92**
Worry Domains Questionnaire (WDQ)			(n=649)						(n=887)			
Relationships	4	4	4	6	9	5.27 ± 1.99	4	7	10	14	18	10.28 ± 4.47
Lack of Confidence	5	5	7	9	14	7.65 ± 3.30	7	12	16	20	24	15.95 ± 5.15
Aimless Future	8	8	10	12	19	11.05 ± 4.00	9	15	21	27	35	21.07 ± 7.79
Work Incompetence	3	3	4	6	9	4.77 ± 1.99	3	6	8	11	14	8.47 ± 3.37
Financial	4	4	5	8	13	6.51 ± 3.04	4	7	10	15	19	10.86 ± 4.92
Health	**6**	**6**	**7**	10	15	8.46 ± 3.31	7	11	14	19	26	15.20 ± 5.94
Total	33	34	39	49	74	43.72±13.62	44	65	81	97	120	81.82 ±23.66
			(n=651)						(n=1231)			
Social Interaction and Anxiety Scale (SIAS)	1	6	10	17	32	12.50 ± 9.34	18	33	44	54	68	43.70 ±14.92
			(n=651)						(n=1237)			
Social Phobia Scale (SPS)	0	2	4	8	19	6.04 ± 6.57	11	22	33	47	64	35.06 ±16.76

Table 5.3: continued.

	ROM reference group						ROM patient group					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Impact of Event Scale – Revised (IES-R)[1]			**(n=1272)**						**(n=390)**			
Intrusions	0	0	1	5	15	3.51 ± 5.03	5	15	20	24	31	19.52 ± 7.36
Avoidance	0	0	0	4	14	2.72 ± 4.80	5	13	18	22	29	17.30 ± 6.96
Hyperarousal	0	0	0	2	8	1.77 ± 3.14	4	11	15	18	22	14.39 ± 5.39
Total	0	0	2	11	36	7.99 ± 11.97	19	43	53	62	78	51.20 ±17.12

SD denotes standard deviation. [1] IES-R scores are sum scores: to yield average scores, divide by number of items

In a sensitivity analysis, we excluded the 9.7% of participants in the ROM reference-group who had a MINI-diagnosis. Among the remaining 1161 participants we found that the median of the changes of the mean scores of the eight anxiety questionnaires was –8% (interquartile range: –5% to –13%). The median of the changes of the P_{95} scores was –9% (interquartile range: –7% to –12%).

To facilitate comparability with the international literature, we also provided means and SDs in Table 5.3. However, we consider these reference values as less valid given that the distributions of all (sub) scores were positively skewed in the ROM reference-group (Figure 1).

Receiver operating characteristic (ROC) curves

Cut-off values, defined by equal sensitivity and specificity, were calculated with ROC analyses (see Table 5.4). The discriminative power of the eight anxiety questionnaires is depicted in Figure 5.1.

ROC analyses, used to discriminate between health and disease, yielded the following cut-off values: 8.5 for the BSA total score, 30,5 for the PI-R total score, 23.5 for the PAI Anticipated Panic subscale score, 21.5 for the total of the PAI Perceived Consequences, and 43.5 for the PAI Perceived Self-efficacy subscale. The cut-off values were as follows: 55.5 for the PSWQ, 55.5 for the WDQ total scale, 24.5 for the SIAS, 14 for the SPS, and 27.5 for the IES-R total scale. AUC values indicated very high discriminatory power for the BSA, the SIAS, the SPS, and the IES-R. Two subscales, PI-R Washing and WDQ Financial, showed clinically useful discriminatory power. All other (sub) scales proved to have moderate discriminatory power. Sensitivity and specificity exceeded 85% for most (sub) scales; for PI-R subscales and WDQ subscales sensitivity and specificity were somewhat lower.

Internal consistency

The internal consistencies of the total scales and subscales of the questionnaires (for all subjects combined) are shown in Table 5.4. The total scales and subscales of all seven self-rating questionnaires showed excellent internal consistencies, with the exception of WDQ subscale Work Incompetence which possessed adequate internal consistency. The internal consistency of the BSA was also adequate.

Table 5.4: Internal consistency and cut-off values in the ROM reference (n=1295) and patient (n=5066) groups for Routine Outcome Monitoring anxiety disorder questionnaires.

	Number of items	Cronbach's Alpha	N	ROC analysis cut-off	Area under the Curve	Sensitivity / specificity (%)
Brief Scale for Anxiety (BSA)	10	0.78	5659	8.5	0.95	87.7 / 87.8
PADUA Inventory Revised (PI-R)			1308			
Impulses	7	0.84		1.5	0.80	68.7 / 81.3
Washing	10	0.95		1.5	0.73	70.2 / 62.5
Checking	7	0.92		6.5	0.86	79.7 / 77.1
Rumination	11	0.93		13.5	0.94	87.5 / 86.3
Precision	6	0.83		2.5	0.83	75.9 / 77.6
Total	41	0.96		30.5	0.94	86.1 / 86.2
Panic Appraisal Inventory (PAI)			2202			
Anticipated panic	15	0.93		23.5	0.94	87.2 / 86.6
Perceived consequences of panic						
-Physical	5	0.89		4.5	0.89	84.1 / 82.9
-Social	5	0.86		4.5	0.86	80.2 / 78.9
-Loss of Control	5	0.84		6.5	0.90	82.6 / 84.1
-Total	15	0.92		21.5	0.93	86.3 / 86.3
Perceived self-efficacy in coping with panic	15	0.96		43.5	0.90	83.1 / 83.4
Penn State Worry Questionnaire (PSWQ)	16	0.95	1544	55.5	0.93	87.4 / 86.5

Table 5.4: continued

	Number of items	Cronbach's Alpha	N	ROC analysis cut-off	Area under the Curve	Sensitivity / specificity (%)
Worry Domains Questionnaire (WDQ)			1536			
Relationships	4	0.87		6.5	0.85	75.5 / 82.4
Lack of Confidence	5	0.91		10.5	0.90	82.5 / 84.4
Aimless Future	8	0.89		13.5	0.88	80.0 / 80.0
Work Incompetence	3	0.79		5.5	0.82	76.6 / 71.3
Financial	4	0.90		7.5	0.76	69.2 / 72.1
Health	6	0.87		10.5	0.86	75.3 / 80.4
Total	30	0.96		55.5	0.92	85.6 / 85.4
Social Interaction and Anxiety Scale (SIAS)	20	0.96	1882	24.5	0.96	88.9 / 89.2
Social Phobia Scale (SPS)	20	0.96	1888	14.0	0.96	90.0 / 90.0
Impact of Event Scale – Revised (IES-R)			1662			
Intrusions	8	0.96		10.5	0.95	88.2 / 88.4
Avoidance	8	0.94		9.0	0.95	87.7 / 87.9
Hyperarousal	6	0.94		6.5	0.96	92.3 / 91.5
Total	22	0.98		27.5	0.96	91.3 / 91.4

The optimal cut-off derived by the ROC analysis is defined by equal sensitivity and specificity scores

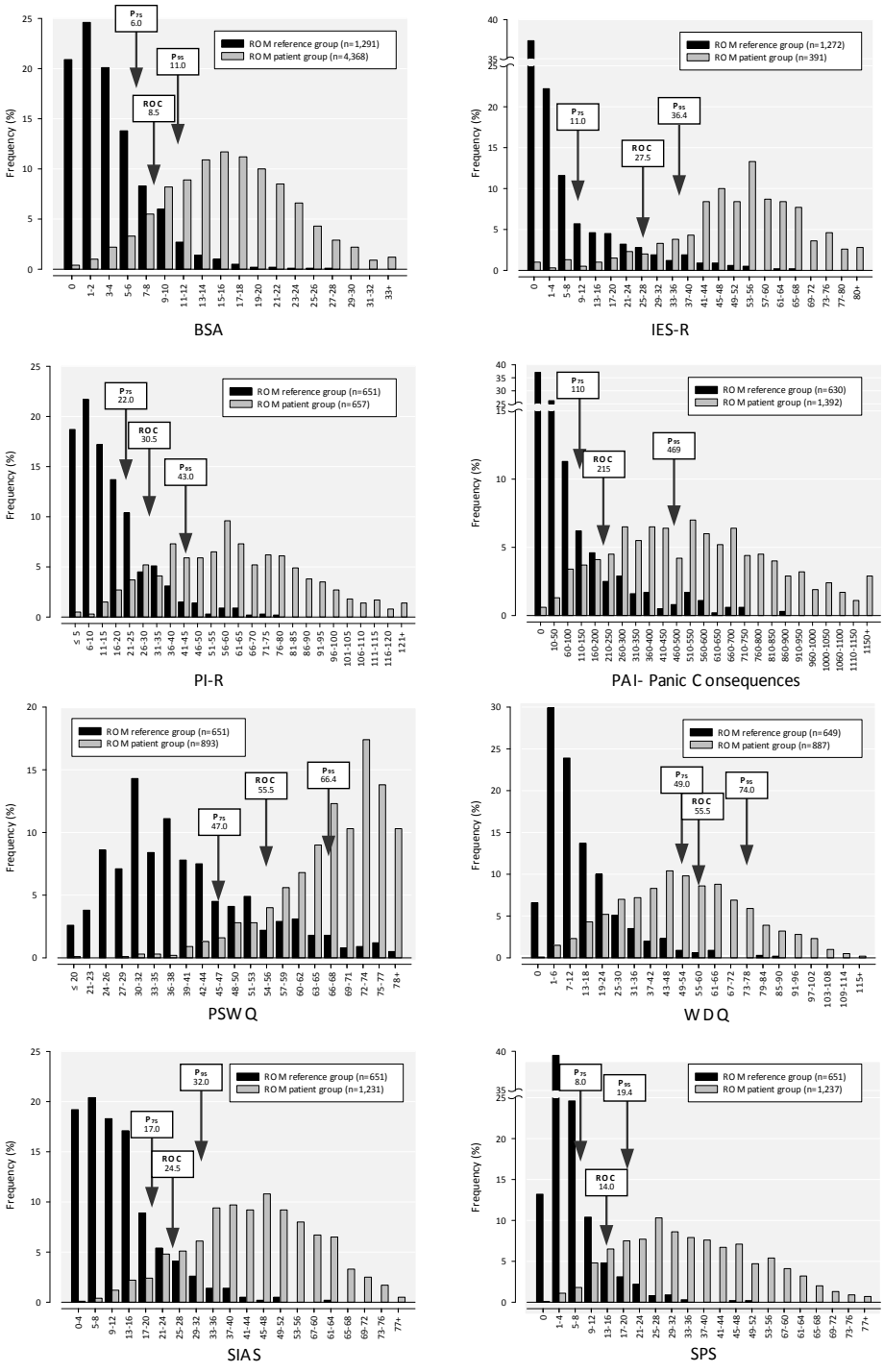


Figure 5.1: Distribution of the scores of the Brief Scale for Anxiety (BSA), PADUA Inventory Revised (PI-R), Panic Appraisal Inventory (PAI) subscale Panic Consequences, Penn State Worry Questionnaire (PSWQ), Worry Domains Questionnaire (WDQ), Social Interaction and Anxiety Scale (SIAS), Social Phobia Scale (SPS), and Impact of Event Scale-Revised (IES-R). Three types of cut-off values are depicted: the 75th percentile score (P_{75}), the 95th percentile score (P_{95}) and the Receiver Operating Characteristics (ROC) cut-off value defined by equal sensitivity and specificity.

DISCUSSION AND CONCLUSION

We reported reference values for a broad range of anxiety questionnaires in two large samples from 'healthy' and 'psychiatrically ill' populations. P_{95} values of the ROM reference-group, cut-off values based on ROC analysis, and P_5 values of the ROM patient-group yielded closely related values. P_{95} values of the ROM reference-group were the highest, ROC values were slightly lower, and P_5 values of the ROM patient-group were the lowest. A pervasive gender-specific pattern in reference values was observed, with higher reference values in women than in men in the ROM reference-group.

The mean PI-R score for our ROM reference-group (M=16.5; SD=13.3) was lower than the mean PI-R scores reported previously, ranging from 21.6 [21] to 37.7 [20]. The mean PSWQ score for the ROM reference-group (M=39.5; SD=13.2) was comparable to the mean PSWQ scores reported by other researchers, ranging from 34.9 to 49.5 [23-28,56], suggesting that our reference-group showed normal levels of pathological worry. The mean WDQ score for the ROM reference-group (M=43.7; SD=13.6) was slightly higher compared to the mean WDQ scores reported in the literature, where it ranged from 24.8 to 38.1 [26-28,30]. This could be explained by only a few participants in our reference-group that showed a high level of non-pathological worry, within the positively skewed distribution. For the ROM reference-group the mean SIAS score (M=12.5; SD=9.3) was slightly lower than the mean SIAS scores reported in other studies, ranging from 14.3 to 19.9 [22,31,32]. The mean SPS score for the ROM reference-group (M=6.0; SD=6.6) was slightly lower than the mean SPS scores reported in literature, ranging from 6.3 to 14.4 [22,31,32]. For the ROM reference-group the mean IES-R score (M=8.0; SD=12.0) was much lower than the mean IES-R score reported by Creamer et al. (M=40.0; SD=23.1) [33]. In sum, the mean scores for our ROM reference-group tended to be lower than the mean scores reported by other researchers, suggesting that our reference-group was relatively healthy. It should however be taken into account that the highly skewed distributions precluded a valid comparison of mean values. For the ROM reference-group the mean scores for the PI-R, the PSWQ, and the IES-R were well below the clinical thresholds as used by other researchers [21,23,33]. This indicated no or only mild anxiety, similar to the previous results. The conducted sensitivity analyses showed slightly lower cut-off values for the reference-group with individuals with a current psychiatric diagnosis excluded. However, these individuals were chosen to be included, in order to prevent producing too strict cut-off values. This would lead to fewer patients considered recovered when P_{95} cut-off scores are used. The high internal consistencies of the PI-R, PAI, PSWQ, WDQ, SIAS, SPS, and IES-R are in accordance with previous studies [25,26,32,36,39,44].

There were some notable differences among the previously published and the present reference values. Health perceptions and health problem expressions vary between cultures [57]. Furthermore, there are differences in study design (e.g., mode of questionnaire administration) [58-60], socio-economic status [58,61], physical functionality [61], health

status varying with area of residence [59], or clinical severity [58,61]. Furthermore, different language versions of the same questionnaire have to measure the same underlying construct where all aspects of this construct (e.g., domain, operational mode, semantics, and psychometric properties) should be similar [60,62]. Two versions of the same questionnaire can be equally sensitive to a given change in functional status yet assign different scores to a given level of distress [61]. Therefore, our reference values should be used with caution in different settings. Further research should evaluate cross-country variability of reference values.

It is noteworthy that a consistent pattern was observed in the 75th and 95th percentile scores of the ROM reference-group, the ROC cut-off values, and the 5th percentile scores of the patient-group. That is, they overlapped considerably, with P_{95} of the ROM reference-group being slightly highest, followed by the ROC cut-off values. The 5th percentile scores of the ROM patient-group had similar values compared to the 75th percentile scores of the ROM reference-group. These values were lower than the 95th percentile scores and ROC cut-off values. This pattern is very similar to the pattern we observed for ROM generic questionnaires [48]. In contrast, for the ROM mood questionnaires the 5th percentile of the ROM patient-group had similar values compared to the 95th percentile of the ROM reference-group [63]. This suggests that there is relatively more subsyndromal anxiety as compared to subsyndromal depression in the ROM reference-group. Mild anxiety may be considered a normal human experience. The ROC cut-off values were rather consistent with the cut-off values derived by other researchers for the PSWQ (55.5 versus 52.3 [23] and for the IES-R (27.5 versus 33 [33].

Furthermore, on average, men from the ROM reference-group scored lower on all eight anxiety scales than did the women from the ROM reference-group. Respectively, for men and women, cut-off (P_{95}) values were 10 and 12 for the BSA, 38 and 44 for the PI-R, 27 and 39 for the PAI Anticipated Panic, 27 and 52 for the PAI Perceived Consequences, 71 and 62 for the PAI Perceived Self-efficacy, 61 and 70 for the PSWQ, 61 and 77 for the WDQ, 27 and 34 for the SIAS, 14 and 22 for the SPS, and 29 and 38 for IES-R. It may be too early to recommend gender-specific reference values because more research is needed in reference populations. Nevertheless, it was striking that reference values from a non-anxious population showed a clinically important gender effect. Most previous studies did not stratify for gender, but those which did [25,26,31,32,43] reported higher means for women than for men, similar to our results.

The results of our study have several clinical implications. The excellent performance of the questionnaires suggests that our reference values are appropriate for various objectives: 1) decisions about treatment termination and referral back to primary care (using the P_{95} of the ROM reference-group); 2) identification of people who may benefit from referral by primary care to specialized mental health care (using the P_5 of the ROM Patient-group), and even 3) diagnostics (using the ROC cut-off values). Regarding diagnostics, these cut-off values might aid in screening for various anxiety disorders, although clinical judgment

and validated diagnostic tools remain the gold standard (e.g., MINI [49,50], Composite International Diagnostic Interview [CIDI; [64]], the Structured Clinical Interview for DSM-III-R [SCID, [65]]). Moreover, cut-off values may be used to classify anxiety. When making decisions about treatment termination or referral to primary care, specificity has to be high [66]. The 75th percentile scores of the ROM reference-group result in few false positives for 'health'. For referral from primary care to specialized mental health care, cut-off values with a high sensitivity are more appropriate, and for that purpose we recommend ROC-based cut-offs or 5th percentile scores from the ROM patient-group because they result in few false positives for 'disease'.

The present study has several strengths. The assessment procedures for both groups were standardized and of high quality (ascertained by training and supervision). Furthermore, the ROM reference-group was large, it was clearly defined, and it resembled the patient-group in all relevant respects (age, gender, level of urbanization) other than those under investigation (i.e., level of psychopathology). The rather precise estimates arising out of the current study are probably attributable to the large sample size. Additionally, the reference-group probably represents the general population quite well. GP registers were used to recruit the reference-group and in the Netherlands the GP registration rate is very high. The ROM patient-group was large as well. Finally, stratification of the ROM reference-group into more homogeneous gender-subgroups may have reduced variation among subgroups, leading to gender-specific reference values, which can be used in clinical practice.

A limitation of the present study includes the relatively high non-response rate in the ROM reference-group, which may have introduced potential selection bias. Additionally, the generalizability of this study is limited by the nature of our ROM reference-group in that it included Dutch-speaking people aged between 18 and 65 years. Reference values may not automatically be applicable to other ethnic groups, to children, and to the elderly. Finally, it is important to recognize that population-based reference values should not be applied rigidly. The choice of cut-off values remains arbitrary and dependent on one's goal (e.g., for confirmation of a diagnosis, specificity should be high and the 95th percentile would be more appropriate than the 75th percentile of the ROM reference-group).

In conclusion, this large-scale population-based study provides reference values and reliability coefficients for the BSA, PI-R, PAI, PSWQ, WDQ, SIAS, SPS, and IES-R. These values increase the utility of these questionnaires, inasmuch as they can be employed as ROM questionnaires to facilitate the assessment of severity of anxiety disorder symptoms. To make responsible decisions about continuing, changing, or terminating therapy, any of these questionnaires can be offered to every patient with MAS disorders. Additionally, these reference values are suitable for indicating which patients have recovered enough to be referred back from specialized mental health care to primary care.

Reference List

1. Baxter AJ, Scott KM, Vos T, Whiteford HA. (2012) Global prevalence of anxiety disorders: a systematic review and meta-regression. *Psychol Med,* 1-14.

2. Bijl RV, Ravelli A, Van Zessen G. (1998) Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc Psychiatry Psychiatr Epidemiol*, 33 (12), 587-595.

3. De Graaf R, Ten Have M, Van Gool C, Van Dorsselaer S. (2012) Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. *Soc Psychiatry Psychiatr Epidemiol,* 47 (2), 203-213.

4. De Vries GJ, Olff M. (2009) The lifetime prevalence of traumatic events and posttraumatic stress disorder in the Netherlands. *J Trauma Stress,* 22 (4), 259-267.

5. De Beurs E, Den Hollander-Gijsman ME, Van Rood YR, Van der Wee NJ, Giltay EJ, Van Noorden MS, Van der Lem R, Van Fenema EM, Zitman FG. (2011) Routine outcome monitoring in the Netherlands: practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clin Psychol Psychother*, 18, 1-12.

6. Schulte-van Maaren YWM, Carlier IVE, Giltay EJ, Van Noorden MS, De Waal MW, Van der Wee NJ, Zitman FG. (2012) Reference values for mental health assessment instruments: objectives and methods of the Leiden Routine Outcome Monitoring Study. *J Eval Clin Pract*.

7. Kazdin AE. (2008) Evidence-based treatment and practice: new opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *Am Psychol*, 63 (3), 146-159.

8. Reed AH, Henry RJ, Mason WB. (1971) Influence of statistical method used on the resulting estimate of normal range. *Clin Chem*, 17 (4), 275-284.

9. Solberg HE. (2008) Establishment and use of reference values. Burtis CA, Ashwood ER, Bruns DE, editors. Fundamentals of clinical chemistry. 6[14], 229-238. St. Louis, Missouri, Saunders Elsevier.

10. Jacobson NS, Truax P. (1991) Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol,* 59 (1), 12-19.

11. Fan J, Upadhye S, Worster A. (2006) Understanding receiver operating characteristic (ROC) curves. *CJEM*, 8 (1), 19-20.

12. Geffre A, Friedrichs K, Harr K, Concordet D, Trumel C, Braun JP. (2009) Reference values: a review. *Vet Clin Pathol*, 38 (3), 288-298.

13. Katayev A, Balciza C, Seccombe DW. (2010) Establishing reference intervals for clinical laboratory test results: is there a better way? *Am J Clin Pathol*, 133 (2), 180-186.

14. Sasse EA, Doumas BT, Miller WG, D'Orazio P, Eckfeldt JH, Evans SA, et al. (2000) How to define and determine reference intervals in the clinical laboratory; approved guideline-Second edition. NCCLS document C28-A2 . 20[13], 1-38. Wayne, PA, NCCLS.

15. Kendall PC, Marrs-Garcia A, Nath SR, Sheldrick RC. (1999) Normative comparisons for the evaluation of clinical significance. *J Consult Clin Psychol,* 67 (3), 285-299.

16. Horn PS, Feng L, Li Y, Pesce AJ. (2001) Effect of outliers and nonhealthy individuals on reference interval estimation. *Clin Chem,* 47 (12), 2137-2145.

17. Wakefield JC. (1992) The concept of mental disorder. On the boundary between biological facts and social values. *Am Psychol*, 47 (3), 373-388.

18. Cohen H. (1981) The evolution of the concept of disease. Caplan AL, Engelhardt HT, McCartney JJ, editors. Concepts of health and disease: Interdisciplinary perspectives. 209-220. Reading, MA, Addison-Wesley.

19. Prusoff BA, Klerman GL, Paykel ES. (1972) Concordance between clinical assessments and patients' self-report in depression. *Arch Gen Psychiatry,* 26 (6), 546-552.

20. Besiroglu L, Yucel AM, Boysan M, Eryonucu B, Gulec M, Selvi Y. (2005) [The assessment of obsessive-compulsive symptoms: the reliability and validity of the Padua inventory in a Turkish population]. *Turk Psikiyatri Derg*, 16 (3), 179-189.

21. Van Oppen P, Hoekstra RJ. (1995) The structure of obsessive-compulsive symptoms. *Behav Res Ther,* 33 (1), 15-23.

22. Brown EJ, Turovsky J, Heimberg RG, Juster HR, Brown TA, Barlow DH. (1997) Validation of the social interaction anxiety scale and the social phobia scale across the anxiety disorders. *Psychological Assessment*, 9 (1), 21-27.

23. Dupuy JB, Beaudoin S, Rheaume J, Ladouceur R, Dugas MJ. (2001) Worry: daily self-report in clinical and non-clinical populations. *Behaviour Research and Therapy*, 39 (10), 1249-1255.

24. Fresco DM, Heimberg RG, Mennin DS, Turk CL. (2002) Confirmatory factor analysis of the Penn State Worry Questionnaire. *Behav Res Ther*, 40 (3), 313-323.

25. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. (1990) Development and validation of the Penn State Worry Questionnaire. *Behav Res Ther*, 28 (6), 487-495.

26. Robichaud M, Dugas MJ, Conway M. (2003) Gender differences in worry and associated cognitive-behavioral variables. *J Anxiety Disord,* 17 (5), 501-516.

27. Stober J. (1998) Reliability and validity of two widely-used worry questionnaires: Self-report and self-peer convergence. *Personality and Individual Differences*, 24 (6), 887-890.

28. Van Rijsoort S, Vervaeke G. (1999) The Penn State Worry Questionnaire and the Worry Domains Questionnaire: structure, reliability and validity. *Clinical Psychology & Psychotherapy,* 6 (4), 297-307.

29. Zlomke KR. (2009) Psychometric properties of internet administered versions of Penn State Worry Questionnaire (PSWQ) and Depression, Anxiety, and Stress Scale (DASS). *Computers in Human Behavior*, 25 (4), 841-843.

30. Verkuil B, Brosschot JF, Thayer JF. (2007) Capturing worry in daily life: are trait questionnaires sufficient? *Behav Res Ther*, 45 (8), 1835-1844.

31. Heimberg RG, Mueller GP, Holt CS, Hope DA, Liebowitz MR. (1992) 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.

32. Mattick RP, Clarke JC. (1998) Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behav Res Ther,* 36 (4), 455-470.

33. Creamer M, Bell R, Failla S. (2003) Psychometric properties of the Impact of Event Scale - Revised. *Behav Res Ther,* 41 (12), 1489-1496.

34. Mattila-Evenden M, Svanborg P, Gustavsson P, Asberg M. (1996) Determinants of self-rating and expert rating concordance in psychiatric out-patients, using the affective subscales of the CPRS. *Acta Psychiatr Scand*, 94 (6), 386-396.

35. Tyrer P, Owen RT, Cicchetti DV. (1984) The brief scale for anxiety: a subdivision of the comprehensive psychopathological rating scale. *J Neurol Neurosurg Psychiatry,* 47 (9), 970-975.

36. Anholt GE, Van Oppen P, Cath DC, Smit JH, Van Dyck R, Van Balkom AJLM. (2009) Measuring obsessive-compulsive symptoms: Padua Inventory-Revised vs. Yale-Brown Obsessive Compulsive Scale. *J Anxiety Disord*, 23 (6), 830-835.

37. De Beurs E, Smit JH, Comijs HC. (2005) Kort Instrumenteel: De Paniek Opinie Lijst (POL). De Betrouwbaarheid en validiteit van een cognitieve maat voor paniekstoornis. *Tijdschrift voor Gedragstherapie*, 2, 141-154.

38. Feske U, De Beurs E. (1997) The Panic Appraisal Inventory: psychometric properties. *Behav Res Ther,* 35 (9), 875-882.

39. Telch MJ, Brouillard M, Telch CF, Agras WS, Taylor CB. (1989) Role of cognitive appraisal in panic-related avoidance. *Behav Res Ther*, 27 (4), 373-383.

40. Kerkhof A, Hermans D, Figee A, Laeremans I, Pieters G, Aardema A. (2000) De Penn State Worry Questionnaire en de Worry Domains Questionnaire: eerste resultaten bij Nederlandse en Vlaamse klinische en poliklinische populaties. *Tijdschrift voor Gedragstherapie,* 2, 135-154.

41. Beck JG, Grant DM, Read JP, Clapp JD, Coffey SF, Miller LM, Palyo SA. (2008) The impact of event scale-revised: psychometric properties in a sample of motor vehicle accident survivors. *J Anxiety Disord*, 22 (2), 187-198.

42. Morina N, Bohme HF, Ajdukovic D, et al. (2010) The structure of post-traumatic stress symptoms in survivors of war: confirmatory factor analyses of the Impact of Event Scale--revised. *J Anxiety Disord,* 24 (6), 606-611.

43. Mystakidou K, Tsilika E, Parpa E, Galanos A, Vlahos L. (2007) Psychometric properties of the Impact of Event Scale in Greek cancer patients. *J Pain Symptom Manage,* 33 (4), 454-461.

44. Weiss DS, Marmar CR. (1997) The Impact of Event Scale-Revised. Wilson JP, Keane TM, editors. Assessing psychological trauma and PTSD: A handbook for practitioners. 399-411. New York, Guildford Press.

45. Sondergaard HP, Ekblad S, Theorell T. (2003) Screening for post-traumatic stress disorder among refugees in Stockholm. *Nord J Psychiatry,* 57 (3), 185-189.

46. Van Noorden MS, Giltay EJ, Den Hollander-Gijsman ME, Van der Wee NJ, Van Veen T, Zitman FG. (2010) Gender differences in clinical characteristics in a naturalistic sample of depressive outpatients: the Leiden Routine Outcome Monitoring Study. *J Affect Disord*, 125 (1-3), 116-123.

47. Poortvliet MC, Lamkadden M, Deville W. Niet op naam ingeschreven (NONI) bij de huisarts. Inventarisatie en gevolgen voor de ziekenfondsverzekerden. Utrecht: NIVEL; 2005.

48. Schulte-van Maaren YWM, Carlier IVE, Zitman FG, Hemert AM, De Waal MW, Van Noorden MS, Giltay EJ. (2012) Reference values for generic instruments used in Routine Outcome Monitoring: the Leiden Routine Outcome Monitoring Study (in press). *BMC Psychiatry.*

49. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC. (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*, 59 Suppl 20, 22-33.

50. Van Vliet IM, De Beurs E. (2007) The MINI-International Neuropsychiatric Interview. A brief structured diagnostic psychiatric interview for DSM-IV en ICD-10 psychiatric disorders. *Tijdschr Psychiatr,* 49 (6), 393-397.

51. Carlier IVE, Meuldijk D, Van Vliet IM, Van Fenema EM, Van der Wee NJ, Zitman FG. (2012) Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *J Eval Clin Pract,* 18 (1), 104-110.

52. Barnabei L, Marazia S, De CR. (2007) Receiver operating characteristic (ROC) curves and the definition of threshold levels to diagnose coronary artery disease on electrocardiographic stress testing. Part I: The use of ROC curves in diagnostic medicine and electrocardiographic markers of ischaemia. *J Cardiovasc Med (Hagerstown),* 8 (11), 873-881.

53. Solberg HE. (1989) Reference values. *Adv Clin Chem,* 27, 1-79.

54. Campbell WW, Robinson LR. (1993) Deriving reference values in electrodiagnostic medicine. *Muscle Nerve*, 16 (4), 424-428.

55. George D, Mallery P. (2003) SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.). Boston, Allyn & Bacon.

56. Brown TA, Antony MM, Barlow DH. (1992) Psychometric properties of the Penn State Worry Questionnaire in a clinical anxiety disorders sample. *Behav Res Ther*, 30 (1), 33-37.

57. Kleinman A, Eisenberg L, Good B. (1978) Culture, illness, and care: clinical lessons from anthropologic and cross-cultural research. *Ann Intern Med,* 88 (2), 251-258.

58. Alonso J, Ferrer M, Gandek B, et al. (2004) Health-related quality of life associated with chronic conditions in eight countries: results from the International Quality of Life Assessment (IQOLA) Project. *Qual Life Res,* 13 (2), 283-298.

59. Bowling A, Bond M, Jenkinson C, Lamping DL. (1999) Short Form 36 (SF-36) Health Survey questionnaire: which normative data should be used? Comparisons between the norms provided by the Omnibus Survey in Britain, the Health Survey for England and the Oxford Healthy Life Survey. *J Public Health Med,* 21 (3), 255-270.

60. Bullinger M, Anderson R, Cella D, Aaronson N. (1993) Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Qual Life Res*, 2 (6), 451-459.

61. Anderson RT, Aaronson NK, Wilkin D. (1993) Critical review of the international assessments of health-related quality of life. *Qual Life Res*, 2 (6), 369-395.

62. Herdman M, Fox-Rushby J, Badia X. (1998) A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Qual Life Res*, 7 (4), 323-335.

63. Schulte-van Maaren YWM, Carlier IVE, Zitman FG, Van Hemert AM, De Waal MW, Van der Does AJW, Van Noorden MS, Giltay EJ. (2012) Reference values for major depression questionnaires: the Leiden Routine Outcome Monitoring Study. *Journal of Affective Disorders*.

64. World Health Organization. (1990) The Composite International Diagnostic Interview (CIDI). Authorized Core Version 1.0. Geneva, WHO.

65. Spitzer RL, Williams JBW, Gibbon M, First MB. (1988) Structured Clinical Interview for DSM-III-R - Patient Version (SCID-P, 4/I/88). New York, Biometrics Research Department, New York State Psychiatric Institute.

66. Bewick V, Cheek L, Ball J. (2004) Statistics review 13: receiver operating characteristic curves. *Crit Care,* 8 (6), 508-512.

Supplementary Tables

Supplementary Table 5.1: Sociodemographic and psychiatric characteristics of the ROM reference (n=1295) patient (n=5066) groups.

Supplementary Table 5.2: Percentile scores and mean values in the ROM reference (n=1291) and patient (n=4368) groups for the subscales and total score of the Brief Scale for Anxiety (BSA).

Supplementary Table 5.3: Percentile scores and mean values in the ROM reference (n=651) and patient (n=657) groups for the subscales and total score of the PADUA Inventory Revised (PI-R).

Supplementary Table 5.4: Percentile scores and mean values in the ROM reference (n=630) and patient (n=1392) groups for the subscales and total score of the Panic Appraisal Inventory (PAI).

Supplementary Table 5.5: Percentile scores and mean values in the ROM reference (n=651/649) and patient (n=893/887) groups for the subscales and total score of the Penn State Worry Questionnaire (PSWQ) and the Worry Domains Questionnaire (WDQ).

Supplementary Table 5.6: Percentile scores and mean values in the ROM reference (n=651) and patient (n=1231/1237) groups for the subscales and total score of the Social Interaction and Anxiety Scale (SIAS) and the Social Phobia Scale (SPS) and.

Supplementary Table 5.7: Percentile scores and mean values in the ROM reference (n=1272) and patient (n=390) groups for the subscales and total score of the Impact of Events Scale (IES-R).

Supplementary Table 5.1: Sociodemographic and psychiatric characteristics of the ROM reference (n=1295) patient and (n=5066) groups.

	ROM reference group (n=1295)				ROM patient group (n=5066)			
	females		males		females		males	
Gender - n (%)	813	(62.8)	482	(37.2)	3260	(64.6)	1806	(35.7)
Age in years: mean (± SD)	39.7	(12.6)	41.2	(12.6)	35.8	(11.8)	37.8	(11.9)
Marital status[1] - n (%)								
Married/cohabitating	552	(67.9)	338	(70.1)	1455	(44.6)	751	(41.6)
Divorced/separated/widow	59	(7.3)	19	(3.9)	401	(12.3)	138	(7.6)
Single	202	(24.8)	125	(25.9)	1047	(32.1)	697	(38.6)
Housing situation[1] - n (%)								
Living alone	132	(16.2)	69	(14.3)	547	(16.8)	435	(24.1)
Living with partner	560	(68.9)	342	(71.0)	1492	(45.8)	767	(42.5)
Living with family	121	(14.9)	71	(14.7)	864	(26.5)	384	(21.3)
Educational status1,3- n (%)								
Lower	189	(23.2)	106	(22.0)	1226	(37.6)	641	(35.5)
Higher	624	(76.8)	376	(78.0)	1676	(51.4)	943	(52.2)
Employment status[1]:- n (%)								
Employed part-time	428	(52.6)	81	(16.8)	854	(26.2)	179	(9.9)
Employed full-time	222	(27.3)	332	(68.9)	402	(12.3)	584	(32.3)
Unemployed/retired	140	(17.2)	57	(11.8)	909	(27.9)	389	(21.5)
Work-related disability	23	(2.8)	12	(2.5)	738	(22.6)	434	(24.0)
Ethnic background[1]: - n(%)								
Dutch	710	(87.3)	440	(91.3)	2259	(69.3)	1246	(69.0)
Other ethnicity	103	(12.7)	42	(8.7)	642	(19.7)	340	(18.8)
MINI diagnoses: - n (%)								
Currently None	723	(88.9)	451	(93.6)	0***		0***	
Anxiety disorder (single)	42	(5.2)	12	(2.5)	1446	(44.4)	800	(44.3)
Anxiety disorder (comorbidity)	15	(1.8)	3	(0.6)	1814	(55.6)	1006	(55.7)
Other psychiatric disorder (without anxiety)	32	(3.9)	16	(3.3)	0***		0***	

*Data not available for 128 (2.4%) to 640 (11.8%) of patients
** Lower education: primary or vocational school; Higher education: college or university
***Selection criterion

Supplementary Table 5.2: Percentile scores and mean values in the ROM reference (n=1291) and patient (n=4368) groups for the subscales and total score of the Brief Scale for Anxiety (BSA).

	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Brief Scale for Anxiety (BSA)												
All participants	0	1	3	6	11	3.91 ± 3.92	6	12	16	21	28	16.36 ± 6.78
- Men	0	0	2	4	10	2.99 ± 3.23	6	11	16	21	29	16.01 ± 6.97
- Women	0	1	4	7	12	4.45 ± 4.19	6	12	16	21	28	16.56 ± 6.67

SD denotes standard deviation.

Supplementary Table 5.3: Percentile scores and mean values in the ROM reference (n=651) and patient (n=657) groups for the subscales and total score of the PADUA Inventory Revised (PI-R).

	ROM reference group (n=651)						ROM patient group (n=657)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
PADUA Inventory Revised (PI-R)												
Impulses												
All participants	0	0	0	1	4	0.84 ± 1.70	0	1	4	8	15	5.09 ± 5.18
- Men	0	0	0	1	4	0.83 ± 1.45	0	2	5	8	17	5.87 ± 5.40
- Women	0	0	0	1	5	0.84 ± 1.83	0	1	3	7	15	4.63 ± 5.00
Washing												
All participants	0	0	1	3	11	2.27 ± 3.83	0	1	6	17	32	9.86 ± 10.72
- Men	0	0	0	2	9	1.72 ± 2.99	0	0	4	13	31	8.41 ± 9.88
- Women	0	0	1	3	13	2.60 ± 4.22	0	1	7	17	34	10.73 ± 11.12
Checking												
All participants	0	1	3	6	12	4.07 ± 4.04	1	8	14	20	26	13.95 ± 7.60
- Men	0	1	3	6	12	4.18 ± 4.06	2	9	15	21	26	14.85 ± 7.29
- Women	0	1	3	6	12	4.01 ± 4.04	1	7	13	19	26	13.41 ± 7.74
Rumination												
All participants	1	3	7	11	18	7.71 ± 5.69	10	19	24	29	38	23.87 ± 8.23
- Men	0	3	6	9	16	6.38 ± 4.80	10	18	24	30	38	23.84 ± 8.24
- Women	1	4	7	12	20	8.50 ± 6.02	10	19	24	29	38	23.89 ± 8.24
Precision												
All participants	0	0	1	2	6	1.57 ± 2.20	0	3	6	11	18	7.38 ± 5.72
- Men	0	0	1	2	6	1.56 ± 2.24	0	2	6	11	17	6.83 ± 5.48
- Women	0	0	1	2	6	1.58 ± 2.17	0	3	7	12	19	7.71 ± 5.85
Total												
All participants	2	7	13	22	43	16.46 ±13.30	20	40	58	78	106	60.15 ± 26.21
- Men	1	6	12	20	38	14.66 ±12.36	21	41	58	77	103	59.81 ± 25.65
- Women	2	8	14	24	44	17.54 ±13.73	20	39	58	79	107	60.36 ± 26.57

Supplementary Table 5.4. Percentile scores and mean values in the ROM reference (n=630) and patient (n=1392) groups for the subscales and total score of the Panic Appraisal Inventory (PAI).

	ROM reference group (n=630)						ROM patient group (n=1392)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Panic Appraisal Inventory (PAI)												
Anticipated panic												
All participants	0	1	7	17	37	10.82 ±12.16	14	32	47	62	82	47.42 ± 20.32
- Men	0	1	5	14	27	8.72 ± 10.34	13	30	44	58	79	44.55 ± 19.94
- Women	0	2	7	19	39	12.05 ±12.98	15	33	49	64	84	49.13 ± 20.36
Perceived consequences of panic - social												
All participants	0	0	0	2	17	2.87 ± 6.62	0	7	18	31	44	19.52 ± 13.99
- Men	0	0	0	1	9	1.91 ± 5.23	0	7	17	30	43	18.81 ± 13.79
- Women	0	0	0	3	21	3.43 ± 7.25	0	8	18	31	45	19.95 ± 14.10
Perceived consequences of panic - social												
All participants	0	0	0	4	14	3.01 ± 5.72	0	6	14	26	40	16.68 ± 12.79
- Men	0	0	0	3	11	2.46 ± 4.26	0	6	13	25	37	15.63 ± 12.17
- Women	0	0	0	4	17	3.34 ± 6.41	0	6	15	27	41	17.31 ± 13.11
Perceived consequences of panic - Loss of control												
All participants	0	0	0	3	17	3.13 ± 5.86	2	9	17	27	40	18.43 ± 11.89
- Men	0	0	0	2	11	2.04 ± 4.01	1	8	16	25	38	17.08 ± 11.39
- Women	0	0	0	5	19	3.78 ± 6.64	2	10	18	28	41	19.22 ± 12.11

Supplementary Table 5.4: continued

	ROM reference group (n=630)						ROM patient group (n=1392)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Perceived consequences of panic - Total												
All participants	0	0	2	11	47	9.01 ± 15.03	10	31	52	75	108	54.63 ± 29.84
- Men	0	0	2	8	27	6.40 ± 11.27	11	29	50	70	104	51.52 ± 28.00
- Women	0	0	2	14	52	10.55 ± 16.68	10	33	54	78	111	56.47 ± 30.75
Perceived self-efficacy in coping with panic												
All participants	0	7	21	36	65	24.19 ± 21.30	29	49	62	76	90	61.48 ± 18.41
- Men	0	5	19	34	71	23.68 ± 23.46	30	46	60	74	88	59.67 ± 18.09
- Women	0	7	23	37	62	24.49 ± 19.94	29	50	63	77	90	62.55 ± 18.52

SD denotes standard deviation.

Supplementary Table 5.5: Percentile scores and mean values in the ROM reference (n=651/649) and patient (n=893/887) groups for the subscales and total score of the Penn State Worry Questionnaire (PSWQ) and the Worry Domains Questionnaire (WDQ).

	ROM reference group						ROM patient group					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Penn State Worry Questionnaire (PSWQ)			n=651							n=893		
All participants	22	30	36	47	66	39.52 ± 13.19	48	62	69	74	79	66.95 ± 9.92
- Men	20	27	32	41	61	35.07 ± 11.15	44	59	67	73	78	65.08 ± 10.25
- Women	24	32	40	51	70	42.17 ± 13.60	48	64	70	75	80	68.05 ± 9.56
Worry Domains Questionnaire (WDQ)			n=649							n=887		
Relationships												
All participants	4	4	4	6	9	5.27 ± 1.99	4	7	10	14	18	10.28 ± 4.47
- Men	4	4	4	5	8	4.85 ± 1.32	4	6	8.5	12	17	9.22 ± 4.11
- Women	4	4	5	6	10	5.53 ± 2.26	4	7	10	14	19	10.89 ± 4.56
Lack of Confidence												
All participants	5	5	7	9	14	7.65 ± 3.30	7	12	16	20	24	15.95 ± 5.15
- Men	5	5	6	8	12	6.65 ± 2.34	6	12	15	19	23	15.21 ± 4.93
- Women	5	5	7	10	16	8.25 ± 3.63	7	13	17	21	24	16.39 ± 5.23
Aimless Future												
All participants	8	8	10	12	19	11.05 ± 4.00	9	15	21	27	35	21.07 ± 7.79
- Men	8	8	9	12	16	10.58 ± 3.13	9	16	22	27	35	21.70 ± 7.55
- Women	8	8	10	13	20	11.34 ± 4.41	9	14	20	27	35	20.70 ± 7.91
Work Incompetence												
All participants	3	3	4	6	9	4.77 ± 1.99	3	6	8	11	14	8.47 ± 3.37
- Men	3	3	4	5	8	4.46 ± 1.56	3	5	8	11	14	8.27 ± 3.34
- Women	3	3	4	6	10	4.96 ± 2.18	3	6	9	11	14	8.58 ± 3.39

Supplementary Table 5.5: continued

	ROM reference group (n=630)						ROM patient group (n=1392)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Financial												
All participants	4	4	5	8	13	6.51 ± 3.04	4	7	10	15	19	10.86 ± 4.92
- Men	4	4	5	7	11	6.21 ± 2.72	4	7	11	15	19	10.99 ± 4.73
- Women	4	4	6	8	14	6.68 ± 3.20	4	6	10	15	20	10.79 ± 5.02
Health												
All participants	6	6	7	10	15	8.46 ± 3.31	7	11	14	19	26	15.20 ± 5.94
- Men	6	6	7	9	14	8.13 ± 2.91	6	10	14	19	26	14.80 ± 5.65
- Women	6	6	7	10	17	8.66 ± 3.51	7	11	14	20	27	15.43 ± 6.10
Total												
All participants	33	34	39	49	74	43.72 ± 13.62	44	65	81	97	123	81.82 ± 23.66
- Men	30	34	38	45	61	40.88 ± 10.00	44	63	80	96	121	80.19 ± 22.93
- Women	30	35	40	52	77	45.43 ± 15.14	44	65	82	99	126	82.78 ± 24.04

SD denotes standard deviation.

Supplementary Table 5.6. Percentile scores and mean values in the ROM reference (n=651) and patient (n=1231/1237) groups for the subscales and total score of the Social Interaction and Anxiety Scale (SIAS) and the Social Phobia Scale (SPS) and.

	ROM reference group						ROM patient group					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Social Interaction and Anxiety Scale (SIAS)			n=651						**n=1231**	n=1231		
All participants	1	6	10	17	32	12.50 ± 9.34	18	33	44	54	68	43.70 ± 14.92
- Men	1	5	9	15	27	11.22 ± 8.09	16	32	43	52	66	41.81 ± 14.67
- Women	1	6	11	18	34	13.27 ± 9.94	20	34	46	56	69	45.15 ± 14.96
Social Phobia Scale (SPS)			n=651						n=1237	n=1237		
Relationships												
All participants	0	2	4	8	19	6.04 ± 6.57	11	22	33	47	64	35.06 ± 16.76
- Men	0	1	3	7	14	4.54 ± 5.17	10	20	31	45	64	33.24 ± 16.59
- Women	0	2	5	9	22	6.93 ± 7.14	12	24	35	48	65	36.46 ± 16.76

SD denotes standard deviation.

Supplementary Table 5.7: Percentile scores and mean values in the ROM reference (n=1272) and patient (n=390) groups for the subscales and total score of **the Impact of Events Scale (IES-R).**

	ROM reference group (n=630)						ROM patient group (n=1392)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
IES-R												
Intrusions												
All participants	0	0	1	5	15	3.51 ± 5.03	5	15	20	24	31	19.52 ± 7.36
- Men	0	0	1	4	13	2.81 ± 4.22	4	15	20	24	31	18.98 ± 7.30
- Women	0	0	1	6	16	3.92 ± 5.41	5	15	20	25	31	19.73 ± 7.39
Avoidance												
All participants	0	0	0	4	14	2.72 ± 4.80	5	13	18	22	29	17.30 ± 6.96
- Men	0	0	0	3	12	2.25 ± 4.10	5	13	17	21	26	16.83 ± 6.41
- Women	0	0	0	4	14	2.99 ± 5.14	4	13	18	22	29	17.48 ± 7.17
Hyperarousal												
All participants	0	0	0	2	8	1.77 ± 3.14	4	11	15	18	22	14.39 ± 5.39
- Men	0	0	0	2	7	1.52 ± 2.58	5	10	14	18	22	14.08 ± 5.09
- Women	0	0	0	2	10	1.91 ± 3.42	2	11	15	18	23	14.51 ± 5.52
Total												
All participants	0	0	2	11	36	7.99 ± 11.97	19	43	53	62	78	51.20 ±17.12
- Men	0	0	2	9	29	6.59 ± 10.03	20	40	52	61	74	49.89 ±16.24
- Women	0	0	3	13	38	8.82 ± 12.92	19	43	53	62	79	51.72 ±17.46

SD denotes standard deviation.

NormQuest

Reference Values for ROM Instruments and Questionnaires

Chapter 6

Reference values for the Body Image Concern Inventory (BICI), the Whitely Index (WI), and the Checklist Individual Strength (CIS-20R)

Yvonne W.M. Schulte-van Maaren
Erik J.Giltay
Albert M. van Hemert
Frans G. Zitman
Margot W.M. de Waal
Yanda van Rood
Ingrid V.E. Carlier

(Submitted)

ABSTRACT

Background:

The Body Image Concern Inventory (BICI), the Whitely Index (WI), and the Checklist Individual Strength (CIS-20R) are three questionnaires often incorporated in Routine Outcome Monitoring (ROM). Respectively, they assess symptom severity in patients with body dysmorphic disorder (BDD), hypochondriasis, and chronic fatigue syndrome (CFS). We aimed to generate reference values for a healthy population (ROM reference-group) and for a population of patients fulfilling diagnostic criteria for at least one of BDD, hypochondriasis, and CFS (ROM patient-group).

Methods: In the ROM reference-group we included 648 subjects recruited through general practitioners. These subjects were matched for age and sex with 823 psychiatric outpatients in the ROM patient-group. To define limits (i.e., cut-off-values) for one-sided reference intervals (5th percentile [P_5] for ROM patient-group and 95th percentile [P_{95}] for ROM reference-group) the outermost 5% of observations were used. Receiver Operating Characteristics (ROC) analyses were used to yield additional cut-off-values.

Results: Cut-off-values (P_{95} ROM reference-group) were 55 for the BICI, 6 for the WI, and 92 for the CIS-20R. These values differed for men and women, being mostly higher for women. P_5 ROM patient-group assessments and ROC analyses yielded slightly lower reference values. The discriminative power of all three somatoform questionnaires was very high.

Conclusions: For the BICI, WI, and CIS-20R a comprehensive set of reference values was obtained. The reference values may facilitate responsible clinical decision-making with respect to adjusting or terminating therapy, and with respect to referring patients from specialized mental health care to primary care and vice versa.

INTRODUCTION

Somatoform disorders are a group of psychiatric disorders in which the patient experiences physical symptoms that are inconsistent with, or cannot be fully explained by, any underlying general medical or neurological condition. The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) includes the following specific somatoform disorders: Somatization Disorder, Undifferentiated Somatoform Disorder, Conversion Disorder, Pain Disorder, Hypochondriasis, and Body Dysmorphic Disorder (BDD) [1]. Patients with these disorders tend to frequently consult general practitioners (GPs) or medical specialists rather than mental healthcare specialists [2]. In the Netherlands, however, such patients do find their way to specialized mental health care due to the availability of evidenced-based and patient-tailored treatment options. Relevant are the use of the maintenance model during intake and special outpatient clinics within the medical setting. Evidenced-based treatments are available for somatization disorder, some of the undifferentiated somatoform disorders (e.g., chronic fatigue syndrome (CFS) [3] and irritable bowel syndrome [4,5]), some of the pain disorders (e.g., low back pain [6] and fibromyalgia [7]), BDD [8,9], and hypochondriasis [10-13].

Routine Outcome Monitoring (ROM) is a system of routine psychometric assessments at baseline (i.e., pre-treatment) and at regular intervals to monitor patients' progress during treatment. DSM-IV-TR Axis I diagnoses are established using the Mini-International Neuropsychiatric Interview-Plus (MINI-Plus) [14]. Together with generic questionnaires, which are completed by all patients, disorder-specific questionnaires are administered to patients who meet the MINI-Plus criteria for a particular disorder [15,16]. These disorder-specific questionnaires assess the severity of symptoms, in order to facilitate the evaluation of treatment effect and clinical decisions about treatment termination. When symptom severity is equivalent to levels found in the general population, second-line treatment can be terminated and referral back to primary care may be indicated.

ROM instruments used to assess symptom severity for a specific disorder need to have good psychometric properties. Preferably, they are also widely used both in research and clinical settings. The availability of the questionnaires in the public domain is also required, given that they are offered to large numbers of patients on numerous occasions. Questionnaires, which fulfill these criteria, are available for the assessment of BDD, hypochondriasis, and CFS. Respectively, the questionnaires are the Body Image Concern Inventory (BICI) [17], the Whitely Index (WI) [18], and the Checklist Individual Strength (CIS-20R) [19].

Reliable ratings from reference populations are required if the ROM results are used for clinical decisions about continuing, altering or terminating treatment [20]. In the present study reference values were established for the BICI, the WI, and the CIS-20R. This set of questionnaires is particularly relevant because it is not easy to ascertain the severity of BDD, CFS and hypochondriasis, and BDD is not easily diagnosed. Some descriptive statistics

(means and standard deviations [SDs]) have been published for healthy controls (see Table 6.1) [17-19,21-23], but we are not aware of studies reporting clinically useful reference values for these scales when administered in the general population. Additionally, we studied a possible gender effect in the reference values.

METHODS

Participants

The reference values were based on two study samples, namely: 1) the ROM reference-group, a sample from the general population; and 2) the ROM patient-group, a sample of psychiatric outpatients diagnosed with BDD (n=130), hypochondriasis (n=226), or CFS (n=481). The ROM patient-group included participants (n=14) with two or more somatoform disorders.

The ROM reference-group is the reference group included in the 'Leiden Routine Outcome Monitoring Study' [16]. Participants in the 'Leiden Routine Outcome Monitoring Study' were randomly selected from the registration systems of eight GPs in the Leiden region, with the aim of recruiting a representative general population sample[1]. Sufficient mastery of the Dutch language and the ability to complete computerized and written questionnaires were required. The response rate was 37.1%, as described previously [16,25]. In all, 1295 participants were included in the 'Leiden Routine Outcome Monitoring Study' [16,16,25,26]. Because of time and financial constraints, 50% of these participants (n=648) were administered the somatoform questionnaires [16]. This group was aged 18 to 65 years (M=40.0 years; SD=12.6) and 62.5% were females. Given that the aim of this study was to generate reference values that can be used to guide decision-making about the continuation or termination of therapy, we excluded those who received treatment for psychiatric disorders and/or were dependent on alcohol or drugs during the six months prior to assessment. The reference-group was matched for gender and age to the ROM patient-group, to ensure it was demographically comparable.

The ROM patient-group of the 'Leiden Routine Outcome Monitoring Study' consisted of a baseline sample of 7840 psychiatric outpatients. This constituted approximately 80% of the total number of referred patients with a tentative diagnosis of mood-, anxiety- and/or somatoform disorder [27]. Inclusion criteria were the diagnosis of at least one somatoform disorder, according to the MINI-Plus, and an age between 18 and 65 years. A sub-sample of 823 patients fulfilled the criteria (mean age=38.6, SD=11.7), of whom 70.5% were females. Depending on their MINI-Plus diagnosis, patients completed the BICI (n=130), the WI (n=226), or the CIS-20R (n=481). They were treated in the Leiden University Medical Center (LUMC) Department of Psychiatry or the Rivierduinen mental health care centres. At baseline, scores represented the severity of symptoms prior to the first treatment session.

[1]In the Netherlands, 99,9% of the general population is registered with a GP [24].

Procedures

Procedures for the web-based ROM program of the LUMC Department of Psychiatry and mental health care centre Rivierduinen are described in detail elsewhere [15]. In short, the baseline ROM assessments comprised a standardized diagnostic interview (Dutch version of the Mini-International Neuropsychiatric Interview Plus, version 5.00-R: MINI-Plus) [14,28], the collection of sociodemographic data, and the administration of generic and disorder-specific instruments for mood, anxiety, and somatoform disorders. The MINI-Plus was used to establish the presence of Axis I symptoms according to the DSM-IV-TR. Disorder-specific self-rating questionnaires were selected on the basis of the MINI-Plus. Participants in the reference-group were assessed in a similar way to those in the patient-group, except that those in the ROM reference-group completed all three questionnaires whereas the participants of the ROM patient-group only completed those questionnaires relevant to their diagnosed disorder(s). The assessments were performed by specially trained and regularly (i.e., monthly) supervised research nurses in the outpatient clinics.

The general study protocol associated with ROM, in which ROM is administered as part of the routine treatment process for patients, was approved by the Medical Ethical Committee of the LUMC. This comprehensive protocol (titled "Psychiatric Academic Registration Leiden Database") safeguards the anonymity of patients and reference-group participants and ensures proper handling of the ROM data. If patients object to the use of their outcome data for scientific purposes, the data are removed. Participants of the ROM reference-group signed informed consent for the purpose of this study.

Questionnaires

Body Image Concern Inventory (BICI)

The BICI measures concerns about appearance [17]. The 19 self-report items are answered on a 5-point Likert scale (1='never', 5='always') and the total score ranges between 19 and 95. Two factors have been identified. Factor 1 (12 items) relates to dissatisfaction and shame regarding one's appearance. Factor 2 (7 items) relates to interference with functioning due to appearance concerns. Because the two factors are highly correlated, Littleton and colleagues [17] suggested using a single total score. A cut-off-value of 72 has been recommended, such that scores above 72 are regarded as clinical concerning. The time frame for the reported symptoms is the past week. The BICI can be used to assess symptom severity. Previous studies have not yielded percentile scores. Means and standard deviations (SDs) for healthy control groups were previously determined, ranging from 42.8 (SD=15.0) to 50.4 (SD=14.2) [17,21,22]. For a BDD patient-group a mean of 80.1 (SD=9.0) was reported [17]. Reliability, validity, and internal consistency (Cronbach's alpha's [Cα] range from 0.91-0.94) of the English-language version are good [17,21,22], as is the Cα (0.93) for the Dutch version [29].

Whitely Index (WI)

The WI is a self-report questionnaire that assesses the severity of symptoms of hypochondriasis [18]. Scores for the 14 dichotomous items are summed to yield a total score (range 0-14). The WI is unifactorial [30]. The time frame for the symptoms is the past week. Previous studies have not assessed percentile scores but they have reported means and SDs for healthy control groups, which ranged from 1.7 (SD=2.4) to 3.0 (SD=2.5) [18,23]. For hypochondriacal patients the mean scores ranged from 7.6 (SD=3.0) to 8.9 (SD=5.2) [18,23]. Internal consistency ranged from 0.76-0.80), stability, concurrent and discriminative validity are adequate [30].

Checklist Individual Strength (CIS-20R)

The CIS-20R was designed to measure the severity of symptoms typical of CFS [19]. Each item is scored on a 7-point Likert scale (1 = 'yes, that is true'; 7 = 'no, that is not true'). The total score is the sum of all items (range 20-140). The four subscales are Subjective Fatigue (8 items), Concentration (5 items), Motivation (4 items), and Physical Activity (3 items). The time frame for the reported symptoms is the past two weeks. The recommended clinical cut-off-value for the CIS-20R is 35 [19]. No percentile scores have been reported yet. Means and SDs for healthy controls and CVA-patient related controls were 41.5 (SD=19.8) and 50.9 (SD=26.6), respectively [19]. Internal consistencies for the CIS-20R total scale and subscales are very good (0.90 for the total scale; from 0.83 to 0.92 for the subscales) [19,31] and psychometric properties (i.e., reliability and validity) are excellent [19].

Table 6.1: Somatoform questionnaires used in Routine Outcome Monitoring

Questionnaire	Domain	number of items	Rating	Range for score	Our sample sizes:	Range for sample sizes in previous studies Reference/ Patient-group	References
BICI	Body dysmorphic disorder	19	1 = never; 5 = always	19-95	645 / 130*	184-1043 / 71	[17,21, 22,32]
WI	Hypochon-driasis	14	0=no; 1=yes		644 / 226+	15-204 / 100-149	[18,23,30]
				0-14			
CIS-20R	Chronic fatigue		1 = yes, completely right; 7 = no, completely wrong				[19,33]
Subj.fatigue		8		8-56			
Concentration		5		5-35			
Motivation		4		4-28			
Activity		3		3-21			
Total		20		20-140	643 / 481‡	43-53 / 758	

BICI denotes Body Image Concern Inventory; CIS-20R denotes Checklist Individual Strength; WI denotes Whitely Index.
* Patients diagnosed with Body Dysmorphic Disorder (BDD)
+ Patients diagnosed with Hypochondriasis
‡ Patients diagnosed with Chronic Fatigue Syndrome (CFS)

Table 6.1 presents the sample sizes, disorder domains, subscales, ratings, and score ranges for each questionnaire, together with the reference values reported in previous studies.

Statistical analyses

Descriptive statistics were derived for the sociodemographic variables and the psychiatric variables, including means and SDs for the continuous variables, and percentages for the categorical variables.

The internal consistency of the questionnaires was determined using Cronbach's alpha (with >0.70 indicating adequate internal consistency).

The 5th, 25th, 50th, 75th, and 95th percentiles were calculated. They were calculated for the entire ROM reference-group and for the sub-set of the ROM patient-group that completed the BICI, WI, or CIS-20R. Furthermore, percentiles were calculated separately for men and women. To facilitate comparability with the international literature, we also calculated means and SDs, although these reference values are less useful in skewed reference-group distributions [34]. Receiver Operating Characteristics (ROC) analyses were used to derive a cut-off-value for each instrument, indicating a neutral discrimination threshold between 'healthy' and 'diseased'. Sensitivity and specificity were chosen to be equal. In this way, an acceptable compromise was reached between as few false positives as possible, and as few false negatives as possible. The discriminatory power of the questionnaire total scales and subscales was assessed using the associated areas under the ROC curve (AUCs), where AUC values above 0.75 were considered clinically useful, with 0.85 showing moderate discriminatory power, and 0.95 showing very high power [35]. To assess the effects on the reference values of individuals in the ROM reference-group with a current psychiatric diagnosis, we performed a sensitivity analysis in the group while excluding participants with any psychiatric diagnosis.

When reference values are calculated and interpreted, attention needs to be paid to sensitivity and specificity, the definition of health, and required sample sizes. Firstly, in the assessment of cut-off-values, sensitivity (i.e., the proportion of actual positives which are correctly identified as such) and specificity (i.e., the proportion of negatives which are correctly identified) play a key role. The 95th percentile (P_{95}) of the reference-group is recommended as a cut-off-value when considering referral back from secondary to primary care. The specificity to assess health is relatively high. The 5th percentile (P_5) of the patient-group is recommended as a cut-off-value when considering referral from primary to secondary care. In this case, the sensitivity to assess disease is relatively high. The 5th percentile (P_5) of the ROM reference-group is generally lower than the 95th percentile (P_{95}) of the patient-group. Secondly, it is preferable that reference values [36] are established in healthy (normal) populations [34] with normality defined statistically rather than as a medical (ab)normality. This statistical definition of normality is based on the distribution of scores from the general population (including all individuals) [37]. Finally, (sub)sample sizes of at least 120 are needed to reduce the amount of uncertainty and error caused by potential outliers [38,39]. For all analyses, SPSS version 20.0 was used (SPSS Inc, Chicago, Illinois).

RESULTS

Sociodemographic and psychiatric characteristics

The sociodemographic and psychiatric characteristics of the ROM reference-group and patient-group are shown in Table 6.2. Characteristics per gender are given in Supplementary Table 6.1.

The ROM reference-group and patient-group were rather well matched for age (M=40.0 years [SD=12.6] and M= 38.6 years [SD=11.8], respectively) and gender distribution (62.5% females and 70.5% females, respectively). Participants from the ROM reference-group were more often married than those from the ROM patient-group (70.5% versus 47.6%) and less often living alone (13.4% versus 16.8%), had higher levels of education (78.7% higher education versus 50.9%), had less work-related disability and unemployment (17.9% versus 52.2%), and were less often from non-Dutch ethnic origin (defined as oneself or both parents not being born in the Netherlands). In the ROM reference-group 5.0% had at least one somatoform disorder according to the MINI-Plus, compared to 100% of the subjects from the ROM patient-group (inclusion criterion). In the ROM reference-group 0.5% fulfilled criteria for BDD, 0.6% for Hypochondriasis, and 2.0% for Undifferentiated Somatoform Disorder. In the patient-group 15.8% fulfilled criteria for BDD, 27.5% for Hypochondriasis, 58.5% for Undifferentiated Somatoform Disorder (of whom 21.9% CFS), and 1.0% for Somatization Disorder. Comorbid BDD and CFS was seen in 0.7% of the patients, 2.9% had hypochondriasis and CFS, and no patients had BDD and hypochondriasis or three diagnoses. In the ROM patient-group, a high proportion of subjects (53.6%) had a co-morbid mood or anxiety disorder.

Table 6.2: Sociodemographic and psychiatric characteristics of the ROM reference (n=648) and patient (n=823) groups.

	ROM reference group (n= 1295)		ROM patient group (n=4627)	
Gender: - n (%)				
Male	243	(37.5)	243	(29.5)
Female	405	(62.5)	580	(70.5)
Age in years: - mean (± SD)	40.0	(12.6)	38.6	(11.7)
Male	40.8	(12.6)	38.0	(12.2)
Female	39.6	(12.6)	38.9	(11.5)
Marital status[1]: - n (%)				
Married/cohabitating	457	(70.5)	392	(47.6)
Divorced/seperated/widow	34	(5.2)	96	(11.7)
No data available			119	(14.5)
Housing situation[1]: - n (%)				
Living alone	87	(13.4)	138	(16.8)
Living with partner	462	(71.3)	403	(49.0)
Living with family	99	(15.3)	163	(19.8)
No data available			119	(14.5)
Educational status1,3: - n (%)				
Lower	138	(21.3)	285	(34.6)
Higher	510	(78.7)	419	(50.9)
No data available			119	(14.5)
Employment status[1]: - n (%)				
Employed part-time	256	(39.5)	157	(19.1)
Employed full-time	276	(42.6)	117	(14.2)
Unemployed/retired	101	(15.6)	200	(24.3)
Work-related disability	15	(2.3)	230	(27.9)
No data available			14	(1.7)
Ethnic background[1]: - n (%)				
Dutch	569	(87.8)	590	(71.6)
Other ethnicity	79	(12.2)	114	(13.9)
No data available			119	(14.5)
MINI diagnoses: - n (%)				
Currently None	590	(91.0)	02	
Somatoform disorder (single)	25	(3.9)	382	(46.4)
Somatoform disorder (comorbidity)	7	(1.1)	441	(53.6)
Other than somatoform disorder	26	(4.0)	02	

[1] Data not available for 14 (1.7%) to 119 (14.5%) of patients
[2] Selection criterion
[3] Lower education: primary or vocational school; Higher education: college or university

Reference values and internal consistencies

Table 6.3 presents the internal consistencies (Cronbach's alpha Cα) and results of the ROC analyses of the BICI, WI, and CIS-20R (sub-) scales for both the ROM reference-group and the patient-group. Table 6.4 presents the percentile scores and mean scores. Results of gender analyses are presented in Supplementary Tables 6.2 through 6.4.

For the ROM reference-group, the distributions of total scores and subscale scores were strongly positively skewed (Figure 6.1). Apparent health was also demonstrated by the substantial percentage of participants rating the lowest possible scores.

Table 6.3: Internal consistency and cut-off-values in the ROM reference (n=648) and patient (n=823) groups for Routine Outcome Monitoring somatoform disorder questionnaires.
BICI denotes Body Image Concern Inventory; CIS-20R denotes Checklist Individual Strength; WI denotes Whitely Index.

	Nr of items	Cron-bach's Alpha	Nr ref.	Nr of patients	ROC analysis cut off*	Area under Curve	Sensitivity / specificity
BICI	19	0.96	645	130	49.5	0.96	0.90 / 0.90
WI	14	0.90	644	226	5.5	0.98	0.95 / 0.93
CIS-20R			643	481			
Checking	8	0.97			42.5	0.96	0.92 / 0.92
Rumination	5	0.93			17.5	0.89	0.83 / 0.83
Precision	4	0.84			10.5	0.85	0.79 / 0.76
Total	3	0.90			9.5	0.89	0.81 / 0.82
	20	0.97			81.5	0.97	0.92 / 0.92

*The optimal cut-off derived by the ROC analysis is defined by equal sensitivity and specificity

Body Image Concern Inventory (BICI)

The internal consistency of the BICI was excellent (Cα=0.96).

For the ROM reference-group, the P_{95} cut-off-value was 55 for the BICI: this is the recommended cut-off-value for the referral of patients in specialized mental health care back to primary care. The P_5 value for the ROM patient-group was 39, which is the recommended cut-off-value when primary care patients should be referred to specialized mental health care. Stratified analyses according to gender indicated that, on average, healthy women reported more symptoms than men in relation to the BICI. reference values were also higher for women relative to men. ROC analyses yielded a BICI cut-off-value of 49.5. The AUC value indicated very high discriminatory power for the BICI. The discriminative power of the

BICI is depicted in Figure 1.

Whitely Index (WI)

The internal consistency of the WI was excellent (Cα=0.90).
For the ROM reference-group, the WI total score showed a P_{95} value of 6, which is the recommended cut-off-value for referral back to primary care of patients in specialized mental health care. The P_5 value for the ROM patient-group was 5. Again, the P_{95} and mean values were higher among healthy women than among healthy men. ROC analyses yielded a WI cut-off-value of 5.5. The AUC value indicated very high discriminatory power for the WI. The discriminative power of the WI is depicted in Figure 6.1.

Checklist Individual Strength (CIS-20R)

The CIS-20R showed excellent internal consistency (Cα= 0.97).
For the CIS-20R total score, the P_{95} cut-off-value for the ROM reference-group was 92. The cut-off-values for the subscales were as follows: 46 for Subjective Fatigue, 26 for Concentration, 20 for Motivation, and 15 for Activity. The P_5 value for the ROM patient-group was 74 for the total score. The P_5 values for the subscales were 38 for Subjective Fatigue, 6 for Concentration, 4 for Motivation, and 3 for Activity. Once again, stratified analyses according to gender indicated that, on average, healthy women reported more symptoms than did healthy men. However, for the CIS-20R subscale Activity, no gender difference was found. ROC analyses yielded a CIS-20R cut-off-value of 81.5. AUC values indicated very high discriminatory power for the CIS-20R total scale and moderate to very high discriminatory power for the subscales. The discriminative power of the CIS-20R total score is depicted in Figure 6.1.

In a sensitivity analysis, all 58 (9%) participants with any MINI-diagnosis were excluded from the ROM reference-group. Among the remaining 590 participants we found that the median of the changes of the mean scores of the three somatoform questionnaires decreased by 5% (interquartile range -4 to -5%). The median of the changes of the P_{95} scores decreased by 7% (interquartile range -4 to -8%). Thus, the inclusion of (non-healthy) participants with symptoms led to slightly higher reference values relative to reference values for a 'supernormal' (i.e., overly healthy) reference-group.

Table 6.4: Percentiles and mean values for Routine Outcome Monitoring somatoform disorder questionnaires in the ROM reference (n=648) and patient (n=823) groups.

	ROM reference group						ROM patient group					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
			(n=645)						n= 130			
BICI	20	26	33	41	55	34.4 ± 10.8	39	60	72	78	87	68.4 ± 13.9
			n= 644						n= 226			
WI	**0**	**1**	**2**	**3**	**6**	**2.2 ± 2.0**	**5**	**8**	**10**	**12**	**13**	**9.8 ± 2.5**
CIS-20R			n= 643						n= 481			
Subjective fatigue	8	9	16	27	46	20.2 ± 12.6	38	50	54	56	56	51.5 ± 7.0
Concentration	5	5	9	14	26	11.0 ± 7.0	6	22	29	35	35	26.7 ± 8.9
Motivation	4	4	7	10	20	8.4 ± 5.1	4	12	17	24	28	17.8 ± 7.3
Activity	3	3	4	9	15	6.2 ± 4.3	3	12	18	21	21	16.0 ± 5.6
Total	20	28	38	60	92	45.8 ± 23.3	74	100	116	128	140	112.0 ± 20.4

ROM denotes Routine outcome monitoring. SD denotes standard deviation.
BICI denotes Body Image Concern Inventory; WI denotes Whitely Index; CIS-20R denotes Checklist Individual Strength.

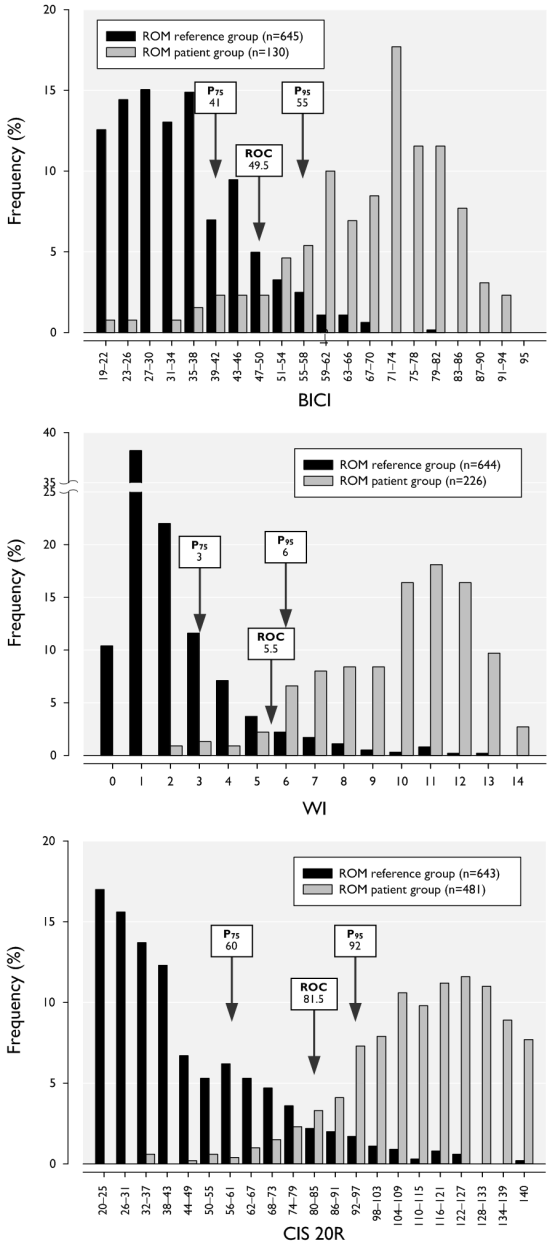


Figure 6.1: Distribution of the scores of the Body Image Concern Inventory (BICI), the Checklist Individual Strength (CIS-20R), and the Whitely Index (WI). Three types of cut-off-values are depicted: the 75th percentile score (P_{75}), the 95th percentile score (P_{95}) and the Receiver Operating Characteristics (ROC) cut-off-value defined by equal sensitivity and specificity.

DISCUSSION AND CONCLUSION

The aim of this study was to determine reference values for the BICI, WI, and CIS-20R based on data from a large sample of 'healthy' participants (defined as not being in specialized mental health care treatment for a psychiatric disorder) and a large 'psychiatrically ill' population. Two clinically relevant types of cut-off-values were generated: the 95th percentile of the ROM reference-group and the 5th percentile of the ROM patient-group. We also derived an additional set of percentile scores and ROC-based cut-off-values for both the ROM reference-group and the patient-group. A gender-specific pattern in reference values was observed for the total scores of all three questionnaires, but not for the CIS-20R subscales. We therefore consider gender-specific reference values to be of clinical relevance for these somatoform questionnaires.

The prevalence rate of any somatoform disorder in the ROM reference-group (5.0%) was comparable to the 4-week prevalence rate (7.5%) in the German general population [40]. As could be expected, the point prevalence rate of BDD in the ROM reference-group (0.5%) was slightly lower than the previously reported 1-year prevalence rates ranging from 0.7 to 2.4 [41-44]. Similarly, the point prevalence rate of hypochondriasis in the ROM reference-group (0.6%) was slightly lower than a previously reported 1-year prevalence rate (4.5%) [41]. The prevalence rate for undifferentiated somatoform disorder in our reference-group was 2.0%, compared with 13.0% in the Dutch treatment-seeking population of De Waal [45]. Our GP population was not necessarily a consulting (i.e., treatment seeking) population. Rather, it was selected from the GP-registration system. The fact that people tend to visit their GP when they have complaints, and that many of these complaints can be classified as undifferentiated somatoform disorder, might explain the large difference in prevalence rate in both studies. Furthermore, it is indeed possible that the MINI-Plus under diagnosed somatoform and other disorders.

The internal consistency of the BICI (0.96) is in accordance with previous studies [17,21,22,29]. The cut-off-values reported in this study (50, P_{95} reference-group; 49.5, ROC based cut-off-value; 39, P_5 patient-group) are substantially lower than the BICI cut-off-value of 72 reported by Littleton and colleagues [17]. This may be explained by Littleton's use of a sample of college students (80% females), where body image concerns appear to be more common [46]. Moreover, they were younger than our reference-group and patient-group and younger people have more body image concerns than older people [47]. The mean BICI score for our ROM reference-group (34) was similar to the mean BICI scores reported by Littleton and colleagues (32 to 43) [21], and lower than the means reported in other studies, ranging from 43 [22] to 50 [17]. This suggests that our reference-group was relatively healthy. However, consideration should be given to the fact that the comparison of mean values of variables with skewed distributions may reflect the strong impact of a few outliers. The internal consistency (α=0.90) of the WI is in accordance with a previous study [23]. The different types of cut-off-values reported in the current study (6, P_{95} reference-group; 5.5,

ROC based cut-off-value; 5, P_5 patient-group) were very similar. To our knowledge, no cut-off-values have previously been reported. The mean WI score of 2.2 for the ROM reference-group was comparable to the mean WI scores reported by Pilowsky (i.e., 1.7 for normal controls) [18]. The mean WI score of 9.8 for our patient-group (10) was very similar to the mean of 8.0 reported by Speckens and colleagues [23] and the mean of 8.5 as reported by Pilowsky [18].

The internal consistency (α=0.97) of the CIS-20R is in accordance with previous studies [19,31]. Vercoulen and colleagues [19] reported decile scores. The P_{50} value for our ROM reference-group of 38 is very close to Vercoulen's P_{50} values of 35 for healthy controls and 42 for controls who are related to somatic (CVA) patients. The mean CIS-20R total score was 46 for the ROM reference-group. The mean Subjective Fatigue score was 20, well below the cut-off of 35 for this subscale [19]. By contrast, the somatoform patient-group had a mean total score of 112 and a mean Subjective Fatigue score of 52. This latter score is well above the cut-off of 35, indicating psychopathology, as was expected.

Gender-effects were analyzed. For the BICI percentile scores were lower for men than for women: e.g., cut-off (P_{95}) values were 45 for men and 57 for women. Luca and colleagues (2011;[22]) found a similar gender effect in their healthy Italian sample [22]. WI data showed that for the ROM reference-group, P_{95} cut-off-values were 5 for men and 7 for women. However, most reference values were equal or close to equal for men and women, both in the reference-group and the patient-group. Pilowski and colleagues (1967;[18]) did not test gender-effects in their healthy sample, but they also reported slightly less symptoms for male non-psychiatric cancer patients compared to their female counterparts [18]. A Dutch study reported no gender differences [30]. Regarding the CIS-20R, ROM reference-group P_{95} cut-off-values for the total score were 89 for men and 97 for women. However, no general gender effect was observed for the subscales. In the ROM reference-group, men reported slightly lower Subjective Fatigue than women, but there was no significant gender-effect for the Concentration, Motivation, and Activation subscales. The developers of the questionnaire found no significant gender-effect [19]. So, at this moment there is not enough evidence to recommend gender specific reference values for the BICI, the WI, nor the CIS-20R.

The excellent (illness-health) differentiating performance of the BICI, WI, and CIS-20R implies that the reference values can be used by clinicians in specialized mental health care to test whether their patient has recovered. Also, the reference values can be used by clinicians in primary care to assess whether referral to specialized mental health care is warranted. Regarding the first point about making decisions about treatment termination, specificity for the assessment of health has to be high. (This contrasts with the normal concept of specificity, which is generally used when ascertaining disorders or dysfunction.) If a treated patient in specialized mental health care displays symptom severity that is equivalent to levels found in the general population, termination of treatment is warranted and referral back to primary care is indicated. The remaining (subsyndromal) symptoms generally do not require specialized treatment anymore.

require specialized treatment anymore. The clinical threshold would be the 95th percentile score (P_{95}) of the reference population (i.e., this results in few false positives). Regarding the second point, referral from primary care to specialized mental health care requires high sensitivity for ascertaining somatoform disorders. The GP has to decide whether the symptoms are so severe that they are equivalent to levels found in the psychiatrically ill population. So, the 5th percentile score (P_5) of the patient population would be the clinical threshold. Severity measures for the BDD, hypochondriasis, and CFS are particularly relevant because these disorders are common but are often unrecognized [48-50].

The present study has several merits. Firstly, the ROM reference-group consisted of individuals without any psychopathological symptoms as well as individuals with psychopathology symptoms who were not receiving treatment in specialized mental health care. In this way, a non-realistic 'supernormal' (i.e., too healthy) reference-group [38] was avoided. This criterion is relevant when the reference values are used to make decisions about the continuation or termination of treatment. It is not necessary that the patient is symptom free; treatment can also be terminated if symptoms have reached a level for which no more specialized care is needed. Secondly, the size of the ROM reference-group sample was large (more than 600 cases). Moreover, the reference-group was clearly defined and it resembled the patient-group in relevant respects (age, gender, level of urbanization). Therefore, our reference values had rather good precision. Thirdly, the ROM reference-group likely represents the general population quite well, because of the very high GP registration rate in the Netherlands. Finally, the assessment procedures for both groups were standardized and of high quality (achieved by training and supervision).

The results should be interpreted in the light of some limitations. Firstly, of the persons (non-consulting GP patients) approached, 63.2% did not want to participate in the study [16]. This large non-response might be due to the extensiveness of the interview. The total time involved was 3 hours which were unpaid. The non-response rate implies potential selection bias, which may have resulted in slightly different (higher or lower) percentile and cut-off-values. Secondly, the patient samples completing the BICI and WI were relatively small in size compared to the reference-group, but they were nevertheless larger than 120. Thirdly, given that our ROM reference-group was aged between 18 and 65 years, Dutch, and taken from a sample of Leiden area GP's, reference values may not necessarily be applicable to children, the elderly, or other ethnic or cultural groups. Fourthly, some demographic data were not collected for about 15% of the ROM patient-group. Finally, only a selection of questionnaires was studied, and thus not every somatoform disorder was investigated.

In conclusion, this large-scale population-based study provides reference values for the BICI, WI, and CIS-20R. This helps improve their usability as ROM questionnaires to differentiate between clinically relevant conditions and normal conditions. These reference values facilitate clinical decisions regarding the continuation, adjustment, or termination of treatment. Additionally, the values allow for the identification of patients in specialized mental health care that have recovered enough in order to be referred back to primary care. Finally, the reference values allow also for the identification of primary care patients that may benefit from specialized mental health care.

Reference List

1. American Psychiatric Association. (1994) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Washington,DC, American Psychiatric Association.

2. Hiller W, Rief W. (2005) Why DSM-III was right to introduce the concept of somatoform disorders. *Psychosomatics,* 46 (2), 105-108.

3. Reme SE, Archer N, Chalder T. (2012) Experiences of young people who have undergone the Lightning Process to treat chronic fatigue syndrome/myalgic encephalomyelitis - a qualitative study. *Br J Health Psychol*.

4. Ford AC, Talley NJ, Schoenfeld PS, Quigley EM, Moayyedi P. (2009) Efficacy of antidepressants and psychological therapies in irritable bowel syndrome: systematic review and meta-analysis. *Gut,* 58 (3), 367-378.

5. Zijdenbos IL, De Wit NJ, Van der Heijden GJ, Rubin G, Quartero AO. (2009) Psychological treatments for the management of irritable bowel syndrome. *Cochrane Database Syst Rev* (1), CD006442.

6. Hoffman BM, Papas RK, Chatkoff DK, Kerns RD. (2007) Meta-analysis of psychological interventions for chronic low back pain. *Health Psychol,* 26 (1), 1-9.

7. Van Koulil S, Effting M, Kraaimaat FW, Van Lankveld W, Van Helmond T, Cats H, Van Riel PL, De Jong AJ, Evers AW. (2007) Cognitive-behavioural therapies and exercise programmes for patients with fibromyalgia: state of the art and future directions. *Ann Rheum Dis,* 66 (5), 571-581.

8. Ipser JC, Sander C, Stein DJ. (2009) Pharmacotherapy and psychotherapy for body dysmorphic disorder. *Cochrane Database Syst Rev* (1), CD005332.

9. Williams J, Hadjistavropoulos T, Sharpe D. (2006) A meta-analysis of psychological and pharmacological treatments for Body Dysmorphic Disorder. *Behav Res Ther,* 44 (1), 99-111.

10. Greeven A, Visser S, Merkelbach JW, Van Rood YR, Van Dyck R, Zitman FG, Spinhoven P. (2007) Cognitive behavior therapy and paroxetine in the treatment of hypochondriasis: a randomized controlled trial. *Am J Psychiatry,* 164 (1), 91-99.

11. Allen LA, Escobar JI, Lehrer PM, Gara MA, Woolfolk RL. (2002) Psychosocial treatments for multiple unexplained physical symptoms: a review of the literature. *Psychosom Med,* 64 (6), 939-950.

12. Kroenke K. (2007) Efficacy of treatment for somatoform disorders: a review of randomized controlled trials. *Psychosom Med,* 69 (9), 881-888.

13. Raine R, Haines A, Sensky T, Hutchings A, Larkin K, Black N. (2002) Systematic review of mental health interventions for patients with common somatic symptoms: can research evidence from secondary care be extrapolated to primary care? *BMJ,* 325 (7372), 1082.

14. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC. (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*, 59 Suppl 20, 22-33.

15. De Beurs E, Den Hollander-Gijsman ME, Van Rood YR, Van der Wee NJ, Giltay EJ, Van Noorden MS, Van der Lem R, Van Fenema EM, Zitman FG. (2011) Routine outcome monitoring in the Netherlands: practical

experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clin Psychol Psychother,* 18, 1-12.

16. Schulte-van Maaren YWM, Carlier IVE, Giltay EJ, Van Noorden MS, De Waal MW, Van der Wee NJ, Zitman FG. (2012) Reference values for mental health assessment instruments: objectives and methods of the Leiden Routine Outcome Monitoring Study. *J Eval Clin Pract*.

17. Littleton HL, Axsom D, Pury CL. (2005) Development of the body image concern inventory. *Behav Res Ther,* 43 (2), 229-241.

18. Pilowsky I. (1967) Dimensions of hypochondriasis. *Br J Psychiatry,* 113 (494), 89-93.

19. Vercoulen JH, Alberts M, Bleijenberg G. (1999) De Checklist Individual Strenght. *Gedragstherapie,* 32, 131-136.

20. Kazdin AE. (2008) Evidence-based treatment and practice: new opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *Am Psychol*, 63 (3), 146-159.

21. Littleton H, Breitkopf CR. (2008) The Body Image Concern Inventory: validation in a multiethnic sample and initial development of a Spanish language version. *Body Image,* 5 (4), 381-388.

22. Luca M, Giannini M, Gori A, Littleton H. (2011) Measuring dysmorphic concern in Italy: psychometric properties of the Italian Body Image Concern Inventory (I-BICI). *Body Image,* 8 (3), 301-305.

23. Speckens AE, Van Hemert AM, Spinhoven P, Bolk JH. (1996) The diagnostic and prognostic significance of the Whitely Index, the Illness Attitude Scales and the Somatosensory Amplification *Scale*. *Psychol Med,* 26 (5), 1085-1090.

24. Poortvliet MC, Lamkadden M, Deville W. Niet op naam ingeschreven (NONI) bij de huisarts. Inventarisatie en gevolgen voor de ziekenfondsverzekerden. Utrecht: NIVEL; 2005.

25. Schulte-van Maaren YWM, Carlier IVE, Zitman FG, Hemert AM, De Waal MW, Van Noorden MS, Giltay EJ. (2012) Reference values for generic instruments used in Routine Outcome Monitoring: the Leiden Routine Outcome Monitoring Study (in press). *BMC Psychiatry*.

26. Schulte-van Maaren YWM, Carlier IVE, Zitman FG, Van Hemert AM, De Waal MW, Van der Does AJW, Van Noorden MS, Giltay EJ. (2012) Reference values for major depression questionnaires: the Leiden Routine Outcome Monitoring Study (submitted). *Journal of Affective Disorders*.

27. Van Noorden MS, Giltay EJ, Den Hollander-Gijsman ME, Van der Wee NJ, Van Veen T, Zitman FG. (2010) Gender differences in clinical characteristics in a naturalistic sample of depressive outpatients: the Leiden Routine Outcome Monitoring Study. *J Affect Disord*, 125 (1-3), 116-123.

28. Van Vliet IM, De Beurs E. (2007) The MINI-International Neuropsychiatric Interview. A brief structured diagnostic psychiatric interview for DSM-IV en ICD-10 psychiatric disorders. *Tijdschr Psychiatr,* 49 (6), 393-397.

29. Van Rood YR, Mulkens S, Bouman T. (2013) Psychometrie van body dysmorphic disorder. Denys D, Vulink N, editors. Body dysmorphic disorder. [11], 163-182. Assen, NL, Van Gorcum.

30. Speckens AE, Spinhoven P, Sloekers PP, Bolk JH, Van Hemert AM. (1996) A validation study of the Whitely Index, the Illness Attitude Scales, and the Somatosensory Amplification

Scale in general medical and general practice patients. *J Psychosom Res*, 40 (1), 95-104.

31. De Vree B, Van der Werf SP, Prins J, Bazelmans E, Servaes P, De Vries M, Bleijenberg G. (2002) Meetinstrumenten bij chronische vermoeidheid. *Gedragstherapie*, 35 (2), 157-164.

32. Dingemans AE, Van Rood YR, De Groot I, Van Furth EF. (2012) Body dysmorphic disorder in patients with an eating disorder: prevalence and characteristics. *Int J Eat Disord,* 45 (4), 562-569.

33. Vercoulen JH, Swanink CM, Fennis JF, Galama JM, Van der Meer JW, Bleijenberg G. (1994) Dimensional assessment of chronic fatigue syndrome. *J Psychosom Res,* 38 (5), 383-392.

34. Solberg HE. (2008) Establishment and use of reference values. Burtis CA, Ashwood ER, Bruns DE, editors. Fundamentals of clinical chemistry. 6[14], 229-238. St. Louis, Missouri, Saunders Elsevier.

35. Barnabei L, Marazia S, De CR. (2007) Receiver operating characteristic (ROC) curves and the definition of threshold levels to diagnose coronary artery disease on electrocardiographic stress testing. Part I: The use of ROC curves in diagnostic medicine and electrocardiographic markers of ischaemia. *J Cardiovasc Med (Hagerstown)*, 8 (11), 873-881.

36. Solberg HE. (1989) Reference values. *Adv Clin Chem,* 27, 1-79.

37. Zimmerman M, Chelminski I, Posternak M. (2004) A review of studies of the Montgomery-Asberg Depression Rating Scale in controls: implications for the definition of remission in treatment studies of depression. *Int Clin Psychopharmacol,* 19 (1), 1-7.

38. Kendall PC, Marrs-Garcia A, Nath SR, Sheldrick RC. (1999) Normative comparisons for the evaluation of clinical significance. *J Consult Clin Psychol,* 67 (3), 285-299.

39. Reed AH, Henry RJ, Mason WB. (1971) Influence of statistical method used on the resulting estimate of normal range. *Clin Chem*, 17 (4), 275-284.

40. Jacobi F, Wittchen HU, Holting C, Hofler M, Pfister H, Muller N, Lieb R. (2004) Prevalence, co-morbidity and correlates of mental disorders in the general population: results from the German Health Interview and Examination Survey (GHS). *Psychol Med,* 34 (4), 597-611.

41. Faravelli C, Salvatori S, Galassi F, Aiazzi L, Drei C, Cabras P. (1997) Epidemiology of somatoform disorders: A community survey in Florence. *Social Psychiatry and Psychiatric Epidemiology,* 32 (1), 24-29.

42. Koran LM, Abuiaoude E, Large MD, Serpe RT. (2008) The prevalence of body dysmorphic disorder in the United States adult population. *Cns Spectrums*, 13 (4), 316-322.

43. Otto MW, Wilhelm S, Cohen LS, Harlow BL. (2001) Prevalence of body dysmorphic disorder in a community sample of women. *American Journal of Psychiatry*, 158 (12), 2061-2063.

44. Rief W, Buhlmann U, Wilhelm S, Borkenhagen A, Brahler E. (2006) The prevalence of body dysmorphic disorder: a population-based survey. *Psychological Medicine,* 36 (6), 877-885.

45. De Waal MW, Arnold IA, Eekhof JA, Van Hemert AM. (2004) Somatoform disorders in general practice: prevalence, functional impairment and comorbidity with anxiety and depressive disorders. *Br J Psychiatry*, 184, 470-476.

46. Bohne A, Keuthen NJ, Wilhelm S, Deckersbach T, Jenike MA. (2002) Prevalence of symptoms of body dysmorphic disorder and its correlates: a cross-cultural comparison. *Psychosomatics*, 43 (6), 486-490.

47. Cororve MB, Gleaves DH. (2001) Body dysmorphic disorder: a review of conceptualizations, assessment, and treatment strategies. *Clin Psychol Rev,* 21 (6), 949-970.

48. Fink P, Sorensen L, Engberg M, Holm M, Munk-Jorgensen P. (1999) Somatization in primary care. Prevalence, health care utilization, and general practitioner recognition. *Psychosomatics*, 40 (4), 330-338.

49. Li CT, Chou YH, Yang KC, Yang CH, Lee YC, Su TP. (2009) Medically unexplained symptoms and somatoform disorders: diagnostic challenges to psychiatrists. *J Chin Med Assoc,* 72 (5), 251-256.

50. Mergl R, Seidscheck I, Allgaier AK, Moller HJ, Hegerl U, Henkel V. (2007) Depressive, anxiety, and somatoform disorders in primary care: prevalence and recognition. *Depress Anxiety,* 24 (3), 185-195.

Supplementary Tables

Supplementary Table 6.1: Sociodemographic and psychiatric characteristics of the ROM reference (n=648) and patient (n=823) groups per gender.

Supplementary Table 6.2: Percentile scores and mean values in the ROM reference (n=645) and patient (n=130) groups for the subscales and total score of the Body Image Concern Inventory (BICI)

Supplementary Table 6.3: Percentile scores and mean values in the ROM reference (n=644) and patient (n=226) groups for the subscales and total score of the Whitely Index (WI)

Supplementary Table 6.4: Percentile scores and mean values in the ROM reference (n=643) and patient (n=481) groups for the subscales and total score of the Checklist Individual Strength (CIS-20R)

Supplementary Table 6.1: Sociodemographic and psychiatric characteristics of the ROM reference (n=648) and patient (n=823) groups per gender.

	ROM reference group (n=648)						ROM patient group (n=823)					
	females		males		total		females		males		total	
Gender: - n (%)	405	(62.5)	243	(37.5)	648	(100)	580	(70.5)	243	(29.5)	823	(100)
Age in years: mean (± SD)	39.6	(12.6)	40.8	(12.6)	40.0	(12.6)	38.9	(11.5)	38.0	(12.3)	38.6	(11.7)
Marital status[1]: - n (%)												
Married/cohabitating	285	(70.4)	172	(70.8)	457	(70.5)	285	(49.1)	107	(44.0)	392	(47.6)
Divorced/separated/widow	28	(6.9)	6	(2.5)	34	(5.2)	79	(13.6)	17	(7.0)	96	(11.7)
Single	92	(22.7)	65	(26.7)	157	(24.2)	137	(23.6)	79	(32.5)	216	(26.2)
Housing situation[1]: - n (%)												
Living alone	55	(13.6)	32	(13.2)	87	(13.4)	98	(16.9)	40	(16.5)	138	(16.8)
Living with partner	287	(70.9)	175	(72.0)	462	(71.3)	291	(50.2)	112	(46.1)	403	(49.0)
Living with family	63	(15.6)	36	(14.8)	99	(15.3)	112	(19.3)	51	(21.0)	163	(19.8)
Educational status1,3: - n (%)												
Lower	90	(22.2)	48	(19.8)	138	(21.3)	204	(35.2)	81	(33.3)	285	(34.6)
Higher	315	(77.8)	195	(80.2)	510	(78.7)	297	(51.2)	122	(50.2)	419	(50.9)
Employment status[1]: - n (%)												
Employed part-time	211	(52.1)	45	(18.5)	256	(39.5)	135	(23.3)	22	(9.1)	157	(19.1)
Employed full-time	111	(27.4)	165	(67.9)	276	(42.6)	42	(7.2)	75	(30.9)	117	(14.2)
Unemployed/retired	73	(18.0)	28	(11.5)	101	(15.6)	157	(27.1)	43	(17.7)	200	(24.3)
Work-related disability	10	(2.5)	5	(2.1)	15	(2.3)	167	(28.8)	63	(25.9)	230	(27.9)
Ethnic background[1]: - n (%)												
Dutch	348	(85.9)	221	(90.9)	569	(87.8)	417	(71.9)	172	(70.8)	590	(71.6)
Other ethnicity	57	(14.1)	22	(9.1)	79	(12.2)	83	(14.3)	31	(12.8)	114	(13.9)
MINI diagnoses: - n (%)												
Currently None	362	(89.4)	228	(93.8)	590	(91.0)	0[2]		0[2]		0[2]	
Somatoform disorder (single)	15	(3.7)	10	(4.1)	25	(3.9)	294	(50.7)	88	(36.2)	382	(46.4)
Somatoform disorder (comorbidity)	6	(1.5)	1	(0.4)	7	(1.1)	286	(49.3)	155	(63.8)	441	(53.6)
Other psychiatric disorder without somatoform	22	(5.4)	4	(3.6)	26	(4.0)	0[2]		0[2]		0[2]	

[1] No data from 119 (14%) patients; [2] Selection criterion; [3] Lower education: primary or vocational school; Higher education: college or university

Supplementary Table 6.2: Percentile scores and mean values in the ROM reference (n=645) and patient (n=130) groups for the subscales and total score of **the Body Image Concern Inventory (BICI)**

	ROM reference group (645)						ROM patient group (130)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Body Image Concern Inventory (BICI)												
All participants	**20**	**26**	**33**	**41**	**55**	**34.4 ± 10.8**	**39**	**60**	**72**	**78**	**87**	**68.4 ± 13.9**
- Men	19	23	27	34	45	29.0 ± 8.4	**29**	**56**	**67**	**73**	**80**	**63.1 ± 14.3**
- Women	22	30	36	43	57	37.4 ± 10.8	45	63	74	80	90	71.6 ± 12.7

SD denotes standard deviation.

Supplementary Table 6.3: Percentile scores and mean values in the ROM reference (n=644) and patient (n=226) groups for the subscales and total score of **the Whitely Index (WI)**

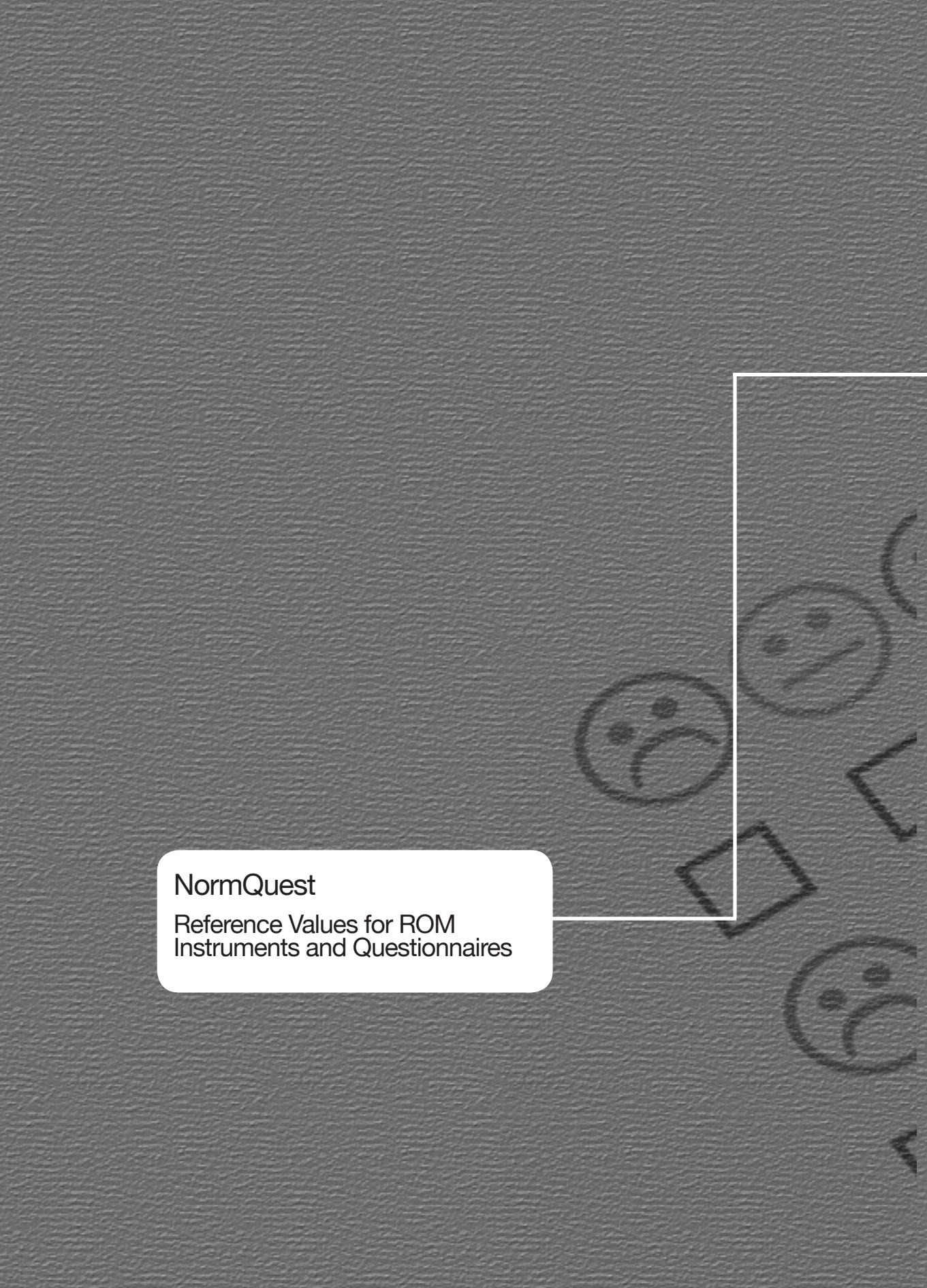
	ROM reference group (644)						ROM patient group (226)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Whitely Index (WI)												
All participants	**0**	**1**	**2**	**3**	**6**	**2.2 ± 2.0**	**5**	**8**	**10**	**12**	**13**	**9.8 ± 2.5**
- Men	0	1	1	3	5	1.9 ± 1.6	**6**	**7**	**10**	**12**	**13**	**9.5 ± 2.5**
- Women	0	1	2	3	7	2.4 ± 2.2	5	9	10	12	13	10.1 ± 2.5

SD denotes standard deviation.

Supplementary Table 6.4: Percentile scores and mean values in the ROM reference (n=643) and patient (n=481) groups for the subscales and total score of the Checklist **Individual Strength (CIS-20R)**

	ROM reference group (644)						ROM patient group (226)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Checklist Individual Strength (CIS-20R)												
Subjective fatigue												
All participants	8	9	16	27	46	20.2 ± 12.6	38	50	54	56	56	51.5 ± 7.0
- Men	8	8	14	25	42	18.1 ± 11.3	33	50	55	56	56	50.8 ± 8.7
- Women	8	10	18	29	50	21.4 ± 13.2	38	50	54	56	56	51.8 ± 6.4
Concentration												
All participants	5	5	9	14	26	11.0 ± 7.0	6	22	29	35	35	26.7 ± 8.9
- Men	5	5	9	14	29	10.9 ± 7.1	13	23	29	35	35	27.6 ± 7.5
- Women	5	5	9	14	26	11.1 ± 7.0	5	21	29	35	35	26.5 ± 9.2
Motivation												
All participants	4	4	7	10	20	8.4 ± 5.1	4	12	17	24	28	17.8 ± 7.3
- Men	4	4	7	10	20	8.1 ± 5.0	10	16	22	25	28	19.9 ± 6.2
- Women	4	4	7	11	20	8.7 ± 5.2	4	10	16	24	28	17.3 ± 7.4
Activity												
All participants	3	3	4	9	15	6.2 ± 4.3	3	12	18	21	21	16.0 ± 5.6
- Men	3	3	5	9	15	6.5 ± 4.4	8	15	19	21	21	17.3 ± 4.3
- Women	3	3	4	8	15	6.0 ± 4.2	3	10	18	21	21	15.6 ± 5.9
Total												
All participants	20	28	38	60	92	45.8 ± 23.3	74	100	116	128	140	112.0 ± 20.4
- Men	20	26	37	57	89	43.5 ± 22.5	73	104	122	129	140	115.5 ± 19.6
- Women	20	28	41	61	97	47.2 ± 23.6	74	98	114	128	140	111.1 ± 20.5

SD denotes standard deviation.



NormQuest

Reference Values for ROM Instruments and Questionnaires

Chapter 7

Development and validation of the 48-item Symptom Questionnaire (SQ-48) in patients with depressive, anxiety and somatoform disorders

Ingrid Carlier
Yvonne Schulte-Van Maaren
Klaas Wardenaar
Erik Giltay
Martijn Van Noorden
Peter Vergeer
Frans Zitman

Psychiatry Research (2012), 200(2-3):904-10

ABSTRACT

Self-report measures of psychological distress or psychopathology are widely used and can be easily implemented as psychiatric screening tools. Positive psychological constructs such as vitality/optimism and work functioning have scarcely been incorporated. We aimed to develop and validate a psychological distress instrument, including measures of vitality and work functioning. A patient sample with suspected depressive, anxiety, and somatoform disorders (N=242) and a reference sample of the general population (N=516) filled in the 48-item Symptom Questionnaire (SQ-48) plus a battery of observer-rated and self-report scales (MINI Plus, MADR, BAS, INH, BSI), using a web-based ROM program. The resulting SQ-48 is multidimensional and includes the following nine subscales: Depression (MOOD, six items), Anxiety (ANXI, six items), Somatization (SOMA, seven items), Agoraphobia (AGOR, four items), Aggression (AGGR, four items), Cognitive problems (COGN, five items), Social Phobia (SOPH, five items), Work functioning (WORK, five items), and Vitality (VITA, six items). The results showed good internal consistency as well as good convergent and divergent validity. The SQ-48 is meant to be available in the public domain for Routine Outcome Monitoring (ROM) and can be used as a screening/monitoring tool in clinical settings (psychiatric and non-psychiatric), as a benchmark tool, or for research purposes.

INTRODUCTION

The measurement of self-reported psychological distress is prominently represented in both the psychological and psychiatric literature. Historically, assessment of the general psychological status of individuals by means of self-report dates back to the First World War, and the development of the so called Personal Data Sheet by Woodworth [1]. Woodworth's scale provided a means for each man to "interview himself" and created a historical benchmark for a new modality of psychological measurement [2]. Nowadays, self-report measures of psychological distress or psychopathology are widely used as psychiatric screening tool in clinical settings and epidemiological studies.

Many validated self-report questionnaires for measuring psychological distress or psychopathology have been developed [3-6]. For instance, Symptom Checklist-90 [7] and its short-form Brief Symptom Inventory [2,8]; General Health Questionnaire [9]; 50-Item Brief Symptom Rating Scale [10]; Talbieh Brief Distress Inventory [11]; Mood and Anxiety Symptoms Questionnaire [12,13] and its short-form MASQ-D30 [14].

Studies concerning the above-mentioned instruments often used multiple related concepts interchangeably: concepts such as psychological distress, emotional distress, affective distress, mental distress, global distress, symptom distress, psychiatric distress, general psychopathology. Notably, however, these instruments have been useful for assessing the aggregate level of nonspecific psychological distress, and not for diagnosing particular psychiatric disorders [3,5,15]. Elevated scores on the scales are an indicator of possible psychopathology and could assist the clinician to predict the probability of individuals meeting criteria for disorder [3,16,17].

More specifically, "psychological distress" can be described as a reaction of an individual to external and internal stresses, characterized by a mixture of psychological symptoms, such as sadness, anxiety, confused thinking, hopelessness, helplessness, dread, and poor self-esteem [6]. In addition, some instruments, such as the BSI, include somatic distress. Psychological distress was originally considered as a uni-dimensional construct. However, more recent research suggested a multidimensional structure of psychological distress. For instance, Schwannauer and Chetwynd [18] found a three-factor model of depression, anxiety, and general psychological distress.

The assessment of psychological distress is important both in health care and mental health care, because of its relevance for compliance, quality of life, prediction of treatment outcome, and planning of treatment [6,19-21]. Research has shown that pervasive distress may affect the course of illness, symptom expression, as well as levels of social relationships and adaptation [6,22-24].

relationships and adaptation [6,22-24].

More recently, there is a growing awareness that, in addition to distress-based measures, attention must also be paid to more positive constructs such as vitality/optimism

[25,26] and work functioning [27,28]. The importance of both constructs has already been demonstrated. For instance, Burdick et al. [27] showed that poor work functioning was significantly related to subsyndromal depression and course of illness. Emotional vitality, on the other hand, seems to be a critical positive psychological factor (related to but separate from optimism) that may promote psychological health as well as physical health [25,29-31]. In addition, (lack of) vitality/optimism has been shown to be an important defining feature of depression, with distinct implications for prognosis [32].

To date, there is no psychological distress instrument available that also measures vitality and work functioning. Another shortcoming is that most self-report instruments are usually not free of charge, which particularly in Routine Outcome Monitoring (ROM) with repeated assessment is a costly matter. In line with these shortcomings, the purpose of this study was to develop and validate a brief psychological distress instrument (SQ - 48), which also includes measures of vitality and work functioning (or study). In addition, the SQ-48 is developed as a public domain questionnaire, freely available to clinicians and researchers. This practical advantage is in line with growing efforts in other scientific areas to develop instruments that are free of charge [33].

The SQ-48 is meant as a screening tool to improve diagnostic recognition in clinical and nonclinical settings. Therefore, the present study used both clinical and nonclinical samples: a patient sample with suspected depressive, anxiety, and somatoform disorders, and a reference sample of the general population. In this way, the SQ-48 could be useful as a monitoring tool in the context of ROM [4,33,34], for benchmark purposes (Hermann et al., 2006; Minami et al., 2008; Cleary et al., 2010), or as a research tool in for instance epidemiological studies.

METHODS

The present study was conducted with patients and non-patients, and consisted of two phases: (1) instrument development of the SQ-48 and (2) its psychometric evaluation.

Participants and procedures

The total sample among which the SQ-48 was developed and evaluated consisted of participants from two large studies: a Routine Outcome Monitoring (ROM) sample of psychiatric outpatients and a ROM reference sample of the general population.

The Medical Research Ethics Committee at the Leiden University Medical Centre (LUMC) approved the general study protocol and documents presented to participants in both phases. A comprehensive protocol safeguards anonymity of ROM-participants and ensures proper handling of the data. This protocol (Psychiatric Academic Registration Leiden database) is available on request for participants, and informed consent is not required for patients. Non-patients provided written and informed consent.

For details about the web-based ROM programme of the LUMC, Department of Psychiatry, we refer to some relevant publications [26,34-36]; see also www.lumc.nl/psychiatry/ROM-instruments).

The ROM patient-group

A total of 242 psychiatric outpatients was included (61.2% females; mean age=38.8 years; SD=14.0), referred with suspected (not necessarily diagnosed) mood, anxiety or somatoform disorders to the LUMC Department of Psychiatry or to Rivierduinen specialized mental healthcare centres. Data were collected during a 2-3 h ROM baseline assessment in the LUMC or at the home of the participant. The assessment consisted of a face-to-face psychiatric interview by a trained psychiatric research nurse and the administration of observer-rated and self-report questionnaires, including the SQ-48.

The ROM reference-group

A total of 516 participants (67.2% females; mean age=38.8 years; SD=12.8) was included in the reference-group, as part of the 'Leiden Routine Outcome Monitoring Study' [36,37]. These participants were randomly selected from the registration systems of general practitioners (GPs) in the Leiden region, in order to recruit a representative general population sample (all Dutchmen are registered with a GP). Because the group was aimed to be used as a healthy reference-group, participants that received treatment for psychiatric problems and/or dependence on alcohol or drugs within six months prior to the assessment were excluded. The inclusion for the ROM reference-group was stratified for gender (62.6% women), age (mean 40.2 years; SD 12.5) and urbanization-level (62.3% urban), to make the group demographically comparable to the ROM patient-group. The participants in the reference-group completed the same assessments as the patient-group.

Instruments

Development of the SQ-48

The SQ-48 was developed to include separate subscales concerning several psychopathological domains matching diagnostic categories in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; [38]. Three main goals were set to achieve during the SQ-48 item-writing and item-selection procedure. First, the items should be easy and unequivocal to understand for everyone, irrespective of level-of-education. Second, the instrument should include measures of functioning which can judge the actual impact of psychiatric problems on daily life. Third, the instrument should cover (lack of) vitality/optimism.

The initial item development followed commonly accepted methods for the creation of patient-reported instruments [39-44]. The questionnaire was drafted by a multidisciplinary team of psychologists and psychiatrics through a comprehensive review of existing screening tools, relevant literature, as well as psychiatric diagnostic criteria for mood, anxiety, and somatoform disorders on the basis of the DSM-IV.

Existing screening tools reviewed in this context were for instance: MASQ (-D30); Dimensional Assessment of Personality Pathology-Short Form [45]; BSI; Short-Form 36 [46]; Outcome Questionnaire 45 [47,48]; Fear Questionnaire [48]; Aggression Questionnaire [49]; Mental Vitality Scale [50]; Work Home and Leisure Activities Scale [51]; MIRECC Version of the Global Assessment of Functioning Scale [52]; Physical Symptom Checklist [53]; Life Orientation Test-revised concerning optimism [54].

The SQ-48 development was based on consensus within the aforementioned multidisciplinary team. It was decided to create a questionnaire covering nine domains or categories: depression, anxiety, somatization, cognitive problems, social phobia, agoraphobia, aggression, work (or study) functioning, and vitality/optimism. Except work functioning and vitality, these general domains cover the most common psychopathological symptoms. So, items were arranged in subscales according to this organization and chosen from a large pool of items. Each item was evaluated to determine whether it was formulated in the simplest way and whether it was unambiguous in its meaning. If there was any disagreement about this within the team, the item was not included. Additional care was taken to prevent redundancy within subscales and to prevent overlap between subscales, to increase the potential discriminant ability of the subscales. The experimental version of the questionnaire was pre-tested in a reduced sample (n=30) of participants in the ROM programme. The aim of the pre-test was to evaluate the practicality and acceptability by collecting comments of participants, clinicians, investigators, in order to better formulate the items.

The final version of the questionnaire included 48 items based on re-evaluation. Re-evaluation consisted of an Exploratory Factor Analysis with oblique Promax-rotation in both non-reduced samples, to check for items with ambiguous factor loads (loads on more than one factor). On the whole, the following seven items were removed because the factor loads indicated poor fit: “I felt confused” (factor load 0.18); “I had the feeling as if something

terrible was going to happen" (factor load 0.21); "I could not relax in the company of others" (factor load 0.03); "I have threatened people I know" (factor load 0.38); "I was incited by people" (factor load 0.24); "I couldn't enjoy my free time" (factor load 0.01); "In the morning I was full of energy" (factor load 0.61).

The final version of the SQ-48

The nine subscales of the SQ-48 corresponded with the abovementioned domains of interest. Five subscales covered aspects of psychopathology: Depression ("MOOD" subscale: items 3, 7, 13, 19, 38, 40), Anxiety, ("ANXI" subscale: items 24, 28, 33, 41, 46, 48), Somatization ("SOMA" subscale: items 1, 5, 11, 17, 22, 26, 31), and Agoraphobia ("AGOR" subscale: items 4, 8, 14, 25). In addition, four subscales were constructed to assess specific aspects of behaviour and/or functioning: Aggression ("AGGR" subscale: items 10, 16, 21, 43), Cognitive problems ("COGN" subscale: items 2, 6, 39, 44, 47), Social Phobia ("SOPH" subscale: items 23, 27, 32, 36, 45), Work ("WORK" subscale: items 9, 15, 20, 30, 35), and Vitality/Optimism ("VITA" subscale: items 12, 18, 29, 34, 37, 42). Each item is rated by the respondent on a 5-points Likert-scale (0: 'Never', 1: 'Rarely', 2 'Sometimes', 3: 'Often', 4: 'Very often'). Mean administration time was 5.4 minutes (S.D.=1.4).

Respondents received the following instruction in the SQ-48: "Try to answer the following statements honestly and accurately. Please indicate what applies best to you. There are no 'right' or 'wrong' answers. Give the answer that best expresses how often you have felt that way in the last week, including today. The answer which comes to your mind first, is often the best answer. Note: If you did not work or study or have not been able to do so, then you can skip questions 9, 15, 20, 30 and 35".

The scoring of the SQ-48 items is as follows. For the score of all subscales, the scores of the relevant items must be added. The 48 items are scored 0-4.

For the purpose of this article, the Dutch SQ-48 was translated into English, according to guidelines for translation and cultural adaptation of questionnaires [55-57]. Both English and Dutch SQ-48 are available as Supplementary material associated with this article, and can be found in the online version.

Other measures

In both groups, the same battery of other measures was administered. The presence of DSM-IV diagnoses was determined by a trained psychiatric research nurse by means of the Mini-International Neuropsychiatric Interview (version: MINI Plus; [58]. General psychopathology was assessed with two generic measures. The first was an observational instrument, the CPRS-SF (Comprehensive Psychopathological Rating Scale-Short Form) consisting of: the Montgomery–Åsberg Depression Rating Scale (MADRS; [59], the Brief Anxiety Scale (BAS; [60], and a scale assessing psychomotor inhibition (INH; [61,62]. The second generic instrument, the BSI, is a self-report instrument that assesses psychopathological symptoms in several domains such as depressive -, anxiety-, somatic symptoms, and hostility [2,63].

Statistical analyses

Analyses were performed in both the patient and reference samples and in the combined datasets. In both samples, data were prepared: missing values were substituted by the mean item-response per subject per subscale. Subjects who had more than three missing values for the total sum score (or more than one per subscale) were excluded.

To evaluate the construct validity, Confirmatory Factor Analysis (CFA) was used on the non-reduced samples (N=516 patient sample, N=242 reference sample). The appropriateness of a 9-factor model (the SQ-48 subscale structure) was evaluated and compared with a 1-factor model. In the input model, all items were set to load freely on their hypothesized factor, except for one item per factor, which had its loading set to 1 in order to fix the scale of the model. Because the items were categorical and non-normally distributed, fit-estimations were based on robust maximum likelihood [64], using polychoric correlation matrices [65]. Fit-indices instead of a traditional $\chi 2$-test were used to assess fit, because the $\chi 2$-test is oversensitive to misfit when testing complex models [66]. The used fit-indices were the Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA). A CFI of at least 0.90 indicates adequate fit, and an RMSEA that is smaller than 0.08 indicates acceptable fit. The CFA was conducted with EQS 6.1 [65].

To investigate internal consistency, Cronbach's alphas were calculated for the subscales and the total scale. To investigate the extent of differentiation between the subscales, Spearman's (ϱ) correlation coefficients were computed for intercorrelations of the SQ-48 subscales. To evaluate convergent/divergent validity, correlations (Spearman's ϱ) between the subscale scores and other instruments (see paragraph 2.2.3) were calculated. ROC (Receiver Operating Characteristic) analysis provided a cut-off score indicating an optimal discrimination threshold between "healthy" and "diseased". Sensitivity and specificity were chosen to be equal, taking into account the trade-off between the two. AUC's (Area Under Curves) were calculated to indicate the predictive capacities of the instrument subscales.

RESULTS

Sociodemographic and clinical characteristics of the samples

The sociodemographic characteristics of the two research groups are shown in Table 7.1.

Table 7.1: Sociodemographic and clinical characteristics of the samples

	Reference group (n=516)	Patient group (n=242)	p-value
Female gender (%)	347 (67.2%)	149 (61.6%)	0.13
Age (yr), mean (SD)	38.8 (12.7)	37.9 (12.9)	0.38
Marital status[1]: - n (%)			
Married/cohabitating	355 (68.4%)	115 (47.5%)	<0.001
Divorced/seperated/widow	23 (4.5%)	29 (12.0%)	
Single	140 (27.1%)	98 (40.5)	
Housing situation[1]: - n (%)			
Living alone	79 (15.3%)	77 (31.8%)	0.02
Living with partner	357 (69.2%)	115 (47.5%)	
Living with family	80 (15.5%)	50 (20.7%)	
Educational status, n (%)			
Lower	126 (24.4%)	73 (30.1%)	0.07
Higher	390 (75.6%)	149 (69.8%)	
Employment status, n (%)			
Employed part-time	218 (42.2%)	54 (22.3%)	<0.001
Employed full-time	199 (38.6%)	53 (21.9%)	
Unemployed/retired	84 (16.3%)	76 (31.4%)	
Work-related disability	15 (2.9%)	59 (24.4%)	
Ethnic background, n (%)			
Dutch	467 (90.5%)	198 (81.8%)	0.07
Other ethnicity	49 (9.5%)	44 (18.2%)	

Both groups were similar with regards to most sociodemographic variables. As expected – because of the sampling procedure – the mean age and gender distribution were comparable between the reference- and patient-groups. Educational status was also roughly similar; the reference-group had 75.6% higher education compared to 69.8% of the patient-group education. However, the groups also differed on some aspects. In the patient-group, participants were less often married and more often unemployment or with work-related disability compared to the reference group.

Table 7.2: Clinical characteristics according to group.

	Range of scores	Reference group (n=516)	Patient group (n=242)	p-value
Psychiatric scales, mean (SD)				
MADRS	0-6	3.0 (4.0)	18.5 (9.1)	<0.001
INH	0-6	0.62 (1.5)	3.4 (3.0)	<0.001
BSI	0-4	0.2 (0.3)	1.2 (0.7)	<0.001
BAS	0-6	4.4 (4.2)	13.9 (6.2)	<0.001
SQ-48 scores, mean (SD)				
MOOD	0-24	2.1 (2.5)	11.3 (6.3)	<0.001
ANXI	0-24	3.8 (3.9)	12.1 (5.8)	<0.001
SOMA	0-28	1.7 (3.1)	6.3 (6.2)	<0.001
AGOR	0-16	0.4 (1.1)	2.8 (3.6)	<0.001
AGGR	0-16	1.2 (1.7)	3.7 (3.4)	<0.001
COGN	0-20	4.1 (3.4)	11.3 (4.6)	<0.001
SOPH	0-20	2.4 (3.0)	8.1 (5.1)	<0.001
WORK	–	–	–	<0.001
VITA	0-24	15.8 (4.6)	9.2 (4.9)	<0.001
MINI-Diagnoses, n (%)				
Depressive disorder		1 (0.2%)	64 (26.4%)	<0.001
Anxiety disorders		35 (6.8%)	31 (12.8%)	<0.001
Comorbid depression & Anxiety		6 (1.2%)	66 (27.3%)	<0.001

BSI denotes the short-form Brief Symptom Inventory, BAS denotes Brief Anxiety Scale, INH denotes the scale assessing psychomotor inhibition, and MADRS denotes Montgomery-Åsberg Depression Rating Scale. SQ-48 subscales: MOOD denotes Depression, ANXI denotes Anxiety, SOMA denotes Somatization, AGOR denotes Agoraphobia, AGGR denotes Aggression, COGN denotes Cognitive problems, SOPH denotes Social Phobia, WORK denotes Work functioning, and VITA denotes Vitality. Because of adjustments made to the final version of instructions to the WORK subscale, insufficient data were available as yet.

The clinical characteristics of the two groups are shown in Table 7.2.
As expected, the scores on all psychopathology ratings were much higher in the patient-group than in the reference-group. Specifically, the mean total SQ-48 score in the patient-group (73.0) was twice as high compared to the reference-group (36.6). The majority of the patient-group met criteria for depression and anxiety disorder (27.3%) versus a neglectable few in the reference-group (1.2 %).

Confirmatory Factor Analyses

CFA was conducted to test the fit of a 9-factor structure to the SQ-48 data. The hypothesized model fitted well with the data in both the reference-group (CFI=0.96; RMSEA=0.05) and the patient-group (CFI=0.97; RMSEA=0.06). In addition, the fit of a simple 1-factor model was worse in both samples (reference group: CFI=0.88; RMSEA=0.08; patient group: CFI=0.88; RMSEA=0.13).

Scale intercorrelations

The Spearman rho's intercorrelations of the SQ-48 subscales are shown in Table 7.3.
The correlations ranged from 0.38 to 0.81, with the highest correlations between MOOD and ANXI (ϱ=0.81), MOOD and COGN (ϱ=0.78), COGN and ANXI (ϱ=0.76), and between ANXI and SOPH (ϱ=0.73). The lowest correlations were found between VITA and AGGR (ϱ=0.38), and between AGOR and AGGR (ϱ=0.39).

Table 7.3: Correlations between the subscales of the SQ-48 in all 758 subjects.

	MOOD	ANXI	SOMA	COGN	SOPH	AGOR	AGGR
ANXI	0.81						
SOMA	0.52	0.59					
COGN	0.78	0.76	0.55				
SOPH	0.69	0.73	0.47	0.72			
AGOR	0.51	0.57	0.50	0.49	0.56		
AGGR	0.60	0.60	0.44	0.56	0.54	0.39	
VITA	–0.66	–0.57	–0.43	–0.60	–0.54	–0.45	–0.38

Data are Spearman's (rho) correlation coefficients. All P-values <0.001.
SQ-48 subscales: MOOD denotes Depression, ANXI denotes Anxiety, SOMA denotes Somatization, AGOR denotes Agoraphobia, AGGR denotes Aggression, COGN denotes Cognitive problems, SOPH denotes Social Phobia, WORK denotes Work functioning, and VITA denotes Vitality.
Because of adjustments made to the final version of instructions to the WORK subscale, insufficient data were available as yet.

Internal consistency

The internal consistency coefficients of the SQ-48 subscales were as follows. In general, the Cronbach's alpha coefficients ranged from 0.78 to 0.98 across the different SQ-48 subscales: 0.97 (Total); 0.93 (MOOD); 0.92 (ANXI); 0.89 (SOMA); 0.89 (COGN); 0.91 (SOPH), 0.84 (AGOR); 0.78 (AGGR); 0.90 (VITA); 0.78 (WORK). So, none of the subscales had alphas below the critical cut-off of 0.70, indicating overall adequate to high internal consistency.

Convergent/divergent validity

Correlations between the SQ-48 subscales and other instruments are shown in Table 7.4.

Table 7.4: Correlations between the subscales of the SQ-48 in all 758 subjects.

Scale	MADRS	INH	BAS	BSI
MOOD	0.77	0.61	0.69	0.82
ANXI	0.73	0.52	0.72	0.84
SOMA	0.51	0.35	0.53	0.59
AGOR	0.47	0.42	0.50	0.58
AGGR	0.47	0.32	0.50	0.60
COGN	0.73	0.54	0.64	0.82
SOPH	0.60	0.49	0.57	0.77
VITA	–0.64	–0.56	–0.59	–0.66

Data are Spearman's (rho) correlations coefficients are presented. All P-values <0.05.
BSI denotes the short-form Brief Symptom Inventory, BAS denotes Brief Anxiety Scale, INH denotes the scale assessing psychomotor inhibition, and MADRS denotes Montgomery-Åsberg Depression Rating Scale. MOOD denotes Depression, ANXI denotes Anxiety, SOMA denotes Somatization, AGOR denotes Agoraphobia, AGGR denotes Aggression, COGN denotes Cognitive problems, SOPH denotes Social Phobia, VITA denotes Vitality. Because of adjustments made to the final version of instructions to the WORK subscale, insufficient data were available as yet.

In line with its coverage of depression-related symptomatology, the MADRS was most strongly correlated with the MOOD subscale (ϱ=0.77), the ANXI and COGN subscales (ϱ=0.73), and the VITA subscale (ϱ=0.64). In line with its anxiety-related symptomatology, the BAS was most strongly correlated with the ANXI subscale (0.72), the MOOD subscale (ϱ=0.69) and the COGN subscale (ϱ=0.64). The INH scale was most strongly correlated with the MOOD subscale (ϱ=0.61) and VITA subscale (ϱ=0.56), in line with its presumed role in both depression and its counterpart vitality. The BSI was moderately to strongly correlated with all subscales, indicating that all subscales are associated with overall psychopathology severity.

Reference values

Finally, percentiles and mean values on the SQ-48 subscales in the ROM reference (n=516) - and patient (n=242) groups are shown in Table 7.5.

Table 7.5 shows the following P_{95} cut-off values for the subscales, i.e., MOOD-8.0; ANXI-11.2; SOMA-8.0; AGOR-2.0; AGGR-5.0; COGN-11.0; SOPH-9.0; and VITA-15.0. These cut-off points are more conservative (with higher specificity but lower sensitivity for MAS disorders) than ROC cut-off points (AUC). Table 5 also shows the cut-off values with almost equal (optimal) sensitivity and specificity values, i.e., MOOD-4.0 (0.91); ANXI-6.5 (0.88); SOMA-1.5 (0.74); AGOR-0.5 (0.75); AGGR-1.5 (0.74); COGN-7.5 (0.89); SOPH-3.5 (0.83); and VITA-10.5 (0.87). Because of adjustments made to the final version of instructions to the WORK subscale, insufficient data were available as yet. These adjustments were related to the fact that many patients no longer worked or could no longer work. As a result, there were also no sufficient data available regarding the total scale of the SQ-48.

Table 7.5: 95th percentiles, mean, and cut off values (with their accompanying sensitivity and specificity) of the ROM reference (n=516) and patient (n=242) groups for subscales of the SQ-48.

	Reference group (n=516)				Receiver Operating Characteristics analysis				Patient group (n=242)
	Mean ± SD	P_{95}	Sensiitivity	Specivicity	Optimal cut-off	AUC	Sensitvity	Specificity	Mean ± SD
Aggression (AGGR)	1.20 ± 1.65	5.0	0.31	0.96	1.5	0.74	0.66	0.70	3.70 ± 3.35
Agoraphobia (AGOR)	0.38 ± 1.14	2.0	0.49	0.94	0.5	0.75	0.62	0.83	2.79 ± 3.59
Anxiety (ANXI)	3.80 ± 3.86	11.2	0.59	0.96	6.5	0.88	0.80	0.80	12.14 ± 5.77
Cognitive complaints (COGN)	4.08 ± 3.40	11.0	0.55	0.96	7.5	0.89	0.79	0.83	11.25 ± 4.57
Depression (MOOD)	2.14 ± 2.50	8.0	0.68	0.96	4.0	0.91	0.86	0.84	11.32 ± 6.27
Somatic complaints (SOMA)	1.71 ± 3.05	8.0	0.34	0.95	1.5	0.74	0.72	0.66	6.25 ± 6.22
Social phobia (SOPH)	2.36 ± 3.02	9.0	0.44	0.96	3.5	0.83	0.79	0.73	8.09 ± 5.06
Vitality/optimism (VITA)	7.66 ± 3.93	15.0	0.54	0.95	10.5	0.87	0.80	0.80	15.04 ± 4.99

In the Receiver Operating Characteristic (ROC) analysis, the optimal cut-off was considered when the sensitivity was equal to the specificity.
Because of adjustments made to the final version of instructions to the WORK subscale, insufficient data were available as yet.

DISCUSSION

The primary purpose of this study was to construct a psychometrically sound self-report measure for psychopathology (depression, anxiety, somatization, agoraphobia, aggression, cognitive problems, social phobia), which also measures vitality and work functioning. The main advantages of the present study were the use of two samples of both patients and non-patients, as well as the broad composition of a naturalistic outpatient population with mood, anxiety, and somatoform disorders.

The resulting SQ-48 is a multidimensional scale with good internal consistency and validity. Our results also indicated that – as intended – the two samples represent quite different populations, which makes them suitable to test the generalizability of the SQ-48 psychometric properties across different population strata. Also, the large range of correlations in both groups indicated that there is sufficient differentiation across the nine subscales.

Outcome assessment is essential in order to determine treatment effectiveness. Preferably, outcome assessment should be implemented as part of an outcomes evaluation programme [67].With the rapidly growing dissemination of computer-based assessment and feedback tools, the monitoring of psychotherapeutic processes and patients' outcome is becoming feasible in routine clinical practice [33,68]. ROM, in the sense of continuous monitoring of patient progress, requires valid measures, which are sensitive to change but also allow inexpensive repeated assessment [4,33]. In this context, self-report questionnaires are a cost-effective option, because they are inexpensive in terms of professional time needed for administration.

The clinical relevance of self-report measures has been demonstrated [6]. A major problem, however, is the fact that the licence policy of many self-report questionnaires is often restrictive. As a result, computerized assessments may not be possible due to copyright regulations. In addition, there can be serious economic obstacles to frequent assessments for the patient [33]. The SQ-48 is partly developed to overcome these problems, and can be used as a public domain questionnaire in both mental health care and general health care. As an example of the latter, Lee et al. [69] described the use of routine distress screening of newly admitted patients to an acute haematology and oncology ward.

Most scales of the SQ-48 measure psychopathology or psychological distress. Psychological distress also incorporates other nonspecific psychological manifestations, has stronger relations with common psychosocial factors, and tends to be milder and more transient than for instance depression [70]. A growing number of studies place specific emphasis on the need to expand the focus from only negative mental health (symptom- or distress-based outcome measures) to also positive mental health [67,71]. Examples of positive mental health outcome measures are work functioning, vitality, dispositional optimism. For this reason, the SQ-48 also assesses work functioning and vitality. Our results showed low correlations between work functioning and vitality, indicating a clear subscale

differentiation. For clinicians it may be helpful to focus on both reducing psychopathology and promoting positive emotions, skills, and engagement with life [29]. Further research in this area is worthwhile.

The endorsing reliability and validity evidence as produced by this study justifies further research on the psychometric properties and utility of the SQ-48. In particular, future research can be pursued in the following five directions (see also [72]. First, cross-validation of the factor analytic solution in an independent sample would enhance confidence in the nine factor structure which was found. Second, it would be informative to compare the results of the SQ-48 to other relevant instruments like for instance the OQ-45, MASQ(-D30). A related research area concerns the possible association between psychological distress measured by the SQ-48 on the one hand and quality of life on the other hand [73]. Third, it would be useful to study the temporal stability of the SQ-48 in a community sample by means of test-retest reliability at for instance one-month interval. Fourth, additional research could explore possible intergroup differences in levels of psychological distress as measured by the SQ-48. More specifically, research could focus on possible differences in psychological distress as a result of for example gender and age [5,74]. Fifth, further research could also determine whether the SQ-48 is useful in predicting treatment outcome.

In summary, the SQ-48 provides a broad and comprehensive survey of psychological distress as well as vitality and work functioning. It has satisfactory psychometric properties and therefore can be used in clinical, research and service settings. Further testing of the utility and validity of the SQ-48 (Dutch and English version) is planned by our department of Psychiatry, including assessment of its use in other cultural settings, psychiatric inpatients, and other diagnostic categories such as personality disorders. Further research is also planned to determine whether the SQ-48 is suitable for measuring changes in symptoms during the course of treatment. Finally, additional data will be collected regarding the subscale WORK and the SQ-48 total scale.

Reference List

1. Woodworth RS. (1918) Personal Data Sheet. Chicago, Stoelting.

2. Derogatis LR, Melisaratos N. (1983) The Brief Symptom Inventory: an introductory report. *Psychol Med*, 13 (3), 595-605.

3. Dohrenwend BP, Shrout PE, Egri G, Mendelsohn FS. (1980) Nonspecific psychological distress and other dimensions of psychopathology. Measures for use in the general population. *Arch Gen Psychiatry,* 37 (11), 1229-1236.

4. Luckett T, Butow PN, King MT, Oguchi M, Heading G, Hackl NA, Rankin N, Price MA. (2010) A review and recommendations for optimal outcome measures of anxiety, depression and general distress in studies evaluating psychosocial interventions for English-speaking adults with heterogeneous cancer diagnoses. *Support Care Cancer*, 18 (10), 1241-1262.

5. Ritsner M, Ponizovsky A. (1998) Psychological symptoms among an immigrant population: a prevalence study. Compr Psychiatry, 39 (1), 21-27.

6. Ritsner M, Modai I, Ponizovsky A. (2002) Assessing psychological distress in psychiatric patients: validation of the Talbieh Brief Distress *Inventory. Compr Psychiatry,* 43 (3), 229-234.

7. Derogatis LR, Lipman RS, Covi L. (1973) SCL-90: an outpatient psychiatric rating scale--preliminary report. *Psychopharmacol Bull,* 9 (1), 13-28.

8. Derogatis LR. (2001) Brief Symptom Inventory (BSI)-18: Administration, scoring and procedures manual. Minneapolis, MN, NCS Pearson.

9. Fichter MM, Elton M, Diallina M, Koptagel-Ilal G, Fthenakis WE, Weyerer S. (1988) Mental illness in Greek and Turkish adolescents. *Eur Arch Psychiatry Neurol Sci*, 237 (3), 125-134.

10. Lee MB, Lee YJ, Yen LL, Lin MH, Lue BH. (1990) Reliability and validity of using a Brief Psychiatric Symptom Rating Scale in clinical practice. *J Formos Med Assoc*, 89 (12), 1081-1087.

11. Ritsner M, Rabinowitz J, Slyuzberg M. (1995) The Talbieh Brief Distress Inventory: a brief instrument to measure psychological distress among immigrants. *Compr Psychiatry*, 36 (6), 448-453.

12. Watson D, Weber K, Assenheimer JS, Clark LA, Strauss ME, McCormick RA. (1995) Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *J Abnorm Psychol*, 104 (1), 3-14.

13. Watson D, Clark LA, Weber K, Assenheimer JS, Strauss ME, McCormick RA. (1995) Testing a tripartite model: II. Exploring the symptom structure of anxiety and depression in student, adult, and patient samples. *J Abnorm Psychol*, 104 (1), 15-25.

14. Wardenaar KJ, van VT, Giltay EJ, de BE, Penninx BW, Zitman FG. (2010) Development and validation of a 30-item short adaptation of the Mood and Anxiety Symptoms Questionnaire (MASQ). *Psychiatry Res*, 179 (1), 101-106.

15. Coyne JC. (1994) Self-reported distress: analog or Ersatz depression? *Psychol Bull,* 116 (1), 29-45.

16. Boulet J, Boss MW. (1991) Reliabiability and validity of the Brief Symptom Inventory. *Journal of Consulting and Clinical Psychology,* 3 (3), 433-437.

17. Oakley Browne MA, Wells JE, Scott KM, McGee MA. (2010) The Kessler Psychological Distress Scale in Te Rau Hinengaro: the New Zealand Mental Health Survey. *Aust N Z J Psychiatry*, 44 (4), 314-322.

18. Schwannauer M, Chetwynd P. (2007) The Brief Symptom Inventory: A validity study in two independent Scottish samples. *Clinical Psychology and Psychotherapy,* 14, 221-228.

19. Awad AG, Voruganti LNP, Heslegrave RJ, Hogan TP. (1996) Assessment of the patient's subjective experience in acute neuroleptic treatment: Implications for compliance and outcome. *International Clinical Psychopharmacology*, 11, 55-59.

20. Buchanan A. (1996) Compliance with treatment in schizophrenia. Londen, UK, Maudsley Monograph, Psychology Press.

21. Kalman TP. (1993) An overview of patient satisfaction with psychiatric treatment. Hospital & Community Psychiatry, 34, 48-54.

22. Barker DA, Orrell MW. (1999) The psychiatric care satisfaction questionnaire: a reliability and validity study. *Social Psychiatry and Psychiatric Epidemiology*, 34 (2), 111-116.

23. Feldman D, Rabinowitz J, Ben YY. (1995) Detecting psychological distress among patients attending secondary health care clinics. Self-report and physician rating. *Gen Hosp Psychiatry*, 17 (6), 425-432.

24. Laatsch L, Shahani BT. (1996) The relationship between age, gender and psychological distress in rehabilitation inpatients. *Disabil Rehabil,* 18 (12), 604-608.

25. Giltay EJ, Geleijnse JM, Zitman FG, Hoekstra T, Schouten EG. (2004) Dispositional optimism and all-cause and cardiovascular mortality in a prospective cohort of elderly elderly dutch men and women. *Arch Gen Psychiatry,* 61 (11), 1126-1135.

26. Van Fenema EM, Van der Wee NJ, Giltay EJ, Den Hollander-Gijsman ME, Zitman FG. (2012) Vitality predicts level of guideline-concordant care in routine treatment of mood, anxiety and somatoform disorders. *J Eval Clin Pract*, 18 (2), 441-448.

27. Burdick KE, Goldberg JF, Harrow M. (2010) Neurocognitive dysfunction and psychosocial outcome in patients with bipolar I disorder at 15-year follow-up. *Acta Psychiatr Scand,* 122 (6), 499-506.

28. Verboom CE, Sentse M, Sijtsema JJ, Nolen WA, Ormel J, Penninx BW. (2011) Explaining heterogeneity in disability with major depressive disorder: effects of personal and environmental characteristics. *J Affect Disord*, 132 (1-2), 71-81.

29. Kubzansky LD, Thurston RC. (2007) Emotional vitality and incident coronary heart disease: benefits of healthy psychological functioning. *Arch Gen Psychiatry,* 64 (12), 1393-1401.

30. Rozanski A, Kubzansky LD. (2005) Psychologic functioning and physical health: a paradigm of flexibility. *Psychosom Med*, 67 Suppl 1, S47-S53.

31. Wyshak G. (2003) Health ratings in relation to illnesses, physical functioning, general mental health and well-being: self-reports of college alumnae, ages <40-80 and older. *Qual Life Res,* 12 (6), 667-674.

32. Giltay EJ, Zitman FG, Kromhout D. (2006) Dispositional optimism and the risk of depressive symptoms during 15 years of follow-up: the Zutphen Elderly Study. *J Affect Disord*, 91 (1), 45-52.

33. Moessner M, Gallas C, Haug S, Kordy H. (2010) The clinical psychological diagnostic system (KPD-38): sensitivity to change and validity of a self-report instrument for outcome monitoring and quality assurance. *Clin Psychol Psychother,* 18 (4), 331-338.

34. Carlier IVE, Meuldijk D, Van Vliet IM, Van Fenema EM, Van der Wee NJ, Zitman FG. (2012) Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *J Eval Clin Pract,* 18 (1), 104-110.

35. De Beurs E, Den Hollander-Gijsman ME, Van Rood YR, Van der Wee NJ, Giltay EJ, Van Noorden MS, Van der Lem R, Van Fenema EM, Zitman FG. (2010) Routine outcome monitoring in the Netherlands: practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clin Psychol Psychother*, 18, 1-12.

36. Van Noorden MS, Giltay EJ, Den Hollander-Gijsman ME, Van der Wee NJ, Van Veen T, Zitman FG. (2010) Gender differences in clinical characteristics in a naturalistic sample of depressive outpatients: the Leiden Routine Outcome Monitoring Study. *J Affect Disord,* 125 (1-3), 116-123.

37. Schulte-van Maaren YWM, Carlier IVE, Giltay EJ, Van Noorden MS, De Waal MW, Van der Wee NJ, Zitman FG. (2012) Reference values for mental health assessment instruments: objectives and methods of the Leiden Routine Outcome Monitoring Study. *J Eval Clin Pract*.

38. American Psychiatric Association. (2000) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Washington, DC, American Psychiatric Association.

39. Ando T, Hashiro M, Noda K, Adachi J, Hosoya R, Kamide R, Ishikawa T, Komaki G. (2006) Development and validation of the psychosomatic scale for atopic dermatitis in adults. *Journal of Dermatology,* 33 (7), 439-450.

40. Diamond G, Levy S, Bevans KB, Fein JA, Wintersteen MB, Tien A, Creed T. (2010) Development, validation, and utility of internet-based, behavioral health screen for adolescents. *Pediatrics*, 126 (1), e163-e170.

41. Johnson I, Kebir O, Ben AO, Dellagi L, Rabah Y, Tabbane K. (2009) The self-assessment scale of cognitive complaints in schizophrenia: a validation study in Tunisian population. *BMC Psychiatry,* 9, 66.

42. Lee MB, Liao SC, Lee YJ, Wu CH, Tseng MC, Gau SF, Rau CL. (2003) Development and verification of validity and reliability of a short screening instrument to identify psychiatric morbidity. *J Formos Med Assoc,* 102 (10), 687-694.

43. Mohr C, Tonge BJ, Einfeld SL. (2005) The development of a new measure for the assessment of psychopathology in adults with intellectual disability. *J Intellect Disabil Res,* 49 (Pt 7), 469-480.

44. Morlock RJ, Williams VS, Cappelleri JC, Harness J, Fehnel SE, Endicott J, Feltner D. (2008) Development and evaluation of the Daily Assessment of Symptoms - Anxiety (DAS-A) scale to evaluate onset of symptom relief in patients with generalized anxiety disorder. *J Psychiatr Res,* 42 (12), 1024-1036.

45. van KD, de BE, Andrea H. (2008) A short form of the Dimensional Assessment of Personality Pathology-Basic Questionnaire (DAPP-BQ): the DAPP-SF. *Psychiatry Res*, 160 (1), 115-128.

46. Razavi D, Gandek B. (1998) Testing Dutch and French translations of the SF-36 Health Survey among Belgian angina patients. *J Clin Epidemiol,* 51 (11), 975-981.

47. Lambert MJ, Hansen LA, Umpress V, Lunnen K, Okiishi J, Burlingame GM, et al. (1996) Outcome Quetionnaire (OQ-45). Salt Lake City, American Professional Credentialing Services L.L.C.

48. Van Zuuren FJ. (1988) The fear questionnaire. Some data on validity, reliability and layout. *Br J Psychiatry,* 153, 659-662.

49. Buss AH, Perry M. (1992) The aggression questionnaire. *J Pers Soc Psychol,* 63 (3), 452-459.

50. Richman LS, Kubzansky LD, Maselko J, Ackerson LK, Bauer M. (2009) The relationship between mental vitality and cardiovascular health. *Psychol Health,* 24 (8), 919-932.

51. Mundt JC, Marks IM, Shear MK, Greist JH. (2002) The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *British Journal of Psychiatry,* 180, 461-464.

52. Niv N, Cohen AN, Sullivan G, Young AS. (2007) The MIRECC version of the Global Assessment of Functioning scale: reliability and validity. *Psychiatr Serv,* 58 (4), 529-535.

53. De Waal MW, Arnold IA, Spinhoven P, Eekhof JA, Assendelft WJ, van Hemert AM. (2009) The role of comorbidity in the detection of psychiatric disorders with checklists for mental and physical symptoms in primary care. *Soc Psychiatry Psychiatr Epidemiol*, 44 (1), 78-85.

54. Scheier MF, Carver CS, Bridges MW. (1994) Distinguishing Optimism from Neuroticism (and Trait Anxiety, Self-Mastery, and Self-Esteem) - A Reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology,* 67 (6), 1063-1078.

55. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25 (24), 3186-3191.

56. McKenna SP, Doward LC. (2005) The translation and cultural adaptation of patient-reported outcome measures. *Value Health*, 8 (2), 89-91.

57. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P. (2005) Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health,* 8 (2), 94-104.

58. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC. (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry,* 59 Suppl 20, 22-33.

59. Montgomery SA, Asberg M. (1979) New Depression Scale Designed to be Sensitive to Change. *British Journal of Psychiatry*, 134 (APR), 382-389.

60. Tyrer P, Owen RT, Cicchetti DV. (1984) The brief scale for anxiety: a subdivision of the comprehensive psychopathological rating scale. *J Neurol Neurosurg Psychiatry*, 47 (9), 970-975.

61. Asberg M, Montgomery SA, Perris C, Schalling D, Sedvall G. (1978) Comprehensive Psychopathological Rating-Scale. *Acta Psychiatrica Scandinavica*, 5-27.

62. Goekoop JG, Knoppert-Van der Klein EA, Hoeksema T, Klinkhamer RA, Van Gaalen HA, Van der Velde EA. (1991) The interrater reliability of a Dutch version of the Comprehensive Psychopathological Rating Scale. *Acta Psychiatr Scand*, 83 (3), 202-205.

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

64. Satorra A, Bentler PM. (1988) Scaling Corrections for Chi-square Statistics in Covariance Structure Analysis. Proceedings of the Business and Economic Statistics Section. Alexandria, VA, American Statistical Association.

65. Bentler PM. (2006) EQS6 Structural Equations Program Manual. Enrico CA, Multivariate Software Inc.

66. Brown TA. (2006) Confirmatory Factor Analysis for Applied Research. New York, The Guilford Press.

67. Zimmerman M, Ruggero CJ, Chelminski I, Young D, Posternak MA, Friedman M, Boerescu D, Attiullah N. (2006) Developing brief scales for use in clinical practice: the reliability and validity of single-item self-report measures of depression symptom severity, psychosocial impairment due to depression, and quality of life. *J Clin Psychiatry*, 67 (10), 1536-1541.

68. Percevic R, Lambert MJ, Kordy H. (2004) Computer-supported monitoring of patient treatment response. *J Clin Psychol*, 60 (3), 285-299.

69. Lee SJ, Katona LJ, De Bono SE, Lewis KL. (2010) Routine screening for psychological distress on an Australian inpatient haematology and oncology ward: impact on use of psychosocial services. *Med J Aust,* 193 (5 Suppl), S74-S78.

70. Coyne JC, Schwenk TL. (1997) The relationship of distress to mood disturbance in primary care and psychiatric populations. J *Consult Clin Psychol,* 65 (1), 161-168.

71. Keller MB. (2003) Past, present, and future directions for defining optimal treatment outcome in depression: remission and beyond. *JAMA*, 289 (23), 3152-3160.

72. Carey KB, Roberts LJ, Kivlahan DR, Carey MP, Neal DJ. (2004) Problems assessment for substance using psychiatric patients: development and initial psychometric evaluation. *Drug Alcohol Depend,* 75 (1), 67-77.

73. Fellinger J, Holzinger D, Dobner U, Gerich J, Lehner R, Lenz G, Goldberg D. (2005) Mental distress and quality of life in a deaf population. *Soc Psychiatry Psychiatr Epidemiol,* 40 (9), 737-742.

74. Ponizovsky A, Ginath Y, Durst R, Wondimeneh B, Safro S, Minuchin-Itzigson S, Ritsner M. (1998) Psychological distress among Ethiopian and Russian Jewish immigrants to Israel: a cross-cultural study. *Int J Soc Psychiatry*, 44 (1), 35-45.

Supplementary Material

Naam patiënt: Datum: Nummer: Geboortedatum:

Instructie:
Probeer de volgende stellingen eerlijk en accuraat te beantwoorden. Geef aan wat op u van toepassing is. Er zijn geen 'goede' of 'foute' antwoorden. U geeft het antwoord dat het beste uitdrukt hoe vaak u zich **de afgelopen week, met vandaag** erbij, zo hebt gevoeld. Wat het eerste in u opkomt, is vaak het beste .

NB: Indien u niet werkt of studeert, of indien u dat de afgelopen week niet hebt kunnen doen, dan kunt u de volgende vragen overslaan: 9, 15, 20, 30 en 35.

HOEVEEL LAST HAD U VAN:	Nooit	Zelden	Soms	Vaak	Zeer Vaak
1. Ik was kortademig zonder dat ik mij inspande	0	1	2	3	4
2. Ik voelde mij vertraagd of langzaam	0	1	2	3	4
3. Ik was ontevreden.	0	1	2	3	4
4. Ik werd angstig in een menigte van mensen	0	1	2	3	4
5. Ik had hartkloppingen.	0	1	2	3	4
6. Ik had moeite met het nemen van beslissingen.	0	1	2	3	4
7. Ik kon nergens van genieten.	0	1	2	3	4
8. Ik durfde open ruimtes, zoals een plein, niet over te steken.	0	1	2	3	4
9. Ik voelde stress op mijn werk of studie.	0	1	2	3	4
10. Ik had onenigheid met anderen.	0	1	2	3	4
11. Ik voelde pijn of druk op de borst.	0	1	2	3	4
12. Ik zag naar dingen uit.	0	1	2	3	4
13. Ik dacht aan mijn dood of zelfmoord.	0	1	2	3	4
14. Ik durfde niet alleen met het openbaar vervoer te reizen.	0	1	2	3	4
15. Mijn werk of studie gaf me geen voldoening.	0	1	2	3	4
16. Ik was opvliegend zonder aanleiding.	0	1	2	3	4
17. Ik voelde mij duizelig of licht in het hoofd.	0	1	2	3	4
18. Ik had zin om dingen te doen.	0	1	2	3	4
19. Ik had geen zin in het leven.	0	1	2	3	4
20. Ik had het gevoel dat ik teveel werkte of studeerde.	0	1	2	3	4
21. Ik had moeite om mijn woede te beheersen.	0	1	2	3	4
22. Ik voelde tintelingen, bijvoorbeeld in mijn handen.	0	1	2	3	4
23. Ik kon moeilijk voor mijn mening uitkomen.	0	1	2	3	4
24. Ik was bang of angstig.	0	1	2	3	4
25. Ik durfde niet alleen naar een drukke winkel te gaan..	0	1	2	3	4
26. Ik trilde of beefde.	0	1	2	3	4
27. Ik was bang om afgewezen te worden in een groep.	0	1	2	3	4

28. Ik was schrikachtig.	0 1 2 3 4
29. Ik was optimistisch over mijn toekomst.	0 1 2 3 4
30. Ik werkte of studeerde minder hard dan voorheen.	0 1 2 3 4
31. Ik voelde mij rillerig.	0 1 2 3 4
32. Ik voelde mij de mindere van anderen.	0 1 2 3 4
33. Ik was zenuwachtig en nerveus.	0 1 2 3 4
34. Ik had plannen of stelde mezelf doelen.	0 1 2 3 4
35. Ik had het gevoel dat het niet goed ging met mijn werk/studie.	0 1 2 3 4
36. Ik voelde mij ongemakkelijk als anderen naar mij keken.	0 1 2 3 4
37. Ik had interesse in dingen.	0 1 2 3 4
38. Ik voelde mij hopeloos.	0 1 2 3 4
39. Ik was vergeetachtig.	0 1 2 3 4
40. Ik voelde mij somber of depressief.	0 1 2 3 4
41. Ik voelde mij onrustig.	0 1 2 3 4
42. Ik voelde me energiek en levenslustig.	0 1 2 3 4
43. Ik wilde mensen het liefst slaan als dat werd uitgelokt.	0 1 2 3 4
44. Ik had moeite om op gang te komen.	0 1 2 3 4
45. Ik voelde mij onzeker in gezelschap.	0 1 2 3 4
46. Ik voelde mij gespannen.	0 1 2 3 4
47. Ik kon mij niet goed concentreren.	0 1 2 3 4
48. Ik piekerde.	0 1 2 3 4

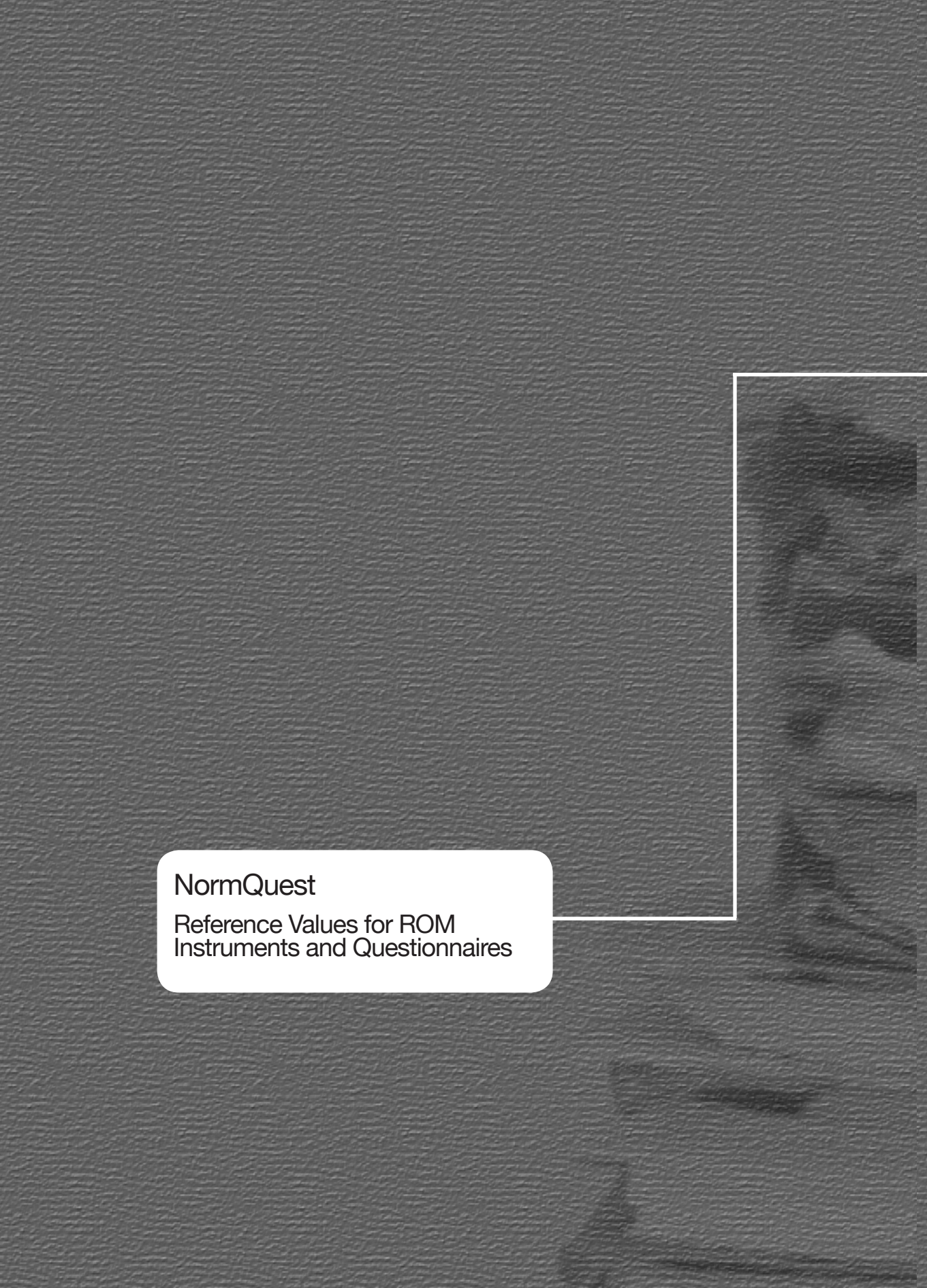
Name of the Patient:
Today's Date:
Number:
Date of Birth:

Instruction:
Try to answer the following propositions fairly and accurately. There are no 'right' or 'wrong' answers. Give the answer that best expresses the number of times you have felt the following ways **last week, including today**. The answer which comes to your mind first is often the best answer.

Note: If you did not work or study or have not been able to do so, then you can skip the questions 9, 15, 20, 30 and 35.

HOW MUCH TROUBLE DID YOU HAVE:	never	rarely	sometimes	often	very often
1. I was short of breath with minimal excursion.	0	1	2	3	4
2. I felt weak or slow.	0	1	2	3	4
3. I was irritable and dissatisfied.	0	1	2	3	4
4. I felt anxious while I was in a crowd (of people).	0	1	2	3	4
5. I felt palpitations.	0	1	2	3	4
6. I had trouble making decisions.	0	1	2	3	4
7. I could not enjoy anything at all.	0	1	2	3	4
8. I did not dare to cross open spaces, such as a public square.	0	1	2	3	4
9. I felt stressed at my work or study.	0	1	2	3	4
10. I argued with others.	0	1	2	3	4
11. I felt chest pain (or pressure).	0	1	2	3	4
12. I looked forward to things.	0	1	2	3	4
13. I considered my death or suicide.	0	1	2	3	4
14. I did not dare to travel on my own using public transport.	0	1	2	3	4
15. I was dissatisfied with my work or study.	0	1	2	3	4
16. I was hot-tempered without good reason.	0	1	2	3	4
17. I felt dizzy or lightheaded.	0	1	2	3	4
18. I felt like doing things.	0	1	2	3	4
19. I did not want to live anymore.	0	1	2	3	4
20. I had the feeling that I have been working or studying very hard.	0	1	2	3	4
21. I had trouble with controlling my anger.	0	1	2	3	4
22. I felt a tingling, for example in my hands.	0	1	2	3	4
23. I could hardly express myself.	0	1	2	3	4
24. I was afraid or anxious.	0	1	2	3	4
25. I did not dare to go alone to a crowded shop.	0	1	2	3	4
26. I was shaking or trembling.	0	1	2	3	4
27. I was afraid of rejection by others.	0	1	2	3	4
28. I was scared.	0	1	2	3	4
29. I was optimistic about my future.	0	1	2	3	4

30. I worked or studied less intensely than before.	0 1 2 3 4
31. I felt shaky or I had shivers.	0 1 2 3 4
32. I felt low and less than others.	0 1 2 3 4
33. I felt jittery and nervous.	0 1 2 3 4
34. I looked forward to my plans and goals for the future.	0 1 2 3 4
35. I had the feeling that I did not do well with my work or study.	0 1 2 3 4
36. I felt uncomfortable when other people looked at me.	0 1 2 3 4
37. I took interest in things.	0 1 2 3 4
38. I felt hopeless.	0 1 2 3 4
39. I was forgetful.	0 1 2 3 4
40. I felt down or depressed.	0 1 2 3 4
41. I felt restless.	0 1 2 3 4
42. I felt energetic and high-spirited.	0 1 2 3 4
43. I wanted to hit people if I was provoked.	0 1 2 3 4
44. I struggled to get the day started.	0 1 2 3 4
45. I felt insecure in the company of others.	0 1 2 3 4
46. I felt tense.	0 1 2 3 4
47. I could not concentrate well.	0 1 2 3 4
48. I worried.	0 1 2 3 4



NormQuest

Reference Values for ROM Instruments and Questionnaires

Chapter 8

Summary, general discussion, and conclusions

1. SUMMARY OF RESULTS

1.1 Aims of our study

The primary aim of the NormQuest study described in this thesis was to generate evidence-based, reference values for 19 self-report and observational questionnaires. The focus was on questionnaires measuring mood, anxiety, and somatoform (MAS) disorders used in Routine Outcome Monitoring (ROM). The set of cut-off values of the ROM reference group ('healthy') can be used in specialized mental health care by therapists to support the decision whether a patient is sufficiently recovered to be considered as a member of the healthy population, and no longer as a member of the patient population. These reference values are suitable as decision support for referral back to primary care physicians. Additionally, the set of the ROM patient group ('clinically ill') cut-off values can be used by primary care physicians as decision support for referral to the specialized mental health care. To allow determination of cut-off points for skewed distributions, percentile scores were used. In addition, we assessed the discriminative power of the questionnaire scores by means of Receiver Operating Characteristics (ROC) analyses. Finally, we calculated reference values in separate strata of gender and age.

The secondary aim of the NormQuest study concerned the need for the development of public domain questionnaires. In the NormQuest study, the generic Symptom Questionnaire-48 (SQ-48), aimed at broad applicability in patients with MAS disorders, was developed. Also, for the SQ-48 reference values were calculated.

1.2 Summary of major findings

This is the first study of this size carried out in the Netherlands to yield reference values for questionnaires measuring MAS disorders. Chapter 2 described the objectives, design, and methodologies. Two groups were included. The first group, the ROM patient group, comprised specialized mental health care (i.e., secondary care) outpatients with one or more MAS-disorders. Patients were screened as part of their routine intake procedure. For the NormQuest study, a group of 5269 outpatients, aged 18-65 years, with complete data were selected. The second group, the ROM reference group, comprised primary care patients, registered with one of 8 participating general practitioners (GPs) but not necessarily seeking treatment. They can be considered to constitute a general population sample since in the Netherlands 99.9% of the general population is registered with a GP [1]. The ROM reference group comprised 1302 participants, aged 18-65 years. The ROM reference group matched the ROM patient group in terms of gender-, age distribution, and the level of urbanization. Data were collected during a baseline assessment comprising a standardized diagnostic interview, administration of rating scales, and completion of several self-report questionnaires by the ROM reference group. For the ROM patient group the baseline assessment was part of the intake procedure. The interviewers were extensively trained and supervised, thus maximizing the inter-rater reliability and validity of the assessment.

In Chapters 3 to 7, we discussed the assessed reference values for the 19 questionnaires. All of the P_{95} ROM reference group and the P_5 patient group cut-off values are summarized in Tables 8.1 and 8.2 of the Appendix of this chapter.

In Chapter 3, reference values for four generic questionnaires were calculated: the Brief Symptom Inventory (BSI), the Mood & Anxiety Symptom Questionnaire – 30-item short adaptation of the MASQ, Dutch translation (MASQ-D30), the Short Form Health Survey 36 (SF-36), and the Dimensional Assessment of Personality Pathology - Short Form (DAPP-SF). Data from 1294 ROM reference group participants were compared with data from 5269 psychiatric outpatients of the ROM patient group. The P_{95} ROM reference group and the P_5 patient group cut-off values are summarized in Table 8.1. The data illustrate gender-specific results. There was a tendency for women in the ROM reference group to have somewhat higher cut-off scores on the BSI and MASQ-D30 than men in the ROM reference group, while the two genders had the opposite pattern of cut-off scores on the DAPP-SF. Men, especially young men, reported better health, reflected in higher scores on several subscales of the SF-36 than young women. The discriminative power of the BSI, MASQ-D30 and SF-36 was good, but it was poor for the DAPP-SF. All analyses of internal consistency were based on a combination of data from the ROM reference group and the ROM patient group. The internal consistency of the subscales ranged from adequate to excellent for all questionnaires.

From Chapter 4 onward, we focused on the reference values for disorder-specific questionnaires.

Chapter 4 concerned major depression, using the Beck Depression Inventory-II (BDI-II), the Inventory of Depressive Symptoms (Self-Report) (IDS-SR), and the Montgomery-Äsberg Depression Rating Scale (MADRS). We compared data from 1295 ROM reference group participants with data from 4627 patients of the ROM patient group diagnosed with major depressive disorder (MDD) or dysthymic disorder. Cut-off values (P_{95} ROM reference group) were significantly higher for women compared to men. The discriminative power of the BDI-II, IDS-SR, and MADRS scores was very high. The internal consistency was excellent for all total scores. For the subscales, internal consistency was satisfactory, with the exception of the IDS-SR subscale Atypical Characteristics, which was poor.

In Chapter 5, we discussed reference values for eight questionnaires measuring anxiety disorders: the Brief Scale for Anxiety (BSA), the PADUA Inventory Revised (PI-R), the Panic Appraisal Inventory (PAI) (with three subscales: the PAI Anticipated Panic, the PAI Perceived Consequences, and the PAI Perceived Self-Efficacy), the Penn State Worry Questionnaire (PSWQ), the Worry Domains Questionnaire (WDQ), the Social Interaction, the Anxiety Scale (SIAS), the Social Phobia Scale (SPS), and the Impact of Event Scale-Revised (IES-R). These questionnaires cover most of the DSM-IV anxiety disorders. We included 1295 ROM reference group participants and 5066 psychiatric outpatients of the ROM patient group diagnosed with at least one specific anxiety disorder. Reference values were generally higher for women than for men. The discriminative power of all eight

were generally higher for women than for men. The discriminative power of all eight questionnaires measuring anxiety disorders was very high. The internal consistency was excellent for the total scores and subscales of all questionnaires, except for the BSA and for the WDQ subscale Work Incompetence: they had adequate internal consistencies.
Chapter 6 included reference values for three disorder-specific questionnaires concerning some of the somatoform disorders: the Body Image Concern Inventory (BICI; for body dysmorphic disorder), the Whitely Index (WI; for hypochondriasis), and the Checklist Individual Strength (CIS20R; for chronic fatigue syndrome). Data were compared from 648 ROM reference group participants and 823 ROM patient group outpatients diagnosed with at least one somatoform disorder. Compared to the sizes of the groups in the previous chapters, the ROM reference group and the ROM patient group were smaller. Somatoform disorders are less prevalent compared to mood- and anxiety disorders. For the BICI, the WI, and the CIS20R total score, the cut-off values differed for men and women, again being higher for women. The discriminative power of all 3 questionnaires was very high and the internal consistency was excellent.

Chapter 7 described the development, validation and reference values of our newly developed public domain questionnaire, the 48-item Symptom Questionnaire (SQ-48). The SQ-48 was developed to be multidimensional, including the following nine subscales: Depression (MOOD, 6 items), Anxiety (ANXI, 6 items), Somatization (SOMA, 7 items), Agoraphobia (AGOR, 4 items), Aggression (AGGR, 4 items), Cognitive problems (COGN, 5 items), Social Phobia (SOPH, 5 items), Work functioning (WORK, 5 items), and Vitality (VITA, 6 items). A part of the ROM reference group (n=516) and a part of the ROM patient group with suspected depressive, anxiety, and somatoform disorders (n= 242) completed the SQ-48 plus a set of observer-rated and self-report scales (MINI-Plus, MADR, BSA, BSI). The discriminative power of the questionnaire was good. The results showed good internal consistency as well as good convergent and divergent validity. The SQ-48 is meant to be available in the public domain for Routine Outcome Monitoring (ROM).

In conclusion, for 19 generic and disorder-specific ROM questionnaires a comprehensive set of reference values was provided. These reference values may support responsible clinical decision-making with respect to initiating, adjusting, or terminating therapy, and with respect to referring patients from mental health care to primary care and vice versa. The main, clinically useful reference values are presented in Tables 8.1 and 8.2 of the Appendix.

2. GENERAL DISCUSSION

In this section, the findings of the NormQuest study will be discussed in a broader perspective. The first topic is the choice of percentile scores as reference values, where the distribution of data guided this choice. The reference group will be outlined, since it provides the characteristics needed for comparison and evaluation of the patient's characteristics

(i.e., severity of psychopathology). The next topic is the representativeness of the reference group and the generalizability of the results. Subsequently, we will elaborate on the implications of our findings for clinical practice, with practical recommendations for referral back to primary care and referral to specialized mental health care. The reference values in separate gender and age strata will be discussed, which followed a consistent pattern for all the questionnaires. We will comment on the discriminative powers of the questionnaire scores by means of ROC analyses. ROM questionnaires are appropriate for the assessment of symptom severity, but our findings suggest that they are also of some value for diagnostic purposes. Finally, the newly developed self-report questionnaire Symptom Questionnaire-48 (SQ-48) will be discussed. We will finish with recommendations for future research.

2.1 Reference values

Reference values of assessment tools are important for different clinical purposes, which were summarized by Solberg [2] as early detection of disease, differential diagnosis, and monitoring response to therapy . Since the questionnaires in this study are measures of symptom severity, rather than diagnostic tools, Solberg's last purpose is the most relevant for the use of our reference values. Other purposes of our reference values are: 1) screening of patients when they first seek treatment by the GP and supporting clinical decisions about possible referral to specialized mental health care; and 2) comparison of individual patients' scores with scores from a similar group (e.g., same gender, same disorder) in order to assess the severity of symptoms. The clinical use of the relevant reference values is described in section 2.4.

The concept of reference values of laboratory measures has been widely accepted in medicine, (e.g., glucose, total cholesterol, serum liver enzymes, and other biochemical analyses) [3-5]. Reference values are widely used in health care [4,6]. In psychiatry, however, reference values still need to be established and applied, to which aim the NormQuest study can contribute.

To derive valid reference values, the reference group needs to have specific characteristics. The COTAN (Commissie Testaangelegenheden Nederland), documentation from the Dutch Institute of Psychologists (NIP), is a leading grading system for test quality in the Netherlands [7]. The COTAN grading system suggests three criteria that are relevant in the context of reference groups. Firstly, the size and representativeness of the groups is evaluated. A group size of $N \geq 400$ is considered good, a group size of $300 \leq N < 400$ is considered adequate, and a group size of $N < 300$ is considered insufficient. We aimed for group sizes (including gender stratification) of at least $N \geq 300$ and succeeded for all generic, mood, and anxiety questionnaires. Thus, according to COTAN criteria, our group sizes for these questionnaires ranged from adequate to good. For the somatoform questionnaires the group sizes were smaller and therefore did, not meet the COTAN criteria. The representativeness of the ROM reference group is discussed in section 2.3. The second COTAN criterion evaluates psychometric measures (e.g., score distribution, means, and standard deviations). We have

met that criterion by providing percentile scores (in view of the skewed distributions), in addition to means and standard deviations, which we considered less appropriate because of the skewed data distributions. Thirdly, data on possible differences between subgroups need to be analyzed properly, according to the COTAN criteria. We used gender stratified sampling for the assessment of reference values for all questionnaires and age stratified sampling for the generic questionnaires and the questionnaires measuring major depression. Therefore we can conclude that our analyses fairly met the COTAN criteria.

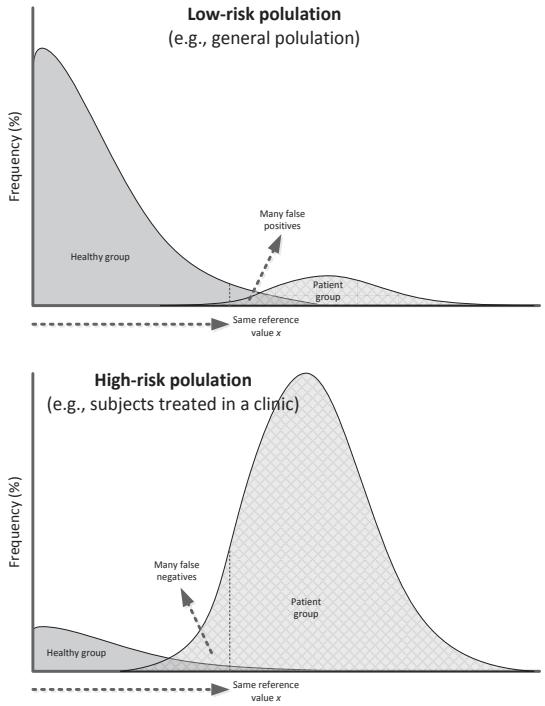


Figure 8.1. Prevalence-dependent cut-off values. Top: low prevalence of the psychiatric disorder: a fixed reference value x results in many false positives. Bottom: high prevalence of the psychiatric disorder in a high-risk population: the same fixed reference value x now results in many false negatives.

Prevalence rates will influence test characteristics of reference values. When the prevalence of the disease is low (i.e., in the general population), the P_{95} cut-off point discriminating clinically ill from healthy will lead to many false positives (see Figure 8.1, top). However, when the prevalence of the disease is high (as it is in a patient population), the same P_{95} cut-off point discriminating clinically ill from healthy will lead to many false negatives (see Figure 8.1, bottom). Therefore, clinicians should always use the test results in conjunction with their clinical judgment when making choices about treatment options and referral.

2.2 Reference group

A reference group consists of a sample of persons who are representative of the population for whom the test is intended. Reference values facilitate the comparison of the individual score to the distribution of scores in a population. The two populations considered in this study are specialized mental health (secondary) care patients (ROM patient group) and 'healthy' members of the general population (ROM reference group) [3,8].

Since our aim was the comparison of these two populations, we chose to ensure similar sociodemographic characteristics. We matched the ROM reference group with the ROM patient group in terms of gender- and age distribution, as well as level of urbanization. With our large ROM reference group of about 1300 persons, the subgroups stratified for gender and age were larger than the required minimum size of 120 that is considered to provide adequate power to yield reference values [9].

Individuals with current psychopathology were not excluded from the ROM reference group, as long as they were not treated in specialized mental health (secondary) care. As noted by Gräsbeck [10] "Absolute health does not exist. Some degree of pathology is present in every individual like entropy in a chemical system". Where reference values are derived from measurements of a so-called healthy population, the 'level of health' of the population should be specified, based upon the criteria for inclusion or exclusion of persons from the 'healthy' population. In this study we chose to not exclude any person, provided that in the past six months they received no treatment for psychiatric problems in specialized mental health (secondary) care. We support the argument made by Kendall et al., [11] that excluding participants with elevated levels of the target psychopathology from the reference group might lead to creating a non-representative, 'supernormal' sample. Comparing the patient group with such a supernormal group would represent an overly stringent criterion with unreasonable narrow reference intervals [12]. The statistical definition of normality is in line with Kendall's argument. This definition is based on the distribution of scores in the general population (including all individuals) [13] where disease is defined as a 'quantitative deviation from the normal' [14]. The statistical definition is opposed to the medical definition. This medical definition equates normality with health and thus with the absence of pathology, which is difficult to quantify [15]. By including all possible participants in the ROM reference group, this group also includes those who may currently be experiencing elevated levels of psychopathology, but are not being treated in specialized mental health (secondary) care.

Reference values are usually based on the middle 95% of the reference population, with the most outlying 5% defined as abnormal. Most often, these outlying observations are split evenly between the ends of the score distributions in the reference group, 2.5% at each end of the distribution. For the ROM questionnaires, only high values are of clinical concern. Therefore, we defined 5% of outlying observations at the high end of the distribution of the ROM reference group scores as abnormal (and 5% at the low end of the distribution for the 'inverted' subscales of the SF-36). This is in line with the practice in laboratory medicine [16]. With a similar argument, the 5% of outlying observations of the ROM patient group at

the low end of the distribution were by definition considered as clinically deviant from the patient population.

According to the MINI-Plus data, about 10% of the ROM reference group reported enough psychiatric symptoms to warrant (at least) one DSM-IV diagnosis. We noted a reduction in the P_{95} ROM reference group values when we excluded these 10% non-healthy subjects from the ROM reference group: for the four generic questionnaires (not for the SQ-48) the decrease was 5% of the P_{95} value [17]; for the three questionnaires measuring major depression the decrease was 15% [18]; for the eight questionnaires measuring anxiety disorders the decrease was 9% [19]; and for the three questionnaires measuring somatoform disorders the decrease was 7% [20].

2.3 Representativeness and generalizability

When deriving reference values, we aimed for generalizability and representativeness. The NormQuest sample was representative for the gender and age distributions of the ROM patient group. Random sampling among persons registered with the participating GPs was used as a strategy for ensuring representativeness. Indeed, in the Netherlands 99.9% of the general population is registered with a GP [1]. There was large variability for many of the demographic variables in the ROM reference group. To yield reference values this variability is recommended, as the reference values need to be applied to a wider population and external validity is required.

Representativeness is related to response rate. The response rate of the present NormQuest study was 37.1%. We used several methods to enhance the possible response rate. These efforts included offering participant-friendly interview conditions, such as choice of venue (at the homes of the participants , at the general practice, or at the academic center LUMC) and time (in the morning, afternoon, or evening), and a personal phone call for further information after an invitation by mail. We have compared the gender and age distributions between the non-respondents and participants. The response rate for women was slightly lower than the response rate for men, implying possible (greater) selective sampling in women. Slightly more persons aged 36-55 years responded compared to those aged 18-35 years. This suggests a slight under-representation of younger participants. Some other populations (i.e., younger fulltime employed men, or persons with (subthreshold) psychopathology) may also have been underrepresented. A possible mechanism of this selective sampling was the contacting of subjects by phone: few mobile phone numbers were registered in the GP information system, thus possibly missing many young subjects. Prevalence rates of mood- and anxiety disorders in 18-24 year olds and prevalence rates of mood disorders in 25-34 year olds are higher than prevalence rates in the older age groups [21]. So, this may have led to a slight underestimation of our reference values. A further possible mechanism of selective sampling was the exclusion of subjects by the GPs. They unlisted subjects who were not able to cope with the effort of the NormQuest interview. Some of those subjects possibly had elevated levels of psychopathology. Again this might have

resulted in a slight under-estimation of our reference values. However, it is also possible that participants, compared to persons who actively refused to participate, are likely to be more interested in their mental health, to be more eager to take actions that improve their health and to have a more favorable clinical course of symptoms [22]. It is unknown whether this has resulted in an overestimation or underestimation of our reference values. Furthermore, it might be relevant that self-report questionnaires are subject to response bias. Previous research suggested that there may be systematic gender differences in self-report bias, with men tending to minimize their depressive symptoms more than women [23]. Therefore, we may have underestimated the prevalence of depression in men, resulting in an under-estimation of reference values for men. In sum, despite our efforts, the ROM reference group may not have been fully representative of the general population. The possible total effect on the calculated reference values is hard to quantify.

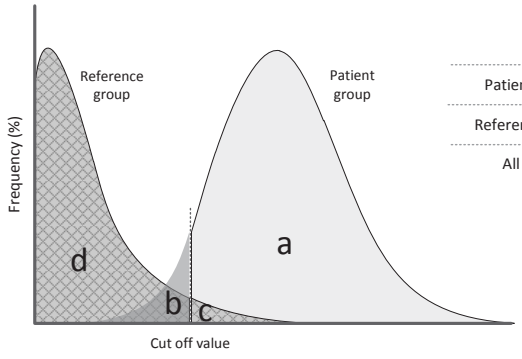
Generalizability of the reference values was another aim in this study. As noted before, the NormQuest sample was representative for the gender and age distributions of the ROM patient group. Therefore, its reference values can be validly used as a comparison against this patient group. However, several reference values calculated in this study differ from reference values in previous studies. In general, our reference values are slightly higher. Why do reference data differ so much over (internationally) different populations? Are the differences culture and language related, or are they design-related? Firstly, the perception of health and the ways health problems are expressed vary from culture to culture [24]. A conceptual distinction exists between disease and illness. Disease relates to malfunctioning or maladaptation of biologic and psychophysiologic processes; illness represents personal, interpersonal, and cultural reactions to disease or discomfort [24]. Semantics may vary between cultures and they may vary between international versions of a questionnaire [25,26]. Self-report questionnaires in particular may reflect the experience of illness and may therefore be culture sensitive. Secondly, design-related differences can emerge when the comparisons are made between 'cheese and chalk' (i.e., differences in terms of patient population [25,27,28], mode of questionnaire administration [25,28,29], socio-economic status [27,29], or clinical severity [27,29]). Especially levels of physical and psychological functioning have to be well-defined. Two versions of the same questionnaire can be equally sensitive to a given change in functional status, yet assign different scores to a given level of distress [27]. Furthermore, using a questionnaire in different national regions may lead to differences: health status may vary by area of residence [28]. Our reference values are regional ones (province of South-Holland). Generalizability to the national level might be not entirely obvious. Further research could legitimize this generalizability. These reference values are appropriate for outpatients referred for MAS disorders. Some caution is appropriate with other patient populations, e.g., inpatients, psychotic or Severe-Mental-Illness-patients, or patients with personality disorders as main diagnosis.

2.4 Clinical use of the reference values

This study yielded reference values, including cut-off values. Reference values allow the determination of the position of the patient in the distribution of the total population as a measure of symptom severity. Reference values can help to indicate when the patient is sufficiently recovered to make a next step in the treatment. In particular, reference values can help to assess whether therapy has moved someone outside the range of the patient population and within the range of the reference population. Clinicians in specialized mental health care can use certain cut-off values to support their decisions concerning the end of treatment and possible referral back to primary care. Vice versa, general practitioners (GPs; primary care) can use a different set of cut-off values to support their decision about referral to specialized mental health (secondary) care. Thus, the choice of cut-off values depends on the purpose for which the cut-off values will be used.

Sensitivity and specificity vary with different cut-off values. Figure 8.2 depicts the proportions of the ROM reference group and of the ROM patient group that scored higher than a certain cut-off value and lower than this cut-off value.

When referral from secondary care to primary care is at order, 'health' is the condition that is to be detected. A cut-off value with high sensitivity for symptomatic health is advised. The proportion of the ROM reference group scoring lower than the cut-off value (d/M_0) will be maximal; the proportion of real patients scoring lower than the cut-off value (b/M_1) will be maximal as well. As we discussed previously, high sensitivity to health is associated with low sensitivity to establish disease: a/M1 is minimal. Vice versa, when referral from primary care to secondary care is at order, 'disease' is the condition that is to be detected. A cut-off value with high sensitivity for disease is advised. The proportion of real patients scoring higher than the cut-off value (a/M_1) will be maximal; the proportion of the ROM reference group scoring higher than the cut-off value (c/M_0) is then maximal as well. High sensitivity to disease is associated with low sensitivity to symptomatic health: d/M_0 is minimal.



	Higher than cut-off	Lower than cut-off	Total
Patient group	a	b	M_1
Reference group	c	d	M_0
All subjects	N_1	N_0	T

Referral from secondary care **to primary care**:
$Sensitivity_{health} = d / M_0$ $Specificity_{health} = a / M_1$

Referral from primary care **to secondary care**:
$Sensitivity_{disease} = a / M_1$ $Specificity_{disease} = d / M_0$

Figure 8.2: The choice of the reference value will determine the sensitivity or specificity of the test, with a trade-off between the two. Sensitivity or specificity also depends on whether health or disease is being assessed. Depicted are proportions of the ROM reference group and of the ROM patient group that scored higher than a certain cut-off value and lower than this cut-off value.

Decisions concerning the end of treatment and possible referral back to primary care

This paragraph is meant for specialized mental health (secondary) care clinicians in order to support their decisions concerning the end of treatment and possible referral back to primary care.

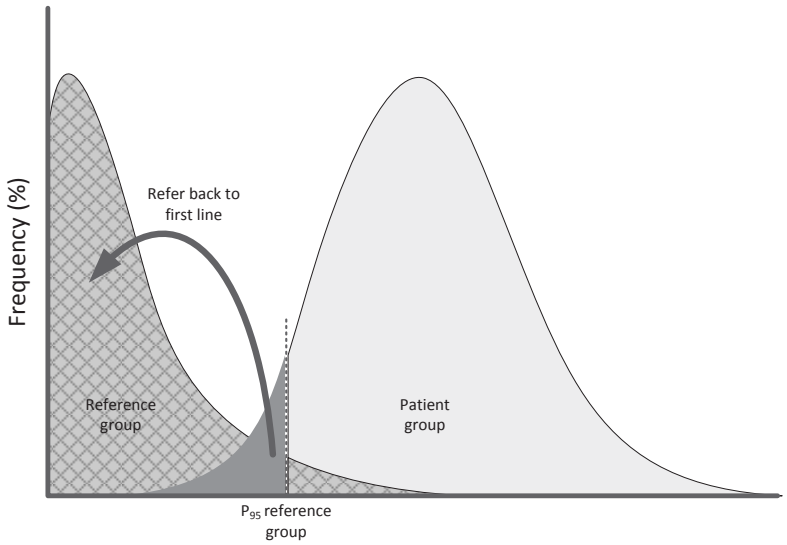


Figure 8.3: Cut-off values relevant for referral back to primary care. Patients depart from treatment when they no longer belong to the patient population, but belong to the reference population instead, below the cut-off value P_{95} ROM reference group.

It can be argued that patients enter treatment when they are part of a patient (clinically ill) population and they depart from treatment when they no longer belong to that population, but belong to the reference ('healthy') population. Referral back to primary care might be indicated when the patient in specialized mental health care has become similar to the reference population (i.e., belongs to the 95% normality range of the ROM reference group). In order to support decisions regarding back referral, a cut-off point can be used. The clinically relevant cut-off point is the point that the patient has to cross at the time of the post-treatment assessment in order to be classified as changed to a clinically significant degree of functionality or health. As can be seen in Figure 8.3, the cut-off value, marking the top 5% of the ROM reference group, is equivalent to the 95th percentile score: P_{95} ROM reference group. This cut-off value is highly sensitive to symptomatic health. It can be considered as a reliable indicator of symptomatic health, since it rarely misses health among those who are actually healthy. However, high sensitivity to health is associated with low sensitivity to establish disease. The cost of low sensitivity to disease or many false negative results might be false reassurance about the absence of disease [30].

Referral back to primary care might be indicated even when the patient in specialized mental health care still has some residual symptoms. Indeed, a substantial part of primary care patients are not without symptoms. Furthermore, referral back to primary care might be indicated for patients with recurrent depression for treatment of any residual anxiety symptoms. [30].

In Table 8.1 of the Appendix the cut-off values, i.e., the P_{95} ROM reference group values, are summarized for the 19 ROM questionnaires. Four sets of questionnaires are available: 1) generic questionnaires; 2) questionnaires measuring mood disorders; 3) questionnaires measuring anxiety disorders; and 4) questionnaires measuring somatoform disorders. When comparing the P_{95} ROM reference group cut-off values with the few cut-off values that were previously published, our values were generally higher. Thus, when our P_{95} ROM reference group cut-off values are used a patient will be eligible for referral back to primary care having more residual symptoms than would be the case if previously published cut-off values were used. Previously published cut-off values were established in groups of recovering patients [31-33] and in control groups with no life-time personal history of psychopathology [13,34]. For the groups of recovering patients in these studies [31-33], the cut-off value was defined as the point of remission, with the total absence of significant signs or symptoms [31-33]. It seems to imply circularity to establish a reference group based on the amount of symptoms. This procedure may have resulted in lower cut-off values compared to our cut-off values, based on patients with some residual symptoms. For the control groups with no life-time personal history the medical definition of normality was used [13,34] thus creating a control group comprising 'supernormal' participants (see section 2.2). Again, this resulted in lower cut-off values compared to our cut-off values, which were based on a reference group with 10% non-healthy subjects (see section 2.2). In yet another control group study, the derived cut-off value provided a high sensitivity (and a lower specificity) [35]. Our P_{95} ROM reference group cut-off values were related to low sensitivity to disease (and high specificity; see section 2.4.1) and therefore they were higher than the previously published values.

Practicing therapists may have specialized mental health (secondary) care patients with continuous high severity scores, despite therapy, for whom treatment is no longer effective. These patients may not have been identified as being ready to be referred back to primary care with conventionally used decision supports, but may be considered ready by our decision supports. On the other hand, therapists do not want to increase the primary care patient population with redundant symptoms, leading to unnecessary risks of recurrence. The P_{95} ROM reference group cut-off values may indicate and aid a proper decision.

Decisions concerning possible referral to specialized mental health (secondary) care

This paragraph is aimed at general practitioners (GPs; primary care) in order to support

decisions about referral of patients to specialized mental health (secondary) care.

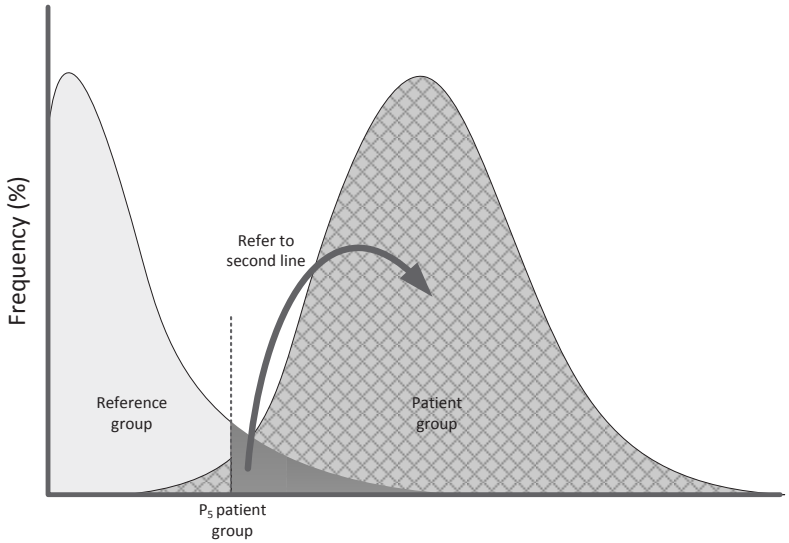


Figure 8.4: Cut-off value relevant for referral to specialized mental health care. Patients enter secondary treatment when they are no longer part of the reference population, but belong to the patient population instead, above the cut-off value P_5 ROM patient group.

Referral to specialized mental health (secondary) care may be indicated when the patient is more similar to the patient population than to the reference population. In this case the clinically relevant cut-off point is the point that the patient has to cross at the time of the assessment in order to be classified as similar to a clinically significant degree of psychiatric illness. As can be seen in Figure 8.4, the cut-off value, marking the bottom 5% of the ROM patient group, is equivalent to the 5th percentile score: the P_5 ROM patient. This cut-off value represents high sensitivity for psychopathology.

In Table 8.2 of the Appendix the cut-off values, the P_5 ROM patient group values, are summarized for the 19 ROM questionnaires. Four sets of questionnaires are available: 1) generic questionnaires; 2) questionnaires measuring mood disorders; 3) questionnaires measuring anxiety disorders; and 4) questionnaires measuring somatoform disorders. The use of reference values is feasible when ROM is available to GPs. Currently, ROM is used by some primary care psychologists [36] but not yet on a large scale by GPs [37].

When using the P_5 ROM patient group for referral to specialized mental health care, some issues have to be considered. The P_5 cut-off value is highly sensitive to disease. It can be considered as a reliable decision support when its result is negative, since true positives (psychopathology) are rarely missed among those who are actually positive – i.e., most sick people are recognized as being ill. However, high sensitivity is related to low specificity in a trade off. Low specificity (i.e., many false positive results) is associated with the burdening

of subjects with the mistaken prospect of facing a disease that they do not have. Furthermore, it may lead to additional tests and possibly to treatments that are not necessary or even detrimental [30,38].

Referral to specialized mental health care can be difficult because of the vague nature of complaints [39]. E.g., persons who are depressed may visit a GP where their disorder remains undetected and untreated [40]. The cause could be that GPs tend to be more responsive to the overall level of distress than to whether patients meet formal criteria for depression [41]. Another obstacle to referral to specialized mental health care could be patient attitudinal barriers to the expected extended treatment [42]. The questionnaires described in this study plus the provided reference values are tools to support clinical decisions about referral to specialized mental health care or counseling in primary care.

Reflection and recommendations on the use of ROM reference values

Reference values have to be used with care. Although it was not a topic of this thesis, the course Reference values have to be used with care. Although it was not a topic of this thesis, the course of questionnaire scores may be a more sensitive indication of the level of responsiveness. Comparison with percentile scores can assist the interpretation of these scores. Cut-off values can be used to support clinical decisions about referral and, at intake, decisions about diagnosis and treatment. Reference values in the present thesis were based on one-time cross-sectional data, whereas the clinician bases his decisions on repeated ROM sessions with his patient and on observed clinical changes. These time-series ROM data likely provide a wealth of information that can assist in better clinical decision making.

The reference values were based on cross-sectional data of subjects without any or with normal (non-treated) symptoms. The limit of 'normality' was determined according to the statistical distribution of the 95th percentile. However, this is an arbitrary assumption and there is no hard evidence that these recommended and statistically derived reference values predict morbidity, relapse, or recurrence [43,44]. Furthermore, cross-sectional data do not provide information about the duration of any of the symptoms [45]. Reference values are relevant factors in decisions about diagnosis and treatment and should therefore be related to prognosis [44]. However, in this study we have not evaluated the prognostic value of our proposed reference values. Mental health studies and physical health studies are not on a par, yet. In somatic medicine it is common practice to study effectiveness and efficacy of reference values [46,47].

2.5 Gender - and age effects

In our analyses, as described in chapters 3 to 7, we observed gender differences in reference values in the ROM reference group with women reporting more severe symptoms on observational and self-report scales for general psychopathology (i.e., BSI, MASQ-D30, and SF-36), depression (i.e., BDI-II, IDS-SR, and MADRS), anxiety (i.e., BSA, PI-R, PAI,

PSWQ, WDQ, SIAS, SPS, and IES-R), and body dysmorphic disorder (i.e., BICI) than men. No gender differences were found for the personality questionnaire DAPP-SF, the hypochondriasis questionnaire WI, and the chronic fatigue questionnaire (CIS-20R). These findings were not unexpected, since gender differences are commonly described in literature for well-defined patient groups [48-52] and for subjects from the general population [21,53]. Women are twice as likely to report depression or anxiety as men [21,53,54]. Gender may be related to a number of environmental causes and other aspects of psychopathology such as the stressors and exposures that influence the onset of disease, how symptoms are expressed [48,50,52,55-59], whether patients seek care [48,49,52,60,61], and how they are treated in the mental health care system [62].

The process of being mentally ill and subsequently seeking help has gender-specific aspects. The issue is what exactly is different between men and women. Are symptoms different or are their standards of acceptable psychological discomfort different? Is their sensitivity to different symptoms different or is their way to present symptoms different? Do women have (or take) more opportunity to report psychological symptoms to mental health care providers? Or are the differences caused by the questionnaires and criteria used in mental health care? [63].

The ROM reference group, a population based, non-treatment-seeking sample, may not completely reflect treatment-seeking patient samples in most of the above mentioned gender studies. Yet, this ROM reference group showed a similar gender effect in the reference values for most generic questionnaires, and questionnaires measuring major depression, anxiety-, and somatoform disorders. We have previously described that participants of the ROM reference group were not necessarily free of psychopathology. Therefore, the gender difference in this group might have been influenced by a relatively larger number of female than male subjects with psychopathology. Indeed, the percentage of participants with psychopathology was higher in women than in men (11.1% versus 6.6%). However, excluding these participants still yielded comparable gender differences in the reference values. For most questionnaires measuring generic symptoms, major depression, and somatoform disorders the gender differences decreased slightly. For some questionnaires measuring anxiety disorders the gender differences were unaffected or increased slightly (data not shown).

The gender effect in the reference values for most questionnaires measuring major depression, anxiety-, and somatoform disorders was similar in the ROM patient group: women reported slightly more symptoms than men. For generic questionnaires, no clear gender effect was found. Our data tentatively suggest that gender-specific reference values might increase precision in the assessment of the clinical state of psychiatric outpatients. However, the use of gender-specific reference values for questionnaires measuring generic symptoms, mood, anxiety, and somatoform disorders is open to debate. The consequence of using gender-specific reference values is illustrated in Figure 8.5. If the cut-off value P_{95}

ROM reference group is assumed to be lower for men, it would imply that women, treated in specialized mental health care, might be referred back to primary care with more residual symptoms compared to men.

Also, the effect of age on the reference values was studied, as described in chapters 3 and 4. For the generic questionnaires BSI, MASQ-D30, and DAPP-SF we showed that advancing age was not clearly associated with more symptoms of psychopathology. Only the results of the SF-36 showed a small negative correlation between age and health. This could be expected on the basis of declining physical health in the elderly. For the self-report BDI-II and IDS-SR, higher age was associated with a higher severity of MDD symptoms in women and men from the ROM reference group, which was not the case for the observer-rated MADRS. Since a clear general age effect was lacking, we decided not to pursue the analyses of age effects in detail.

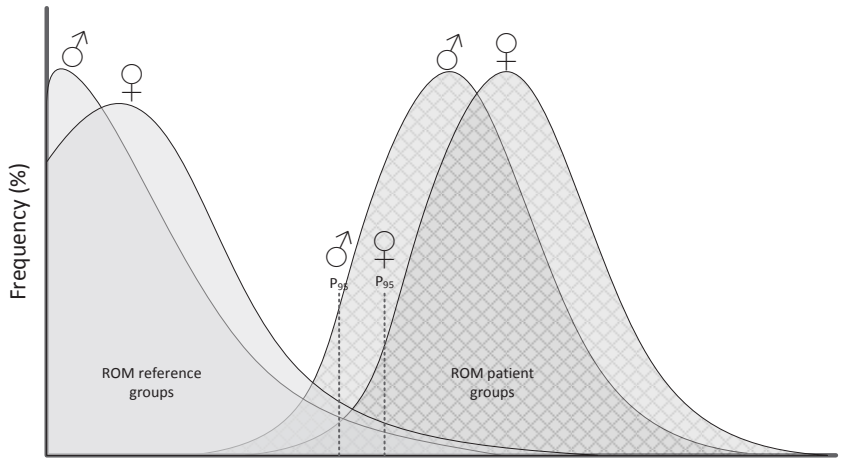


Figure 8.5. Hypothetical distribution of the scores of a questionnaire measuring psychopathology within the ROM reference group and within the ROM patient group. Scores are gender-specific: for women and men separately.

2.6 Discriminatory power of the questionnaires

The ROM questionnaires that are used to assess the level of (dys-) functionality in the ROM reference group and the ROM patient group are primarily designed for assessment of severity of MAS disorders. An additional aim of the NormQuest study was to test if these questionnaires can support the diagnostic process. By means of Receiver Operating Characteristics (ROC) and subsequent Area Under the Curve (AUC) analyses we investigated the discriminative power, which is indicative for the diagnostic capability of the 19 ROM questionnaires. For the generic questionnaires BSI, SF-36, and MASQ-D30, which assess general Axis-I psychopathology or distress, the discriminative power was good. This was

very satisfactory, given the fact that they are applicable for patients with more than one condition and irrespective of specific disorders. For all 14 disorder-specific (i.e., 3 major depression related, 8 anxiety related, and 3 somatoform disorder related) questionnaires the discriminative power was excellent. Only for the DAPP-SF subscales the discriminative power was poor. The DAPP-SF measures Axis-II personality traits that are thought to be stable and less affected by current psychopathology and treatment. So, although the questionnaires in this study were not designed for diagnostics but for severity assessment, the good discriminatory performance of the scales suggests that these questionnaires (except the DAPP-SF) can aid the diagnostics process.

Although the discriminatory power of the disorder-specific questionnaires are very good, these questionnaires cannot replace the MINI-Plus used for diagnosis. Most are self-report questionnaires and focus on particular symptoms relevant to a single disorder and are more sensitive to changes in outcome due to treatment as they assess the intensity of the symptoms that the patient suffers from [64,65]. The MINI-Plus, however, focuses on general psychopathology, distress, or general functioning and is a structured diagnostic interview, incorporating clinical judgment. It allows statements about the therapy effect regardless of the diagnosis and it is applicable for patients with more than one condition [66].

2.7 The SQ-48

To allow broad implementation, ROM questionnaires should ideally be free of copyright. Regrettably, some publishers claim copyrights for some ROM questionnaires. Therefore, the need has arisen to develop and validate freely available alternatives. As a first initiative, we developed and validated a 48-item psychological distress questionnaire, the Symptom Questionnaire (SQ-48; [67]), including measures of vitality and work functioning. This self-report questionnaire is intended as a tool for screening in clinical settings (psychiatric and non-psychiatric), monitoring during treatment in the context of ROM, and benchmarking. Reference values were derived and psychometric characteristics (e.g., internal consistency, convergent validity, and divergent validity) were validated. For the Depression subscale the discriminative power was good; for the subscales Anxiety, Cognitive complaints, Social phobia, and Vitality/optimism the discriminative power was moderate, for the subscales Aggression, Agoraphobia, and Somatic complaints the discriminative power was not clinically useful. By developing and validating the SQ-48 we have paved the way for further research that is aimed at the sensitivity to change due to treatment.

2.8 Recommendations for future research

The NormQuest study presented in this thesis can be seen as the overture to the establishment of reference values for all ROM questionnaires, used for the assessment of MAS disorders. Several additions and adjustments may further improve the quality and implementation of these reference values.

- External validity of the reference values for certain subgroups can be improved. Replication of this study with children, the elderly, and ethnic minorities is needed. Furthermore, the presented reference values are not necessarily generalizable to other language versions of the questionnaires or to other countries and cultures [68-70]. So, international and cross-country studies are recommended to develop internationally valid outcome measures, including reference values.
- The definition of 95% of a population as being normal, and 5% as being abnormal, is common practice but an arbitrary choice. Future research has to evaluate how well this definition and subsequent cut-off is in sink with the objective to provide an adequate tool to support clinical decisions on referral back to primary care.
- The size of the ROM reference group and of the ROM patient group that completed the questionnaires measuring somatoform disorders was suboptimal. Replication of the study with larger samples would enhance the validity and precision of the reference values.
- It might be possible to improve the specificity of a questionnaire without compromising a high sensitivity by sequencing questionnaires. By requiring a sequence of positive test results before taking further diagnostic action or starting treatment, the specificity of the questionnaire might be improved [30]. This would apply to patients with mild to moderately severe symptoms. Furthermore, either the sensitivity or the specificity of a questionnaire might be improved by using it in combination with a second questionnaire. Requiring a positive result from two questionnaires increases the specificity but decreases the sensitivity. Conversely, if a positive result on either questionnaire is taken to indicate the presence of the disease the sensitivity will become higher but the specificity will become lower [30]. In this study we focused on individual questionnaires. The effect of specific combinations of questionnaires on sensitivity and specificity could be further studied.
- Reference values are widely applied and recognized in laboratory medicine [4], but not in mental health care yet. The clinical application of test scores would have to be further evaluated. Subsequently, following laboratory medicine routine, a comprehensive approach should ideally be developed to implement the reference values of this study nation-wide. This would include an information development plan, summaries of reference values and clinical guidelines (i.e., elaboration of the guidelines in section 2.6.), and national reporting. Stakeholders (e.g.,psychiatrists, GPs, mental health nurses, managers, and insurance companies) would have to be engaged and motivated. Because ROM is getting implemented in several organizations, this seems feasible. Studies on implementation and factors influencing implementation are needed but lacking, as far as we know.
- It is imperative to have an optimal (not maximal) set of questionnaires in ROM. The set of 19 questionnaires we provided reference values for may not constitute this optimum. Further research will have to decide whether questionnaires have to be added, removed or replaced. Newly added questionnaires will need rigorously assessed reference values similar to the ones we provided.
- Future research could evaluate whether the extension of ROM with extra

questionnaires regarding (additional) somatoform disorders and subsequent derivation of reference values would increase the utility of ROM.

- With the introduction of the DSM-V (APA, 2011), revisions for some diagnostic categories may warrant adaptations of some questionnaires. These adaptations and any newly developed questionnaires will require (new) reference values.

3. GENERAL CONCLUSIONS

We have gathered reference data in a larger group of population based controls and in a larger number of MAS outpatients than in any other Dutch or international study. Reference values, including cut-off scores, were calculated for 19 questionnaires.

When collecting reference data, it is important to match the ROM reference group to the ROM patient group in terms of gender- and age distribution, as well as level of urbanization. To minimalize selective sampling the response rate has to be optimalized (e.g., by offering the possibility of home-based completion of questionnaires, a larger monetary incentive, personalized invitational letters, stamped return envelopes, contacting participants before sending questionnaires, sending non-respondents a second invitational letter). Furthermore, clinical interpretations of symptoms and complaints have to be reliable. Therefore, and to minimalize inter-rater variability between interviewers, interviewers should be trained and supervised.

The provided reference values can be used to support decisions of referral to or from specialized mental health care. When a therapist considers treatment termination and subsequent referral back to primary care, the P_{95} ROM reference group can be used to support the decision. When a GP regards referral to specialized mental health care a feasible option, the P_5 ROM patient group can support his decision.

Reference data have to be used with care. Percentile scores are clear but strict; the practical use of these reference values should not be that strict. Purely statistical approaches are unsatisfactory. Additional information regarding comorbidity, personal functioning, and motivation for treatment is needed. A treatment strategy is most likely to succeed when it combines effective therapy and a strong therapeutic relationship, with ROM and its reference values.

Reference list

1. Poortvliet MC, Lamkadden M, Deville W. Niet op naam ingeschreven (NONI) bij de huisarts. Inventarisatie en gevolgen voor de ziekenfondsverzekerden. Utrecht: NIVEL; 2005.

2. Solberg HE. (1987) International Federation of Clinical Chemistry. Scientific committee, Clinical Section. Expert Panel on Theory of Reference Values and International Committee for Standardization in Haematology Standing Committee on Reference Values. Approved recommendation (1986) on the theory of reference values. Part 1. The concept of reference values. *Clin Chim Acta,* 165 (1), 111-118.

3. Geffre A, Friedrichs K, Harr K, Concordet D, Trumel C, Braun JP. (2009) Reference values: a review. *Vet Clin Pathol,* 38 (3), 288-298.

4. Henny J, Hyltoft Petersen P. (2004) Reference values: from philosophy to a tool for laboratory medicine. *Clin Chem Lab Med*, 42 (7), 686-691.

5. Park HN, Sinn DH, Gwak GY, et al. (2012) Upper normal threshold of serum alanine aminotransferase in identifying individuals at risk for chronic liver disease. *Liver Int*, 32 (6), 937-944.

6. Kallner A, Gustavsson E, Hendig E. (2000) Can age and sex related reference intervals be derived for non-healthy and non-diseased individuals from results of measurements in primary health care? *Clin Chem Lab Med,* 38 (7), 633-654.

7. Evers A, Van Vliet-Mulder JC, Groot CJ. (2000) Documentatie van tests en testresearch in Nederland. Deel 1 testbeschrijvingen. Van Gorcum, Assen.

8. Solberg HE. (2008) Establishment and use of reference values. Burtis CA, Ashwood ER, Bruns DE, editors. Fundamentals of clinical chemistry. 6[14], 229-238. St. Louis, Missouri, Saunders Elsevier.

9. Reed AH, Henry RJ, Mason WB. (1971) Influence of statistical method used on the resulting estimate of normal range. *Clin Chem*, 17 (4), 275-284.

10. Gräsbeck R, Saris N-E. (1969) Establishment and use of normal values. *Scan J Clin Lab Invest*, 26 (110), 62-63.

11. Kendall PC, Marrs-Garcia A, Nath SR, Sheldrick RC. (1999) Normative comparisons for the evaluation of clinical significance. *J Consult Clin Psychol*, 67 (3), 285-299.

12. Horn PS, Feng L, Li Y, Pesce AJ. (2001) Effect of outliers and nonhealthy individuals on reference interval estimation. *Clin Chem*, 47 (12), 2137-2145.

13. Zimmerman M, Chelminski I, Posternak M. (2004) A review of studies of the Montgomery-Asberg Depression Rating Scale in controls: implications for the definition of remission in treatment studies of depression. *Int Clin Psychopharmacol,* 19 (1), 1-7.

14. Cohen H. (1981) The evolution of the concept of disease. Caplan AL, Engelhardt HT, McCartney JJ, editors. Concepts of health and disease: Interdisciplinary perspectives. 209-220. Reading, MA, Addison-Wesley.

15. Wakefield JC. (1992) The concept of mental disorder. On the boundary between biological facts and social values. *Am Psychol*, 47 (3), 373-388.

16. Boyd JC. (2010) Defining laboratory reference values and decision limits: populations, intervals, and interpretations. *Asian J Androl,* 12 (1), 83-90.

17. Schulte-van Maaren YWM, Carlier IVE, Zitman FG, Hemert AM, De Waal MW, Van Noorden MS, Giltay EJ. (2012) Reference values for generic instruments used in Routine Outcome Monitoring: the Leiden Routine Outcome Monitoring Study (in press). *BMC Psychiatry.*

18. Schulte-van Maaren YWM, Carlier IVE, Zitman FG, Van Hemert AM, De Waal MW, Van der Does AJW, Van Noorden MS, Giltay EJ. (2012) Reference values for major depression questionnaires: the Leiden Routine Outcome Monitoring Study. *Journal of Affective Disorders.*

19. Schulte-van Maaren YWM, Giltay EJ, Van Hemert AM, Zitman FG, De Waal MW, Carlier IVE. (2013) Reference values for anxiety questionnaires: the Leiden Routine Outcome Monitoring Study (in press). *Journal of Affective Disorders.*

20. Schulte-van Maaren YWM, Giltay EJ, Van Hemert AM, Zitman FG, De Waal MW, Van Rood YR, Carlier IVE. (2013) Reference values for the Body Image Concern Inventory (BICI), the Whitely Index (WI), and the Checklist Individual Strength (CIS-20R): the Leiden Routine Outcome Monitoring Study (submitted). *Journal of Psychosomatic Research.*

21. De Graaf R, Ten Have M, Van Gool C, Van Dorsselaer S. (2012) Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. *Soc Psychiatry Psychiatr Epidemio*l, 47 (2), 203-213.

22. Rothman KJ. (2012) Dealing with bias. Rothman KJ, editor. Epidemiology: An Introduction. 2nd[7], 124-147. Oxford, UK, Oxford University Press.

23. Hunt M, Auriemma J, Cashaw AC. (2003) Self-report bias and underreporting of depression on the BDI-II. *J Pers Assess*, 80 (1), 26-30.

24. Kleinman A, Eisenberg L, Good B. (1978) Culture, illness, and care: clinical lessons from anthropologic and cross-cultural research. *Ann Intern Med*, 88 (2), 251-258.

25. Bullinger M, Anderson R, Cella D, Aaronson N. (1993) Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Qual Life Res*, 2 (6), 451-459.

26. Herdman M, Fox-Rushby J, Badia X. (1998) A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Qual Life Res,* 7 (4), 323-335.

27. Anderson RT, Aaronson NK, Wilkin D. (1993) Critical review of the international assessments of health-related quality of life. Qual Life Res, 2 (6), 369-395.

28. Bowling A, Bond M, Jenkinson C, Lamping DL. (1999) Short Form 36 (SF-36) Health Survey questionnaire: which normative data should be used? Comparisons between the norms provided by the Omnibus Survey in Britain, the Health Survey for England and the Oxford Healthy Life Survey. *J Public Health Med*, 21 (3), 255-270.

29. Alonso J, Ferrer M, Gandek B, et al. (2004) Health-related quality of life associated with chronic conditions in eight countries: results

from the International Quality of Life Assessment (IQOLA) Project. *Qual Life Res*, 13 (2), 283-298.

30. Rothman KJ. (2012) Epidemiology in Clinical Settings. Rothman KJ, editor. Epidemiology: an introduction. 2nd[13], 235-253. Oxford, UK, Oxford University Press.

31. Hawley CJ, Gale TM, Sivakumaran T. (2002) Defining remission by cut off score on the MADRS: selecting the optimal value. *Journal of Affective Disorders*, 72 (2), 177-184.

32. Kumar G, Steer RA, Teitelman KB, Villacis L. (2002) Effectiveness of Beck Depression Inventory-II subscales in screening for major depressive disorders in adolescent psychiatric inpatients. *Assessment*, 9 (2), 164-170.

33. Trivedi MH, Rush AJ, Ibrahim HM, et al. (2004) The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation. *Psychol Med,* 34 (1), 73-82.

34. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. (1996) The Inventory of Depressive Symptomatology (IDS): psychometric properties. *Psychol Med*, 26 (3), 477-486.

35. Creamer M, Bell R, Failla S. (2003) Psychometric properties of the Impact of Event Scale - Revised. *Behav Res Ther,* 41 (12), 1489-1496.

36. Ashworth M, Robinson SI, Godfrey E, Shepherd M, Evans C, Seed P, Parmentier H, Tylee A. (2005) Measuring mental health outcomes in primary care: the psychometric properties of a new patient-generated outcome measure, 'PSYCHLOPS' ('psychological outcome profiles'). *Primary Care Mental Health*, 3, 261-270.

37. Addington D, Kyle T, Desai S, Wang J. (2010) Facilitators and barriers to implementing quality measurement in primary mental health care: Systematic review. *Can Fam Physician,* 56 (12), 1322-1331.

38. Bewick V, Cheek L, Ball J. (2004) Statistics review 13: receiver operating characteristic curves. *Crit Care*, 8 (6), 508-512.

39. Huibers MJ, Beurskens AJ, Bleijenberg G, van Schayck CP. (2007) Psychosocial interventions by general practitioners. *Cochrane Database Syst Rev* (3), CD003494.

40. Coyne JC, Thompson R, Klinkman MS, Nease DE, Jr. (2002) Emotional disorders in primary care. *J Consult Clin Psychol*, 70 (3), 798-809.

41. Klinkman MS, Coyne JC, Gallo S, Schwenk TL. (1998) False positives, false negatives, and the validity of the diagnosis of major depression in primary care. *Arch Fam Med,* 7 (5), 451-461.

42. Coyne JC, Palmer SC, Sullivan PA. (2002) Is case-finding an inefficient way of addressing depression as a public health problem? *Pharmacoepidemiol Drug Saf,* 11 (7), 545-547.

43. Flegal KM, Ogden CL, Wei R, Kuczmarski RL, Johnson CL. (2001) Prevalence of overweight in US children: comparison of US growth charts from the Centers for Disease Control and Prevention with other reference values for body mass index. *Am J Clin Nutr,* 73 (6), 1086-1093.

44. Ohkubo T, Imai Y, Tsuji I, Nagai K, Ito S, Satoh H, Hisamichi S. (1998) Reference values for 24-hour ambulatory blood pressure

monitoring based on a prognostic criterion: the Ohasama Study. *Hypertension*, 32 (2), 255-259.

45. Cook S, Weitzman M, Auinger P, Nguyen M, Dietz WH. (2003) Prevalence of a metabolic syndrome phenotype in adolescents: findings from the third National Health and Nutrition Examination Survey, 1988-1994. *Arch Pediatr Adolesc Med*, 157 (8), 821-827.

46. Alberti KG, Zimmet PZ. (1998) Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med*, 15 (7), 539-553.

47. Cundy T. (2012) Proposed new diagnostic criteria for gestational diabetes--a pause for thought? *Diabet Med*, 29 (2), 176-180.

48. Angst J, Gamma A, Gastpar M, Lepine JP, Mendlewicz J, Tylee A. (2002) Gender differences in depression. Epidemiological findings from the European DEPRES I and II studies. *Eur Arch Psychiatry Clin Neurosci*, 252 (5), 201-209.

49. Kinrys G, Wygant LE. (2005) Anxiety disorders in women: does gender matter to treatment? *Rev Bras Psiquiatr*, 27 Suppl 2, S43-S50.

50. Nolen-Hoeksema S. (2012) Emotion regulation and psychopathology: the role of gender. *Annu Rev Clin Psychol*, 8, 161-187.

51. Pigott TA. (1999) Gender differences in the epidemiology and treatment of anxiety disorders. *J Clin Psychiatry*, 60 Suppl 18, 4-15.

52. Seeman MV. (1997) Psychopathology in women and men: focus on female hormones. *Am J Psychiatry,* 154 (12), 1641-1647.

53. Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler KS. (1994) Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Arch Gen Psychiatry*, 51 (1), 8-19.

54. Bijl RV, Ravelli A, Van Zessen G. (1998) Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc Psychiatry Psychiatr Epidemiol*, 33 (12), 587-595.

55. Bekker MH. (1996) Agoraphobia and gender: a review. *Clinical Psychology Review*, 16, 129-146.

56. Bekker MH, Vingerhoets AJ. (2001) Male and female tears: Swallowing versus shedding? The relationship between crying, biological sex and gender. Vingerhoets AJJM, Cornelius RR, editors. Adult Crying: A Biopsychosocial Approach. 91-114. Hove, United Kingdom, Brunner-Routledge.

57. Bekker MH, Van Mens-Verhulst J. (2007) Anxiety disorders: sex differences in prevalence, degree, and background, but gender-neutral treatment. Gend Med, 4 Suppl B, S178-S193.

58. Chambless DL. (1988) Sex and phobia [in Dutch]. *Gedragstherapie*, 21, 283-293.

59. MacKinaw-Koons B, Vasey MW. (2000) Considering sex differences in anxiety and its disorders across the life span: A construct-validation approach. *Applied and Preventive Psychology*, 9, 191-209.

60. Almqvist F. (1986) Sex differences in adolescent psychopathology. *Acta Psychiatr Scand,* 73 (3), 295-306.

61. Scheibe S, Preuschhof C, Cristi C, Bagby RM. (2003) Are there gender differences in major depression and its response to antidepressants? *J Affect Disord,* 75 (3), 223-235.

62. Johnson J, Stewart DE. (2010) DSM-V: toward a gender sensitive approach to psychiatric diagnosis. *Arch Womens Ment Health*, 13 (1), 17-19.

63. Van Mens-Verhulst J. (1996) Gender-specifieke ontwikkelingen in de klinische psychologie. Top T, Heesink J, editors. Psychologie en sekse. [12], 191-208. Houten, NL, Bohn Stafleu Van Loghum.

64. De Beurs E, Den Hollander-Gijsman ME, Van Rood YR, Van der Wee NJ, Giltay EJ, Van Noorden MS, Van der Lem R, Van Fenema EM, Zitman FG. (2011) Routine outcome monitoring in the Netherlands: practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clin Psychol Psychother,* 18, 1-12.

65. McKay R, Coombs T, Pirkis J. (2012) A framework for exploring the potential of routine outcome measurement to improve mental health care. *Australas Psychiatry*, 20 (2), 127-133.

66. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC. (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry,* 59 Suppl 20, 22-33.

67. Carlier I, Giltay E, Vergeer P. (2012) Development and validation of the 48-item Symptom Questionnaire (SQ-48) in patients with depressive, anxiety and somatoform disorders. *Psychiatry Res*, 200 (2-3), 904-910.

68. Fisher CE, Spaeth-Rublee B, Alan PH. (2013) Developing mental health-care quality indicators: toward a common framework. *Int J Qual Health Care*, 25 (1), 75-80.

69. Hermann RC, Palmer RH. (2002) Common ground: a framework for selecting core quality measures for mental health and substance abuse care. *Psychiatr Serv*, 53 (3), 281-287.

70. O'Brien AP, Boddy JM, Hardy DJ. (2007) Culturally specific process measures to improve mental health clinical practice: indigenous focus. *Aust N Z J Psychiatry,* 41 (8), 667-674.

APPENDIX

Table 8.1: Cut-off values for specialized mental health (secondary) care supporting decisions concerning referral back to primary care - P_{95} ROM reference group

Table 8.2: Cut-off values for primary care supporting decisions concerning referral to specialized mental health (secondary) care - P_5 ROM patient group

Table 8.1: Cut-off values for specialized mental health (secondary) care supporting decisions concerning referral back to primary care - P_{95} ROM reference group

Questionnaire	Domain	Cut-off
Symptom Questionnaire 48 items (SQ-48)	Generic	
Aggression (AGGR)		5.0
Agoraphobia (AGOR)		2.0
Anxiety (ANXI)		11.2
Cognitive complaints (COGN)		11.0
Depression (MOOD)		8.0
Somatic complaints (SOMA)		8.0
Social phobia (SOPH)		9.0
Vitality/optimism (VITA)		15.0
Brief Symptom Inventory (BSI)	Generic	0.68
Mood & Anxiety Symptom Questionnaire 30-item (MASQ-D30)	Generic	
General distress (GD)	depression	23
Anhedonic depression (AD)	anxiety	29
Anxious arousal (AA)		17
Short Form 36 (SF36)*	Generic	
Physical Functioning	health status	65
Role-Physical	well-being	5
Bodily Pain		54
Social Functioning		63
Mental Health		56
Role-Emotional		33
Vitality		40
General Health		45
Dimensional Assessment of Personality Pathology–short form (DAPP-SF)	Generic	
Submissiveness	personality	3.50
Cognitive Distortion		2.33
Identity Problems		2.70
Affective Lability		3.50
Stimulus Seeking		3.38
Compulsivity		4.00
Restricted Expression		3.63
Callousness		2.60
Oppositionality		3.20
Intimacy Problems		3.38
Rejection		3.75
Anxiousness		3.50
Conduct Problems		2.13
Suspiciousness		2.15
Social Avoidance		3.33
Narcissism		3.50
Insecure Attachment		3.33
Self-Harm		1.50
Beck Depression Inventory-II (BDI-II)	MDD	13
Inventory of Depressive Symptomatology - Self-Report (IDS-SR)	MDD	20
Montgomery Åsberg Depression Rating Scale (MADRS)	MDD	11
Brief Scale for Anxiety (BSA)	Anxiety Disorder	11
PADUA Inventory Revised (PI-R)	OCD	43
Panic Appraisal Inventory (PAI)	Panic Disorder	37
Anticipated panic		47
Perceived consequences of Panic (Total):		
Perceived self-efficacy in coping with panic		65
Penn State Worry Questionnaire (PSWQ)	Worry (pathological)	66
Worry Domains Questionnaire (WDQ)	Worry	74
Social Interaction and Anxiety Scale (SIAS)	Social Anxiety	32
Social Phobia Scale (SPS)	Social Anxiety	19
Impact of Event Scale – Revised (IES-R)[1] Total	PTSD	36
Body Image Concern Inventory (BICI)	BDD	55
Whitely Index (WI)	Hypochondriasis	6
CIS20R	Chronic Fatigue	92

ROM, routine outcome monitoring; MDD denotes major depressive disorder; OCD denotes obsessive compulsive disorder; PTSD denotes posttraumatic stress disorder; BDD denotes body dysmorphic disorder.
*: P_5 ROM reference group and P_{95} ROM patient group, as high scores indicate better functioning.

Table 8.2: Cut-off values for primary care supporting decisions concerning referral to specialized mental health (secondary) care - P_5 ROM patient group

Questionnaire	Domain	Cut-off
Symptom Questionnaire 48 items (SQ-48)	Generic	
Aggression (AGGR)		0.0
Agoraphobia (AGOR)		0.0
Anxiety (ANXI)		2.0
Cognitive complaints (COGN)		3.0
Depression (MOOD)		1.0
Somatic complaints (SOMA)		0.0
Social phobia (SOPH)		0.0
Vitality/optimism (VITA)		6.0
Brief Symptom Inventory (BSI)	Generic	0.34
Mood & Anxiety Symptom Questionnaire 30-item (MASQ-D30)	Generic	
General distress (GD)	depression	17
Anhedonic depression (AD)	anxiety	17
Anxious arousal (AA)		18
Short Form 36 (SF36)*	Generic	
Physical Functioning	health status	100
Role-Physical	well-being	100
Bodily Pain		100
Social Functioning		88
Mental Health		76
Role-Emotional		100
Vitality		65
General Health		90
Dimensional Assessment of Personality Pathology–short form (DAPP-SF)	Generic	
Submissiveness	personality	1.25
Cognitive Distortion		1.00
Identity Problems		1.33
Affective Lability		1.63
Stimulus Seeking		1.00
Compulsivity		1.38
Restricted Expression		1.75
Callousness		1.00
Oppositionality		1.40
Intimacy Problems		1.13
Rejection		1.13
Anxiousness		1.67
Conduct Problems		1.00
Suspiciousness		1.00
Social Avoidance		1.17
Narcissism		1.10
Insecure Attachment		1.00
Self-Harm		1.00
Beck Depression Inventory-II (BDI-II)	MDD	14
Inventory of Depressive Symptomatology - Self-Report (IDS-SR)	MDD	18
Montgomery Åsberg Depression Rating Scale (MADRS)	MDD	11
Brief Scale for Anxiety (BSA)	Anxiety Disorder	6
PADUA Inventory Revised (PI-R)	OCD	20
Panic Appraisal Inventory (PAI)	Panic Disorder	
Anticipated panic		14
Perceived consequences of Panic (Total):		10
Perceived self-efficacy in coping with panic		29
Penn State Worry Questionnaire (PSWQ)	Worry (pathological)	48
Worry Domains Questionnaire (WDQ)	Worry	44
Social Interaction and Anxiety Scale (SIAS)	Social Anxiety	18
Social Phobia Scale (SPS)	Social Anxiety	11
Impact of Event Scale – Revised (IES-R)[1] Total	PTSD	19
Body Image Concern Inventory (BICI)	BDD	39
Whitely Index (WI)	Hypochondriasis	5
CIS20R	Chronic Fatigue	74

ROM, routine outcome monitoring; MDD denotes major depressive disorder; OCD denotes obsessive compulsive disorder; PTSD denotes posttraumatic stress disorder; BDD denotes body dysmorphic disorder.
*: P_5 ROM reference group and P_{95} ROM patient group, as high scores indicate better functioning

Samenvatting en algemene discussie

In het kort

NormQuest is een studie die ten doel had referentiewaarden (normscores) voor de in Routine Outcome Monitoring (ROM) gebruikte vragenlijsten in een huisartspopulatie vast te stellen. Die referentiewaarden waren voor de meeste ROM vragenlijsten niet bekend, maar ze zijn wel van belang voor een juiste interpretatie van de scores bij patiënten met stemmings-, angst- en somatoforme (SAS-) stoornissen in de tweede en de derde lijn. In NormQuest stelden we deze referentiewaarden vast bij een grote steekproef van ca. 1300 patiënten uit huisartspraktijken voor negentien vragenlijsten die in ROM-SAS gebruikt worden.

Inleiding

In de geestelijke gezondheidszorg is het gebruikelijk dat de resultaten van een behandeling geëvalueerd worden door de behandelaar en de patiënt. ROM is een methode om die resultaten meer objectief vast te stellen door regelmatig de aard, de ernst en het beloop van klachten van patiënten te meten. Feedback over de behandelresultaten is gunstig gebleken voor de behandeling: deze werkt naast informerend vooral ook beïnvloedend en motiverend voor de patiënt. Een behandelaar kan met deze feedback nagaan of de behandeling aanslaat. ROM blijkt een gunstig effect te hebben op de communicatie tussen patiënt en behandelaar. Deze feedback wordt mogelijk gemaakt door ROM-scores te vergelijken met referentiewaarden. Behalve voor het meten van de werkzaamheid van de behandeling kunnen referentiewaarden gebruikt worden ter ondersteuning van besluiten over voortzetting, wijziging of beëindiging van de behandeling.

Een 64-jarige vrouw werd opgenomen met ernstige depressieve en angstklachten. Haar problemen waren 7 jaar geleden nogal abrupt begonnen na echtelijke problemen die resulteerden in een echtscheiding. In haar medische verleden was er sprake van agorafobie en orthostatische hypotensie (plotselinge bloeddrukdalingen met duizeligheid of flauwvallen als gevolg). Meerdere malen werd ze behandeld voor angst en depressie met psychotherapie en verschillende antidepressiva, soms klinisch, soms poliklinisch. Vanwege ernstige depressie met psychotische kenmerken en resistentie tegen behandeling met antidepressiva werd ze opgenomen in de psychiatrische universiteitskliniek (PUK) van het Leids Universitair Medisch Centrum (LUMC). Ze werd gedurende enige tijd behandeld met elektroconvulsietherapie (ECT) en haar depressie verminderde. Tijdens de behandeling werd de ernst van de depressie wekelijks beoordeeld door haar psychiater. Ook werden er wekelijks ROM-scores bepaald door een onafhankelijk onderzoeksverpleegkundige.

De ROM-scores in het bovenstaande geval behoeven interpretatie. Is de patiënt verbeterd of hersteld (in remissie), verslechterd, of niet veranderd? Een belangrijke vraag voor de therapeut is: wanneer is de patiënt voldoende hersteld om de volgende stap in de behandeling te maken? Een aanpak die een dergelijke beslissing kan ondersteunen is het vergelijken van de ROM-scores met die van een normale of referentiepopulatie. Wanneer ROM-scores onder een bepaalde cut-off-waarde komen, is de patiënt qua ernst van de depressieve symptomen niet meer te onderscheiden van de normale referentiepopulatie. De behandeling zou dan kunnen verschuiven naar het voorkomen van een terugval en de patiënt zou terugverwezen kunnen worden naar de eigen huisarts. Evidence based cut-off-waarden voor ROM-vragenlijsten kunnen klinische beslissingen ondersteunen. Deze cut-off-waarden kunnen worden afgeleid uit de verdeling van scores van de gezonde bevolking en van patiëntenpopulaties. Cut-off-waarden en aanvullende scoreverdelingen worden hier aangeduid als referentiewaarden.

Doel van deze NormQuest-studie

Voor diverse ROM-vragenlijsten zijn er nog geen referentiewaarden beschikbaar. Om deze referentiewaarden te ontwikkelen, werd in 2008 de NormQuest-studie geïnitieerd door de afdeling Psychiatrie van het LUMC en de Zuid-Hollandse instelling voor geestelijke gezondheidszorg Rivierduinen. Dit proefschrift presenteert referentiewaarden voor 19 ROM-vragenlijsten (18 bestaande en 1 nieuw ontwikkelde) die kunnen worden gebruikt om klinische beslissingen te ondersteunen bij de behandeling en verwijzing van patiënten met SAS-stoornissen. Referentiewaarden bestaan uit afkap (cut-off-) waarden die het verschil markeren tussen de patiëntenpopulatie (‘psychisch ziek’) en de referentiepopulatie (‘gezond’).

Verder wordt het onderscheidend vermogen van de vragenlijsten besproken aan de hand van Receiver Operating Characteristics (ROC-) analyses. Ook wordt het effect van geslacht en leeftijd op referentiewaarden gepresenteerd. Het secundaire doel van de NormQuest-studie betreft de noodzaak van de ontwikkeling van publiek domein vragenlijsten. De generieke Symptom Questionnaire-48 (SQ-48) is ontwikkeld en referentiewaarden worden gepresenteerd.

ROM-vragenlijsten

Voor het meten van de klachten binnen ROM is een reeks van objectieve, standaard vragenlijsten (zowel voor zelfrapportage als observaties) een essentieel onderdeel. ROM-vragenlijsten moeten klinisch relevant, gevoelig voor verandering en minimaal belastend voor patiënten, personeel en organisatie zijn. Derhalve dient de keuze van vragenlijsten gebaseerd te zijn op validiteit, betrouwbaarheid en op beschikbaarheid in het publieke domein. Vragenlijsten voor ROM zijn generiek of specifiek. Generieke vragenlijsten worden gebruikt voor het meten van algemene psychopathologie, angst of algemeen functioneren. Omdat ze in principe geschikt zijn voor alle patiënten met psychische problemen, maken ze uitspraken mogelijk over het behandeleffect, onafhankelijk van de diagnose. Verder zijn

ze geschikt voor patiënten met meer dan een aandoening. Bovendien vergemakkelijken ze vergelijkingen tussen verschillende patiëntengroepen. Ziekte-specifieke vragenlijsten zijn gericht op specifieke symptomen van een enkele psychiatrische aandoening of groep stoornissen en worden alleen aangeboden aan patiënten die voldoen aan de criteria voor die bepaalde stoornis. Zij zijn gevoeliger voor veranderingen door de behandeling, aangezien ze de intensiteit van de symptomen waarvoor behandeld wordt meten. Voor bijna alle ROM-vragenlijsten geldt dat een hogere score op een vragenlijst méér psychopathologie betekent. Omdat ROM-vragenlijsten bij veel patiënten afgenomen worden, vaak ook meerdere keren, is het belangrijk dat er geen of minimale kosten verbonden zijn aan het gebruik van de vragenlijsten. Een belangrijk criterium voor selectie van onze vragenlijsten is daarom geweest dat deze beschikbaar waren in het publieke domein.

Stemmings-, Angst- en Somatoforme (SAS-) stoornissen

De meerderheid van de patiënten van het LUMC en een aanzienlijk aantal patiënten van Rivierduinen worden behandeld voor SAS-stoornissen. Schattingen van de life-time prevalentie in Nederland bedragen ruim 19% voor de groep stemmings- en angststoornissen. Minder duidelijk is de prevalentie van somatoforme stoornissen, maar schattingen van de prevalentie zijn ook relatief hoog (± 16%). SAS-aandoeningen komen vaak voor als comorbide stoornissen: depressie gaat vaak gepaard met angst en patiënten met een angststoornis zijn ook vaak somber gestemd. Ook kunnen bijvoorbeeld persoonlijkheidsstoornissen en middelenmisbruik naast SAS-stoornissen voorkomen.

SAS-aandoeningen zijn de meest frequent waargenomen psychische stoornissen in de eerste lijn (bij de huisarts). De ziektelast is zeer groot, met de depressieve stoornis als de meest belangrijke bijdrage aan de wereldwijde ziektelast.

Referentiewaarden

Referentiewaarden worden gebruikt voor variabelen die kwantitatief worden beoordeeld, zoals lichaamstemperatuur of ernst van de depressie. Referentiewaarden worden bepaald in een referentiepopulatie. De selectie van de referentiepopulatie en de definitie van de referentiewaarden zijn van de referentiepopulatie en de definitie van de referentiewaarden zijn belangrijk. De referentiepopulatie moet bestaan uit individuen met een goed gedefinieerde gezondheidstoestand. Gezondheid kan worden gedefinieerd op verschillende manieren: medisch en statistisch. De medische benadering beschouwt gezondheid als afwezigheid van pathologie, in absolute termen, of op zijn minst van een bepaald type van pathologie. Zo worden mensen met die aandoening uitgesloten van de referentiepopulatie. Wanneer volgens deze medische benadering referentiewaarden voor bijvoorbeeld depressie bepaald worden, zullen depressieve patiënten uitgesloten worden van de referentiepopulatie. De statistische benadering is gebaseerd op de verdeling van de scores van een variabele in een totale referentiepopulatie, inclusief personen die toevallig hoog scoren op die variabele. In

de statistische benadering worden scores in de centrale 95% meestal beschouwd als gezond; extreem hoge of lage scores worden als afwijkend gezien. Voor veel ROM-variabelen, zoals de ernst van een depressie, is alleen een extreem hoge score afwijkend. (Een hogere score op een vragenlijst betekent immers méér psychopathologie. Heel lage scores hebben daarom geen speciale betekenis.) In dergelijke gevallen wordt 'afwijkend' beperkt tot de top 5% van de verdeling. Personen met verhoogde niveaus van psychopathologie worden niet uitgesloten van de referentiegroep, want anders zou een te gezonde ('supernormale') steekproef gecreëerd worden. De resulterende referentiewaarden zouden overdreven streng zijn. Van de psychiatrisch zieke populatie echter, kan de onderste 5% van de scores worden beschouwd als 'afwijkend': hun symptomen zijn subsyndromal geworden – er zijn nog wel klachten, maar men is niet meer ernstig ziek. In deze studie werd de statistische benadering gevolgd.

We berekenden percentiel scores als referentiewaarden. In een scheve verdeling, zoals we die (terecht, naar later bleek) verwachtten voor de ROM-referentiegroep, zijn deze meer zinvol dan de vaak gebruikte gemiddeldes en standaarddeviaties. Om vergelijking met de internationale literatuur mogelijk te maken, hebben we echter gemiddeldes en standaarddeviaties wel bepaald.

Samenvatting van de belangrijkste bevindingen

Dit is de eerste in Nederland uitgevoerde studie van deze omvang die als doel heeft referentiewaarden te genereren voor SAS-vragenlijsten. Hoofdstuk 2 beschrijft de doelstellingen, het design en de methodieken van deze studie. Twee groepen werden geïncludeerd. De eerste groep, de ROM-patiëntengroep, bestond uit 5.269 patiënten die voor een of meer SAS-stoornissen behandeld werden op de polikliniek van het LUMC of bij de Zuid-Hollandse instelling voor geestelijke gezondheidszorg Rivierduinen. De patiënten, in de leeftijd van 18-65 jaar, werden gescreend als onderdeel van de routine intakeprocedure. De tweede groep, de ROM-referentiegroep, bestond uit 1302 huisartspatiënten, ook tussen 18 en 65, geregistreerd bij een van de acht deelnemende huisartsen, maar niet per se onder behandeling. Deze personen kunnen worden beschouwd als een steekproef van de algemene bevolking, aangezien in Nederland 99,9% van de bevolking staat ingeschreven bij een huisarts. De ROM-referentiegroep en de ROM-patiëntengroep waren vergelijkbaar in termen van geslacht, leeftijdsopbouw en het wonen in een stad of een dorp. Gegevens die werden verzameld bestonden uit een gestandaardiseerd diagnostisch interview (MINI-Plus 5.0.0) en uit enkele beoordelingsschalen (BAS, MADRS). Daarnaast werden door de patiënt of deelnemer een aantal zelfrapportage vragenlijsten ingevuld. De interviewers werden uitgebreid getraind en begeleid, waardoor de interbeoordelaars-betrouwbaarheid en validiteit van de metingen geoptimaliseerd werden.

In de hoofdstukken 3 tot en met 6 worden de gevonden referentiewaarden voor de 18 (reeds bestaande) vragenlijsten besproken. De belangrijkste referentiewaarden betreffen het 95ste percentiel van de huisartspatiënten (P_{95} ROM-referentiegroep) en het 5de percentiel

van de psychiatrische patiënten (P_5 ROM-patiëntengroep). Deze zijn weergegeven in Tabel I en Tabel II van de Appendix. Deze referentiewaarden kunnen een verantwoordelijke, klinische besluitvorming met betrekking tot het initiëren, aanpassen of beëindigen van de behandeling ondersteunen. Daarnaast kunnen ze beslissingen ondersteunen met betrekking tot het doorverwijzen van patiënten uit de GGZ (gespecialiseerde geestelijke gezondheidszorg) terug naar de huisarts en andersom van de huisartspraktijk naar de GGZ.

De 18 vragenlijsten omvatten de volgende 4 clusters:
1) Vier generieke vragenlijsten: Brief Symptom Inventory (BSI), Mood & Angst Symptom Questionnaire 30-item (MASQ-D30), Short Form 36 (SF36) en Dimensional Assessment of Personality Pathology–short form (DAPP-SF)

2) Drie vragenlijsten die depressie meten: Beck Depressie Inventory-II (BDI-II), Inventory of Depressive Symptomatology - Self-Report (IDS-SR) en Montgomery Åsberg Depressie Rating Scale (MADRS);

3) Acht vragenlijsten die angst meten: Brief Scale for Angst (BSA), PADUA Inventory Revised (PI-R), Panic Appraisal Inventory (PAI), Penn State Worry Questionnaire (PSWQ), Worry Domains Questionnaire (WDQ), Social Interaction and Angst Scale (SIAS), Social Phobia Scale (SPS), and Impact of Event Scale – Revised (IES-R);

4) Drie vragenlijsten die somatoforme stoornissen meten: Body Image Concern Inventory (BICI), Whitely Index (WI) en Checklist Individual Strength (CIS20R).

Het eerste en tweede cluster vragenlijsten werden aangeboden aan alle ROM-referentiegroep deelnemers. Vervolgens beantwoordde 50% van deze groep het cluster met de vragenlijsten die angst meten, de andere 50% vulde de vragenlijsten in die somatoforme stoornissen meten. Alle ROM-patiënten werd het cluster generieke vragenlijsten aangeboden, ongeacht hun stoornis. De stoornisspecifieke clusters werden alleen aangeboden aan die patiënten die met de betreffende stoornis(sen) gediagnosticeerd werden.

100%	
Generieke vragenlijsten: - Brief Symptom Inventory (BSI) - Mood & Anxiety Symptom Questionnaire 30-item (MASQ-D30) - Short Form 36 (SF36) - Dimensional Assessment of Personality Pathology–short form (DAPP-SF) **Depressie vragenlijsten:** - Beck Depression Inventory-II (BDI-II) - Inventory of Depressive Symptomatology - Self-Report (IDS-SR) - Montgomery Åsberg Depression Rating Scale (MADRS);	
50%	**50%**
Angst vragenlijsten: - Brief Scale for Anxiety (BSA) - PADUA Inventory Revised (PI-R) - Panic Appraisal Inventory (PAI) - Penn State Worry Questionnaire (PSWQ) - Worry Domains Questionnaire (WDQ) - Social Interaction and Anxiety Scale (SIAS) - Social Phobia Scale (SPS) - Impact of Event Scale – Revised (IES-R)	**Somatoforme vragenlijsten:** - Body Image Concern Inventory (BICI) - Whitely Index (WI) - Checklist Individual Strength (CIS20R)

De gegevens illustreren sekse-specifieke resultaten. Er was een tendens voor vrouwen in de ROM-referentiegroep om hoger te scoren (en dus meer klachten te rapporteren) op alle vragenlijsten, behalve op de DAPP-SF. Het onderscheidend vermogen van de generieke vragenlijsten was goed met uitzondering van dezelfde DAPP-SF, die daar niet voor ontworpen is. De stoornisspecifieke vragenlijsten lieten een zeer goed onderscheidend vermogen zien. De interne consistentie varieerde van voldoende tot uitstekend voor alle totaalscores.

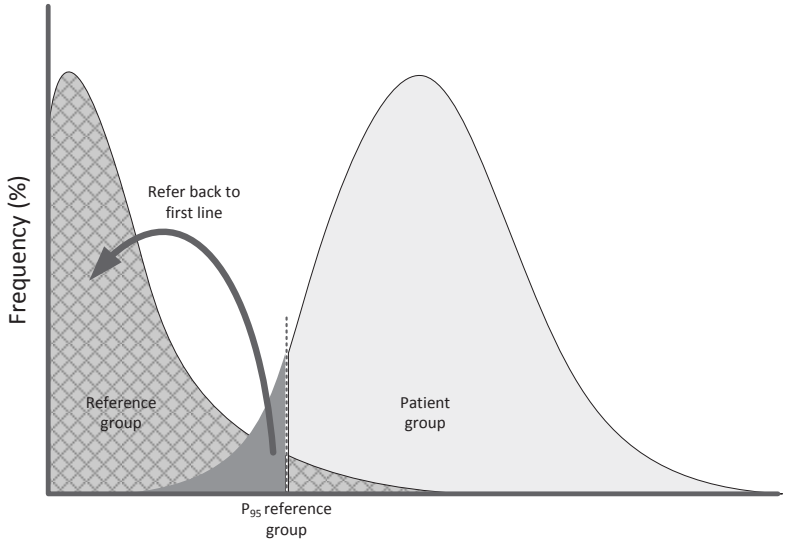
Hoofdstuk 7 beschrijft de ontwikkeling en validatie van onze nieuw ontwikkelde publieke domein vragenlijst, de 48-item Symptom Questionnaire (SQ-48). De SQ-48 werd ontwikkeld als generieke, multidimensionale vragenlijst. Het onderscheidend vermogen van de vragenlijst blijkt goed te zijn. De resultaten tonen een goede interne consistentie en goede convergente en divergente validiteit. Uiteraard werden ook voor deze SQ-48 referentiewaarden bepaald. De belangrijkste zijn weergegeven in de eerder genoemde Tabel I en Tabel II van de Appendix.

Klinisch gebruik van de referentiewaarden

Referentiewaarden kunnen aangeven of een patiënt voldoende hersteld is zodat de volgende stap in de behandeling gezet kan worden. Clinici in de GGZ kunnen bepaalde cut-off-waarden gebruiken om hun beslissingen te ondersteunen met betrekking tot beëindiging van de behandeling en mogelijke terugverwijzing naar de huisarts. Andersom kunnen huisartsen een (andere) set van cut-off-waarden gebruiken die hun beslissing met betrekking tot doorverwijzing naar de GGZ ondersteunen. Zo is de keuze van de cut-off-waarden afhankelijk van het doel waarvoor de cut-off-waarden gebruikt gaan worden.

Beslissingen betreffende het beëindigen van de GGZ behandeling en het mogelijk terugverwijzen naar de huisarts

Men zou kunnen stellen dat patiënten in behandeling gaan wanneer ze deel zijn gaan uitmaken van een patiënten-(klinisch zieke) populatie. Andersom kan beëindiging van de behandeling en terugverwijzing worden overwogen wanneer de patiënt niet langer deel uitmaakt van die zieke populatie, maar gaat behoren tot de referentie-('gezonde') populatie qua klachtenpatroon. Daarom is 'gezondheid' de conditie die moet worden opgespoord. Dit is het geval wanneer de patiënt behoort tot de range van 95% 'normalen' van de ROM-referentiegroep, want we definiëren 5% van de uiterste waarnemingen aan de hoge kant van de verdeling van de ROM-referentiegroep scores als afwijkend. De cut-off-waarde die de top 5% van de ROM-referentiegroep markeert, is gelijk aan de 95ste percentiel score: de P_{95} ROM-referentiegroep (zie Figuur 9.1). Deze cut-off-waarde kan worden beschouwd als een betrouwbare indicator voor de symptomatische gezondheid, omdat die zelden gezondheid mist bij diegenen die daadwerkelijk psychiatrisch gezien gezond zijn. Echter, hoge sensitiviteit voor gezondheid is gekoppeld aan lage sensitiviteit voor ziekte. Het nadeel van de lage sensitiviteit voor ziekte (met veel fout-negatieve resultaten) kan valse geruststelling over de afwezigheid van ziekte zijn. In Tabel 9.1 zijn de P_{95} ROM-referentiegroep cut-off-waarden van de 18 bestaande ROM-vragenlijsten en van de SQ-48 weergegeven.



Figuur 9.1: De cut-off-waarde die relevant is voor terugverwijzing naar de huisarts. GGZ-behandeling van een patiënt kan beëindigd worden wanneer de patiënt niet langer tot de patiëntenpopulatie behoort maar in plaats daarvan tot de referentiepopulatie: dit is onder de cut-off-waarde P_{95} ROM-referentiegroep.

Terugverwijzing naar de huisarts kan geïndiceerd zijn zelfs wanneer de patiënt nog enige restsymptomen heeft. Dit strookt met het feit dat een aanzienlijk deel van de huisartspatiënten wel enige klachten of symptomen heeft. Daarnaast kan terugverwijzing naar de huisarts worden geïndiceerd voor patiënten met terugkerende depressies voor de behandeling van eventuele resterende stemmingsklachten.

In Tabel 9.1 van de Appendix zijn de belangrijkste cut-off-waarden, de P_{95} ROM-referentiegroepwaarden, weergegeven voor de 18 reeds bestaande ROM- vragenlijsten plus de SQ-48. In vergelijking met eerder gepubliceerde cut-off-waarden zijn deze waarden over het algemeen hoger. Dit betekent dat een patiënt eerder in aanmerking zal komen voor terugverwijzing naar de huisarts en met meer restverschijnselen dan voorheen.

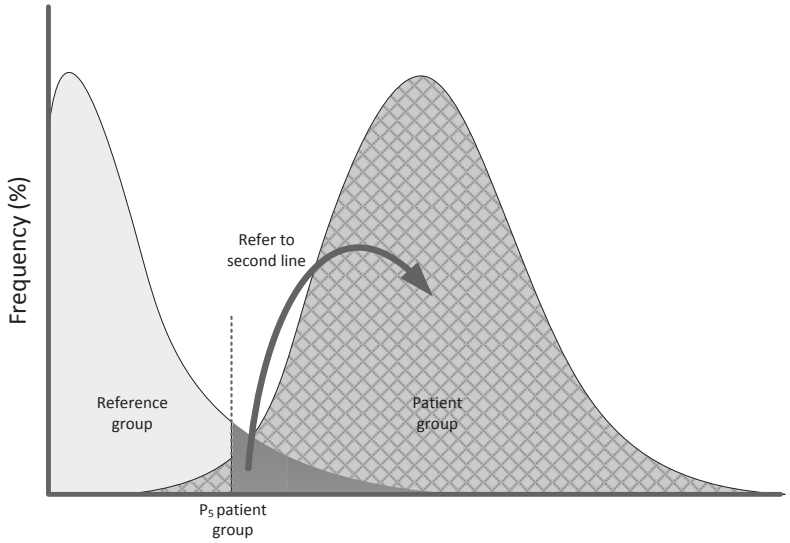
Er zijn patiënten die ondanks behandeling continu hoge ernstscores blijven houden. Het kan zijn dat voor hen de behandeling niet (langer) effectief is. Eerder gepubliceerde cut-off-waarden zouden suggereren dat deze patiënten nog niet klaar zijn om te worden terugverwezen naar de huisarts, terwijl onze cut-off-waarden een dergelijke beslissing wel zouden steunen. Aan de andere kant wil men niet dat de huisartspatiëntenpopulatie vergroot wordt met patiënten met veel restsymptomen en het daarbij horende verhoogde risico op terugval. De ROM-referentiegroep cut-off-waarden kunnen het maken van een goede afweging ondersteunen.

Beslissingen betreffende eventuele doorverwijzing vanuit de huisarts naar de GGZ

Doorverwijzing naar de GGZ kan aan de orde zijn wanneer de huisartspatiënt meer vergelijkbaar is met de patiëntenpopulatie dan met de referentiepopulatie. 'Ziekte' is dan de conditie die moet worden gedetecteerd. Dit is het geval wanneer de huisartspatiënt behoort tot de range van 95% 'zieken' van de ROM-patiëntengroep, want we definiëren 5% van de uiterste waarnemingen aan de lage kant van de verdeling van de ROM-patiëntengroepscores als afwijkend, ofwel gezond. De cut-off-waarde die de laagste 5% van de ROM-patiëntengroep markeert, is gelijk aan de 5^{de} percentielscore: de P_5 ROM-patiëntengroep (zie Figuur 9.2). Deze cut-off-waarde vertegenwoordigt hoge sensitiviteit voor psychopathologie. In Tabel 9.2 zijn de P_5 ROM-patiëntengroep cut-off-waarden van de 18 bestaande ROM vragenlijsten en van de SQ-48 weergegeven.

Bij gebruik van de P_5 ROM-patiëntengroepwaarden voor verwijzing naar de GGZ moet een aantal zaken in overweging worden genomen. De P_5 cut-off-waarde is zeer gevoelig voor de ziekte. Het kan worden beschouwd als een betrouwbare ondersteuning van de besluitvorming als het resultaat negatief is, omdat echte positieven (met psychopathologie) zelden worden gemist onder degenen die een positieve testuitslag hebben (d.w.z. dat de meeste zieke mensen ook als zodanig herkend worden). Echter, een hoge sensitiviteit is steeds gerelateerd aan een lage specificiteit. Lage specificiteit (d.w.z. veel fout positieve resultaten) houdt in dat relatief veel personen verdacht worden van een diagnose terwijl zij de ziekte niet

hebben. Bovendien kan het leiden tot extra tests en eventuele behandelingen die niet nodig of zelfs schadelijk kunnen zijn. of zelfs schadelijk kunnen zijn.



Figuur 9.2: De cut-off-waarde die relevant is voor verwijzing naar de GGZ. Patiënten kunnen daar behandeld worden wanneer ze niet langer deel uitmaken van de referentiepopulatie maar van de patiëntenpopulatie: dit is boven de cut-off-waarde P_5 ROM-patiëntengroep.

De afweging of doorverwijzing naar de GGZ nodig is, kan moeilijk zijn vanwege de vaak relatief subjectieve aard van psychiatrische klachten. Bij depressieve klachten kan de stoornis onopgemerkt en onbehandeld blijven. De oorzaak zou kunnen zijn dat huisartsen geneigd zijn meer in te spelen op het algemene niveau van onwelzijn dan de formele criteria voor depressie te testen. Een ander obstakel voor doorverwijzing naar de GGZ zouden bezwaren van de patiënt tegen de verwachte langdurige behandeling kunnen zijn. De referentiewaarden van de -studie kunnen klinische beslissingen ondersteunen m.b.t. een eventuele doorverwijzing naar de GGZ.

Algemene discussie

In onze analyses zagen we sekseverschillen in referentiewaarden in de ROM-referentiegroep: in vergelijking met mannen rapporteerden vrouwen ernstigere symptomen, zoals gemeten met de meeste observationele en zelfrapportageschalen voor algemene psychopathologie, depressie, angst en de stoornis in de lichaamsbeleving. Er werden geen sekseverschillen gevonden voor de persoonlijkheidsvragenlijst, de hypochondrievragenlijst en de chronischevermoeidheidsvragenlijst. Deze bevindingen waren niet onverwacht, aangezien sekseverschillen vaak worden beschreven in de literatuur zowel voor goed gedefinieerde

patiëntengroepen als voor de algemene bevolking. Vrouwen rapporteren twee keer zoveel depressieve of angstklachten als mannen. Geslacht kan worden gerelateerd aan een aantal omgevingsfactoren en andere aspecten van psychopathologie zoals stressoren en de mate van blootstelling hieraan. Deze beïnvloeden een eventueel begin van de ziekte, hoe symptomen worden uitgedrukt, of patiënten hulp zoeken en hoe ze worden behandeld in de geestelijke gezondheidszorg. De vraag is wat er precies verschillend is tussen mannen en vrouwen. Zijn het de symptomen zelf, de gevoeligheid voor symptomen, de manier van presenteren, of de maatstaven voor wat nog acceptabel is m.b.t. psychische klachten? Krijgen (of nemen) vrouwen meer de gelegenheid om psychische klachten te bespreken bij de GGZ? De literatuur lijkt erop te wijzen dat het traditionele masculiene genderrol stereotype de belangrijkste factor is. Of zijn de verschillen veroorzaakt door de vragenlijsten en de criteria die worden gehanteerd in de geestelijke gezondheidszorg? Het gebruik van sekse-specifieke referentiewaarden voor generieke en SAS-vragenlijsten is echter open voor discussie. Als wordt aangenomen dat de cut-off-waarde P_{95} ROM-referentiegroep lager voor mannen zou zijn, zou het impliceren dat vrouwen, behandeld in de GGZ, eerder zouden kunnen worden terugverwezen naar de eerste lijn. Ze hebben dan gemiddeld meer restverschijnselen dan mannen.

Ook het effect van leeftijd op de referentiewaarden werd bestudeerd. Aangezien een duidelijk algemene leeftijdseffect ontbrak, hebben we besloten dit niet verder in detail te beschrijven.

De ROM-vragenlijsten zijn primair bedoeld voor de ernstmeting van de SAS-aandoeningen. Een bijkomend doel van de NormQuest-studie was om te testen of deze vragenlijsten het diagnostische proces kunnen ondersteunen. Het onderscheidend vermogen, indicatief voor de diagnostische mogelijkheden van vragenlijsten, bleek voor de meeste generieke vragenlijsten (behalve de DAPP-SF) goed te zijn en voor alle 14-stoornis-specifieke vragenlijsten was het uitstekend. Dus, hoewel de vragenlijsten in deze studie niet ontworpen zijn voor hun diagnostische waarde maar voor ernstmeting, suggereren de goede prestaties toch dat deze vragenlijsten (behalve DAPP-SF) het diagnostische proces wel kunnen ondersteunen. Deze vragenlijsten kunnen echter de MINI-Plus als diagnoseinstrument niet vervangen. De MINI-Plus richt zich op algemene psychopathologie, distress of algemeen functioneren. Het klinisch oordeel blijft natuurlijk een essentieel onderdeel vormen van deze observatielijst voor het stellen van de klinische hoofddiagnose. De MINI-Plus kan bovendien gebruikt worden bij patiënten met meer dan één aandoening.

Onze bevindingen zijn gedaan bij volwassenen tussen de 18 en 65 jaar oud. Toekomstig onderzoek zal daarom referentiewaarden voor kinderen en ouderen moeten bepalen. Daarnaast zou verder onderzocht kunnen worden of de arbitraire grens van 95% van een populatie als ‘normaal’ te definiëren (en 5% als ‘abnormaal’) gerechtvaardigd is. Referentiewaarden zijn al veel gebruikt in een groot deel van de medische wereld, maar minder in de geestelijke gezondheidszorg. Verder onderzoek naar de implementatie van ROM en de geleverde referentiewaarden zijn noodzakelijk.

Tabel 9.1: Cut-off-waarden voor de GGZ die beslissingen t.a.v. terugverwijzing naar de huisarts kunnen ondersteunen - P_{95} ROM-referentiegroep.

Vragenlijst	Domein	Cut-off
Symptom Questionnare 48 items (SQ-48)	Generiek	
Aggressie (AGGR)		5.0
Agorafobie (AGOR)		2.0
Angst (ANXI)		11.2
Cognitieve klachten (COGN)		11.0
Depressie (MOOD)		8.0
Somatische klachten (SOMA)		8.0
Sociale fobie (SOPH)		9.0
Vitaliteit/optimisme (VITA)		15.0
Brief Symptom Inventory (BSI)	Generiek	0.68
Mood & Anxiety Symptom Questionnaire 30-item (MASQ-D30)	Generiek depressie/angst	
Non-specifieke symptomen (NA)	depressie	23
Anhedonie (PA)	angst	29
Fysiologische hyperarousal (SA)		17
Short Form 36 (SF36 of RAND36)*	Generiek	
Fysiek Functioneren	Kwaliteit van leven	65
Beperkingen functioneren tgv lichamelijke klachten		13
Pijnklachten		54
Sociaal Functioneren		63
Geestelijke gezondheid		56
Beperkingen functioneren tgv emotionele klachten		33
Vitaliteit		40
Algemene gezondheidsbeleving		45
Dimensional Assessment of Personality Pathology–short form (DAPP-SF)	Generiek	
Volgzaam (bedeesdheid)	Persoonlijkheid	3.50
Cognitieve vertekening		2.33
Identiteitsproblemen		2.70
Affectieve labiliteit/onstabiliteit		3.50
Behoefte aan prikkels		3.38
Dwangmatigheid		4.00
Gesloten		3.63
Gebrekkige empathie		2.60
Passief-agressief		3.20
Niet gesteld op intimiteit		3.38
Dominantie		3.75
Bezorgdheid/angstig		3.50
Gedragsproblemen		2.13
Gebrek aan vertrouwen in de medemens		2.15
Sociale ontwijking		3.33
Narcisme		3.50
Onveilige hechting		3.33
Zelfbeschadiging		1.50
Beck Depression Inventory-II (BDI-II)	Depressie	13
Inventory of Depressive Symptomatology - Self-Report (IDS-SR)	Depressie	20
Montgomery Åsberg Depression Rating Scale (MADRS)	Depressie	11
Brief Anxiety Scale (BAS)	Angststoornis	11
PADUA Inventory Revised (PI-R)	OCS	43
Paniek Opinie Lijst (POL)	Paniekstoornis	37
Paniekverwachting		47
Catastrofale gevolgen van paniek (Total):		
Maten van zelfvertrouwen in omgaan met paniek		65
Penn State Worry Questionnaire (PSWQ)	Piekeren (patholog.)	66
Worry Domains Questionnaire (WDQ)	Piekeren (normaal)	74
Social Interaction and Anxiety Scale (SIAS)	Sociale angst	32
Social Phobia Scale (SPS)	Sociale angst	19
Impact of Event Scale – Revised (IES-R)[1] Total	PTSS	36
Body Image Concern Inventory (BICI)	Lichaamsbeleving	55
Whitely Index (WI)	Hypochondrie	6
Checklist Individuele Spankracht (CIS20R)	Chron. vermoeidheid	92

ROM - routine outcome monitoring; OCS - obsessief compulsieve stoornis; PTSS - posttraumatische stress stoornis.
*: P_5 ROM referentiegroep en P_{95} ROM patiëntengroep, aangezien hogere scores een beter functioneren aangeven

Table 9.2: Cut-off-waarden voor huisartsen die beslissingen t.a.v. verwijzing naar de GGZ kunnen ondersteunen - P_5 ROM-patiëntengroep

Vragenlijst	Domein	Cut-off
Symptom Questionnare 48 items (SQ-48)	Generiek	
Aggressie (AGGR)		0.0
Agorafobie (AGOR)		0.0
Angst (ANXI)		2.0
Cognitieve klachten (COGN)		3.0
Depressie (MOOD)		1.0
Somatische klachten (SOMA)		0.0
Sociale fobie (SOPH)		0.0
Vitaliteit/optimisme (VITA)		6.0
Brief Symptom Inventory (BSI)	Generiek	0.34
Mood & Anxiety Symptom Questionnaire 30-item (MASQ-D30)	Generiek	
Non-specifieke symptomen (NA)	depressie	17
Anhedonie (PA)	angst	17
Fysiologische hyperarousal (SA)		18
Short Form 36 (SF36 of RAND36)*	Generiek	
Fysiek Functioneren	Kwaliteit van leven	100
Beperkingen functioneren tgv lichamelijke klachten		100
Pijnklachten		100
Sociaal Functioneren		88
Geestelijke gezondheid		76
Beperkingen functioneren tgv emotionele klachten		100
Vitaliteit		65
Algemene gezondheidsbeleving		90
Dimensional Assessment of Personality Pathology–short form (DAPP-SF)	Generiek	
Volgzaam (bedeesdheid)	persoonlijkheid	1.25
Cognitieve vertekening		1.00
Identiteitsproblemen		1.33
Affectieve labiliteit/onstabiliteit		1.63
Behoefte aan prikkels		1.00
Dwangmatigheid		1.38
Gesloten		1.75
Gebrekkige empathie		1.00
Passief-agressief		1.40
Niet gesteld op intimiteit		1.13
Dominantie		1.13
Bezorgdheid/angstig		1.67
Gedragsproblemen		1.00
Gebrek aan vertrouwen in de medemens		1.00
Sociale ontwijking		1.17
Narcisme		1.10
Onveilige hechting		1.00
Zelfbeschadiging		1.00
Beck Depression Inventory-II (BDI-II)	Depressie	14
Inventory of Depressive Symptomatology - Self-Report (IDS-SR)	Depressie	18
Montgomery Åsberg Depression Rating Scale (MADRS)	Depressie	11
Brief Anxiety Scale (BAS)	Angststoornis	6
PADUA Inventory Revised (PI-R)	OCS	20
Paniek Opinie Lijst (POL)	Paniekstoornis	
Paniekverwachting		14
Catastrofale gevolgen van paniek (Total):		10
Maten van zelfvertrouwen in omgaan met paniek		29
Penn State Worry Questionnaire (PSWQ)	Piekeren (patholog.)	48
Worry Domains Questionnaire (WDQ)	Piekeren (normaal)	44
Social Interaction and Anxiety Scale (SIAS)	Sociale angst	18
Social Phobia Scale (SPS)	Sociale angst	11
Impact of Event Scale – Revised (IES-R)[1] Total	PTSS	19
Body Image Concern Inventory (BICI)	Lichaamsbeleving	39
Whitely Index (WI)	Hypochondrie	5
Checklist Individuele Spankracht (CIS20R)	Chron. vermoeidheid	74

ROM - routine outcome monitoring; OCS - obsessief compulsieve stoornis; PTSS - posttraumatische stress stoornis.

*: P_5 ROM referentiegroep en P_{95} ROM patiëntengroep, aangezien hogere scores een beter functioneren aangeven

LIST OF ABBREVIATIONS

AA	Anxious Arousal
AD	Anhedonic Depression
AD	Agoraphobia
AGO	Agoraphobia Scale
ANX	Anxiety
ASD	Acute stress disorder
AUC	Area Under the ROC Curve
BDI-II	Beck Depression Inventory version II
BICI	Body Image Concern Inventory
BSA	Brief Scale for Anxiety
BSI	Brief Symptom Inventory
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CFS	Chronic Fatigue Syndrome
CIDI	Composite International Diagnostic Interview
CIS20r	Checklist Individual Strength
COTAN	Commissie Testaangelegenheden Nederland
CPRS-SF	Comprehensive Psychopathological Rating Scale-Short Form
CVA	Cerebro vascular accident
Cα	Cronbach's alpha
DAPP-BQ	Dimensional Assessment of Personality Pathology – Basic Questionnaire
DAPP-sf	Dimensional Assessment of Personality Pathology - Short Form
DD	Dysthymic disorder
DEP	Depression
DSM-IV-TR	Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revision
e.g.	Exempli gratia (meaning: for example)
ECT	Electroconvulsive Therapy
ERB	Ethical Review Board
GAD	Generalized anxiety disorder
GAF	Global Assessment of Functioning Scale
GD	General Distress
GP	General practitioner
HOS	Hostility
HPA	Hypothalamic-pituitary-adrenal
i.e.	Id est (meaning: that is)
IDS-SR	Inventory of Depressive Symptoms (Self Report)
IES-R	Impact of Event Scale – Revised

I-S	Interpersonal Sensitivity
LUMC	Leiden University Medical Center
M	Mean
MADRS	Montgomery-Äsberg Depression Rating Scale
MAS	Mood, anxiety, and somatoform
MASQ-D30	Mood & Anxiety Symptom Questionnaire -30
MDD	Major depressive disorder
MINI Plus 5.0.0. *	Mini International Neuropsychiatric Interview Plus 5.0.0.
MOS	Rand Medical Outcome Study
NA	Negative Affect
NEMESIS	Netherlands Mental Health Survey and Incidence Study
O-C	Obsessive-Compulsive
OCD	Obsessive-compulsive disorder
PA	(lack of) Positive Affect
PADUA/PI-r	PADUA Inventory revised
PAI	Panic Appraisal Inventory
PAR	Paranoid Ideation
PHOB	Phobic Anxiety
PSWQ	Penn State Worry Questionnaire
PSY	Psychoticism
PTSD	Posttraumatic stress disorder
RD	Rivierduinen
RMSE	Root Mean Square Error of Approximation
ROC	Receiver Operating Characteristics
ROM	Routine Outcome Monitoring
SA	Somatic Arousal
SCID	Structured Clinical Interview for Diagnostic and Statistical Manual
SD	Standard deviation
SD	Standard deviation
SF-36	Short Form Health Survey 36
SIAS	Social Interaction and Anxiety Scale
SOM	Somatization
SoPD	Social phobia
SpPD	Specific phobia
SPS	Social Phobia Scale
SQ-48	Symptom Questionnaire -48 Items
vCPRS *	Abbreviated Comprehensive Psychopathological Rating Scale
WDQ	Worry Domains Questionnaire
WI	Whitely Index
WSQ	Web Screening Questionnaire for common mental disorders

Over de auteur

Curriculum Vitae

Yvonne Schulte-van Maaren was born on the 14th of January, 1954 in Oosterhout. In 1971 she graduated from secondary school (HBS-B) and started studying Chemistry and Physics at the University of Utrecht. Subsequently, she taught chemistry, physics and mathematics at secondary school in Eindhoven, Bladel, and Rotterdam. In 1984-'85 Yvonne lived in Houston (USA). From 1988 till 1993 she lived in Aberdeen (Scotland) where she taught physics and mathematics at a local College (Dutch Stream). Back in the Netherlands she started working for an educational institute. From 1995 till 1998 she was director of another educational institute. In 1998 she started her studies of psychology at the University of Leiden as a part-time student. In 2004 she obtained her degree cum laude, in Clinical & Health Psychology and in Child & Adolescent Psychology. From 2004 till 2008 she worked at the Department of Psychology of the University of Leiden, where she was involved in a literature study of psychosocial problems related to (cardio-) vascular diseases, teaching and supervising psychology students. She initiated and set up a PhD study of post traumatic stress disorder in traumatized train drivers. When funding of this study was stopped, in 2008, Yvonne was invited by the Department of Psychiatry at the Leiden University Medical Centre to start a PhD-project aimed at developing reference values for Routine Outcome Monitoring instruments. Since 2006, Yvonne trains cardiac patients in stress- and lifestyle management in the Rijnlands Rehabilitation Centre, Leiden. She also trains Master students psychology in basic therapeutic skills. Yvonne is married and has three children and three grandchildren.

The NormQuest team

Publications

1) Schulte-van Maaren YWM, Carlier IV, Giltay EJ, van Noorden MS, de Waal MW, van der Wee NJ, Zitman FG (2013). Reference values for mental health assessment instruments: objectives and methods of the Leiden Routine Outcome Monitoring Study. Journal of Evaluation in Clinical Practice, 19, 342-350.

2) Schulte-van Maaren YWM, Carlier IV, Zitman FG, van Hemert AM, de Waal MW, van Noorden MS, Giltay EJ (2012). Reference values for generic instruments used in routine outcome monitoring: the Leiden Routine Outcome Monitoring Study. BMC Psychiatry. DOI: 10.1186/1471-244X-12-203.

3) Schulte-van Maaren YWM, Carlier IV, Zitman FG, van Hemert AM, de Waal MW, van der Does AJW, van Noorden MS, Giltay EJ (2013). Reference values for major depression questionnaires: the Leiden Routine Outcome Monitoring Study. Journal of Affective Disorders, 149(1-3):342-9.

4) Schulte-van Maaren YWM, Giltay EJ, van Hemert AM, Zitman FG, de Waal MW, Carlier IV (2013). Reference values for anxiety questionnaires: the Leiden Routine Outcome Monitoring Study. Journal of Affective Disorders 150 (3), 1008–1018.

5) Schulte-van Maaren YWM, Giltay EJ, van Hemert AM, Zitman FG, de Waal MW,van Rood YR,. Carlier IV. Reference values for somatoform instruments: the Leiden Routine Outcome Monitoring Study. submitted

6) Carlier IV, Schulte-Van Maaren YWM, Wardenaar KJ, Giltay EJ, Van Noorden MS, Vergeer P, Zitman FG (2012). Development and validation of the 48-item Symptom Questionnaire (SQ-48) in patients with depressive, anxiety and somatoform disorders. Psychiatry Research, 200 (2), 904-910.

7) Blackwell SE, Rius-Ottenheim N, Schulte-van Maaren YWM, Carlier IV, Middelkoop VD, Zitman FG, Spinhoven P, Holmes EA, Giltay EJ (2013). Optimism and mental imagery: A possible cognitive marker to promote wellbeing? Psychiatry Research, 206 (1), 56-61.

8) De Boer J, Boersma SN, De Gucht VMJ, Maes S, Schulte-van Maaren YWM. Psychosociale problemen bij hart- en vaatziekten. Den Haag: Nederlandse Hartstichting. 2006.

Dankwoord

Het NormQuest onderzoek, waarvan dit proefschrift de resultaten presenteert, was een groot en bijzonder project. Ik wil heel graag de mensen bedanken die deze NormQuest periode tot een heel bijzondere hebben gemaakt.

Hooggeleerde Zitman, beste Frans, toen je me 5 jaar geleden vroeg om het NormQuest onderzoek op te zetten, had ik er nog geen idee van hoe uitgebreid dit onderzoek uiteindelijk zou worden. Veel meer deelnemers, meer vragenlijsten, meer interviewers dan we oorspronkelijk hadden gepland: geweldig dat dat kon. Het was bijzonder om met je samen te werken. Hooggeleerde van Hemert, beste Bert, ik vind het heel jammer dat je niet eerder bij dit onderzoek betrokken was. Ik had graag nog veel meer van je geleerd, vooral op het gebied van schrijven. Ingrid, het was heel speciaal hoe jij iedere keer weer de puntjes op de i wist te zetten. Ik denk met veel plezier terug aan onze gesprekken. Dank hier voor. Erik, het was heel fijn om met jou samen te werken. Dank voor je inspirerende en motiverende begeleiding. Voor mij was het heel belangrijk dat ik me door jou gerespecteerd wist en dat je me steeds weer steunde, zowel bij de analyses en het schrijven van alle manuscripten als op het persoonlijke vlak. Prof. dr. J. Gussekloo, Prof. dr. W. Heiser en dr. E de Beurs: veel dank dat jullie in mijn promotiecommissie zitting hebben willen nemen.

Gea, het was geweldig om samen met jou het praktische deel van het onderzoek zelf op te zetten en ons team van interviewers te begeleiden. Ik heb heel goede herinneringen aan onze bijzondere dicussies over inhoudelijke en minder inhoudelijke zaken en jouw luisterend oor in heftige tijden. Margot de Waal, het was heel fijn om met jou en met de afdeling PHEG samen te werken. Samen kijken naar de te verwachten response rates en daarop onze lijsten met prospects te baseren was iedere keer weer een plezier. Mirjam, ook van jouw praktische ervaringen met het uitvoeren van een groot onderzoek heb ik dankbaar gebruik gemaakt. Dank daar voor. Marion, fijn dat jij onze enorme hoeveelheid data zo goed beheert en dat ik, keer op keer, weer die data kreeg die ik nodig had. Marieke, dank voor de QuestManager coördinatie. Nic, ik heb jouw support bij het begeleiden van een groot team enorm gewaardeerd. Vivian, Alice en Lia, ook jullie secretariele hulp was heel fijn.

Arianne en Leonie, mijn hele promotieperiode waren jullie beiden er, met jullie steun, inhoudelijke adviezen en gezelligheid. Het was heel fijn samenwerken met jullie tijdens de ‘hartgroepen’. Maar bovenal ben ik heel blij met jullie beider vriendschap. Onze gastronomische en nautische ervaringen hebben zeker bijgedragen aan de wetenschappelijke verdieping van onze discussies! En dan de collega’s van de kantoortuin en van ‘boven’: Nienke, Denise, Anke, Steven, Justine, Jessica, Moji, Marloes, Viki, Sumayah, Liora en Monique. Dank voor jullie goede gezelschap en jullie collegialiteit.

Ik vond het geweldig dat ruim 1300 mensen, huisarts patienten die zelf niet onder behandeling waren voor psychische problemen, mee hebben willen doen aan dit onderzoek. Door 3 uur lang diverse vragen, mondeling en schriftelijk, te beantwoorden is dit onderzoek werkelijkheid geworden. Voor hun hulp bij het selecteren van deze deelnemers dank ik de huisartspraktijken: VEUR, Bootsma en Joustra (Leidschendam), Van Haastert en Roelen (Roelofarendsveen), De Lange (Leiden), Barnhoorn (Leiden), Schinkelshoek (Voorhout), Zaaijer, Zaaijer & Hensing (Leiden), Hammerstein & Meskers (Leiden) en Boenders (Leiden).

NormQuest was NormQuest niet geworden zonder het NormQuest-team, mijn geweldige interviewers en bellers. Met de interviewers van het eerste uur ging ik naar de verschillende deelnemers thuis. Het was dikke pret met jullie: Lennie Geerlings, Anna Coppoolse, Bart Huisman, Roy de Kleijn en Nathalie Boot. Mirjam vd Hoorn, Marian Lucas en Marieke de Wit, fijn dat jullie op het LUMC zoveel mensen hebben gezien. Na een half jaar hebben we het team uitgebreid met een 2de lichting interviewers, al even enthousiast en geweldig: Lilo Gerdessen, Charlotte Flierboom, Anne Raterman, Elise van Holsteijn, Karlijn Noest, Lotte van Reuler, Margina Ruiter, Nicole Billingy, Cyntha Bogaart en Vera Atema. De avonden boven op C8, samen met Gea, jullie interviewers en de bellers Tessa Kooistra, Jordy Fransz, Lisette Bedijn, Charlotte Pronk, Nikki Elbers en Brenda Riegman waren oergezellig! En dat kwam mede door de gastvrijheid van de C8-afdeling. Dank aan Barbara Romson, Nadine Schonberger en het verplegend personeel.

Zonder dierbare vrienden zou promoveren niet half zo leuk zijn geweest. David Heyne, dank je wel voor het corrigeren van mijn Engelstalige artikelen. Maar vooral: dank je voor de stimulerende gesprekken en de gezellige etentjes. Annemarie Smith, ik heb genoten van onze discussies over onderzoek doen, over genderverschillen in gerapporteerde psychopathologie, over promoveren en over lekker eten. Marijke Heijloo, het was heel fijn om tijdens het hardlopen over alle promotieperikelen met je te praten. Top dat jij zo snel en zo goed mijn Nederlandse samenvatting hebt bekeken. Arnold van Emmerik, het had zomaar gekund dat jij hier als copromotor was genoemd, als ons PTSS-project anders was uitgepakt. Dank voor je vriendschap en dank voor goede eerste ervaringen op het gebied van promotieonderzoek opzetten. Promotieplezier heb ik van heel nabij meegemaakt bij mijn vriendinnen, de “meisjes van de ondergang” van FSW: Marieke Tolenaar, Rimke Haringsma, Ilke Jellema, Jacobien van Peer, Emma Massey, Maarten Fischer (als enig niet-meisje) en Anja Greeven. Het plezier werkte aanstekelijk!

En dan heel belangrijk: Alle vrienden en familie, die de afgelopen 5 jaar veel te weinig tijd/aandacht van mij hebben gehad. Dank voor jullie geduld en begrip. Het wordt weer beter, dat beloof ik jullie!

Nienke en Jasper, ik ben heel erg blij dat jullie tijdens de verdediging naast mij staan als mijn paranimfen. Nienke, jouw niet aflatende steun en jouw bezorgdheid dat ik wel eens te veel hooi op mijn vork nam hebben me steeds weer laten voelen wat een lief mens jij bent. Jasper, jouw trots op wat ik deed en je relativerend vermogen, naast je humor, hebben me steeds weer veel goed gedaan.

Mijn geweldige andere dochter, Jolien, en mijn drie evenzo geweldige 'schoonkids' Marco, Boudewijn en Marielle en natuurlijk mijn allerliefste kleinkinderen Isabel, James en Matthijs: jullie hebben deze 5 jaren voor veel energie en plezier gezorgd.

Willem, mijn lief. Ruim 35 jaar geleden hebben we samen jouw promotietraject doorlopen. Het was een heftige periode waarin ik je op allerlei mogelijke manieren heb geprobeerd steunen. Die periode heeft tot veel goeds geleid. Nu, tijdens mijn promotietraject, was jij er steeds met praktische steun, met advies, met heerlijke maaltijden plus een kaarsje als ik moe thuis kwam, met humor en, vooral, met veel liefde. Op naar ons volgende project!

Yvonne, Gea en een aantal interviewers

Referentiewaarden ROM: Generieke lijsten

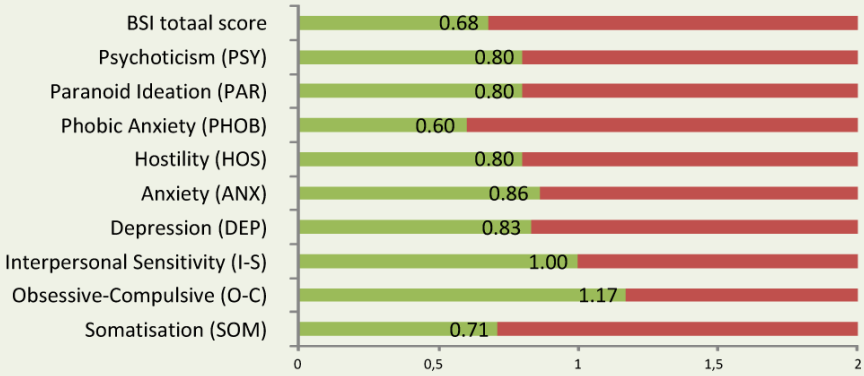
SQ48 (48-item Symptom Questionnaire)

Psychiatry Research 2012; 200: 904–910

Bereik
0 – 24
0 – 20
0 – 28
0 – 24
0 – 20
0 – 24
0 – 16
0 – 16

BSI (Brief Symptom Questionnaire)

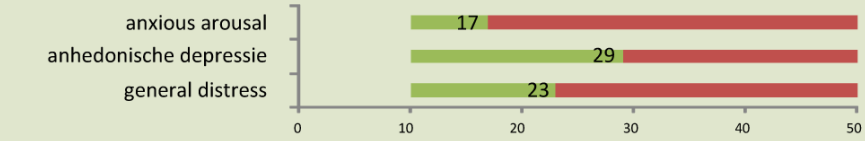
BMC Psychiatry 2012: 12; 203



Bereik
0 – 2
0 – 2
0 – 2
0 – 2
0 – 2
0 – 2
0 – 2
0 – 2
0 – 2
0 – 2

MASQ (Mood & Anxiety Symptom Questionnaire)

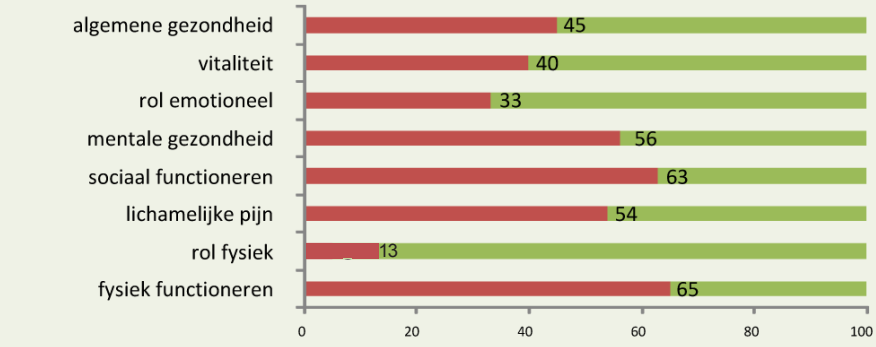
BMC Psychiatry 2012: 12; 203



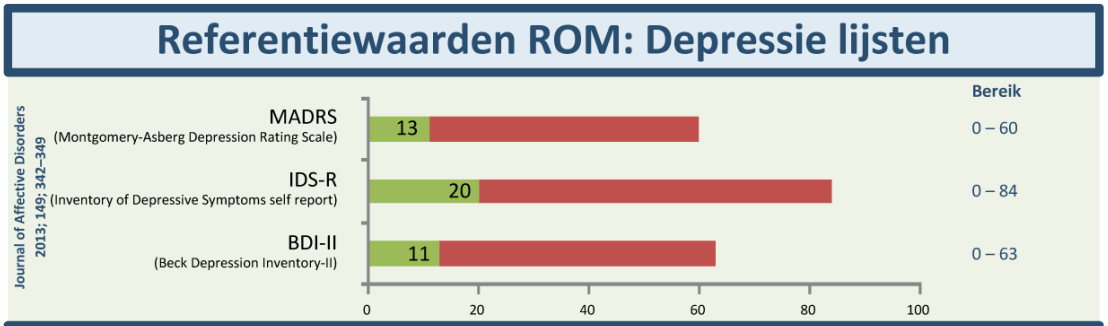
Bereik
10 – 50
10 – 50
10 – 50

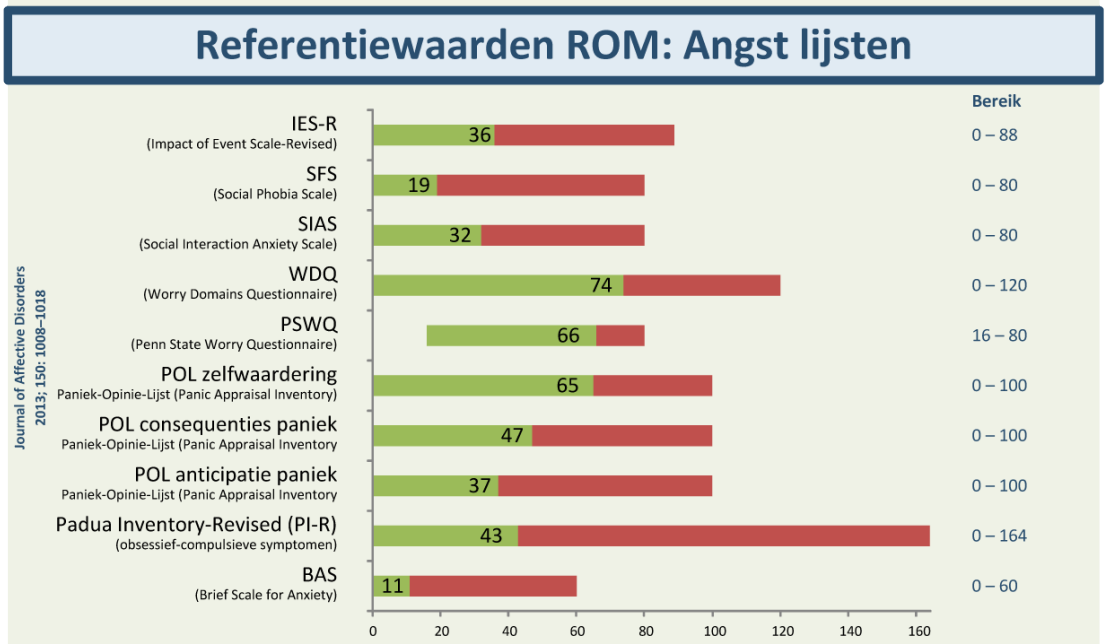
SF-36 (Short Form 36, kwaliteit van leven)

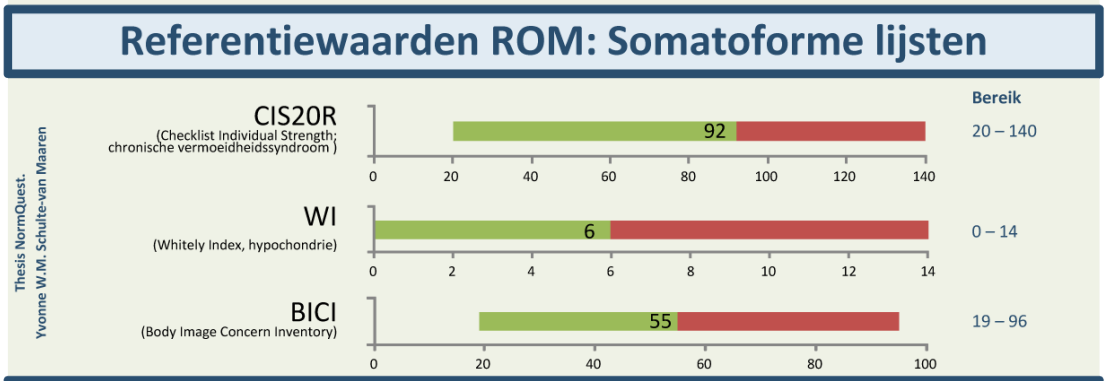
BMC Psychiatry 2012: 12; 203

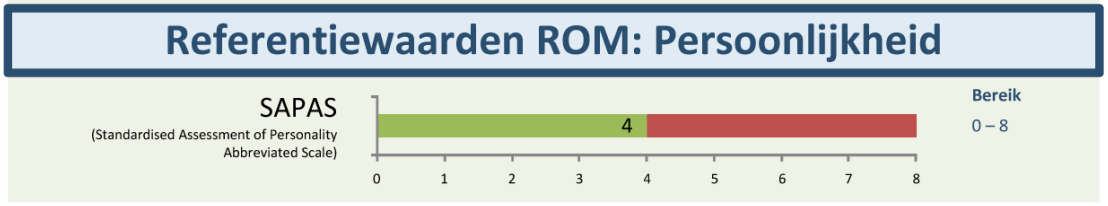


Bereik
0 – 100
0 – 100
0 – 100
0 – 100
0 – 100
0 – 100
0 – 100
0 – 100









Met dank aan A. Schat voor de opmaak van dit overzicht.