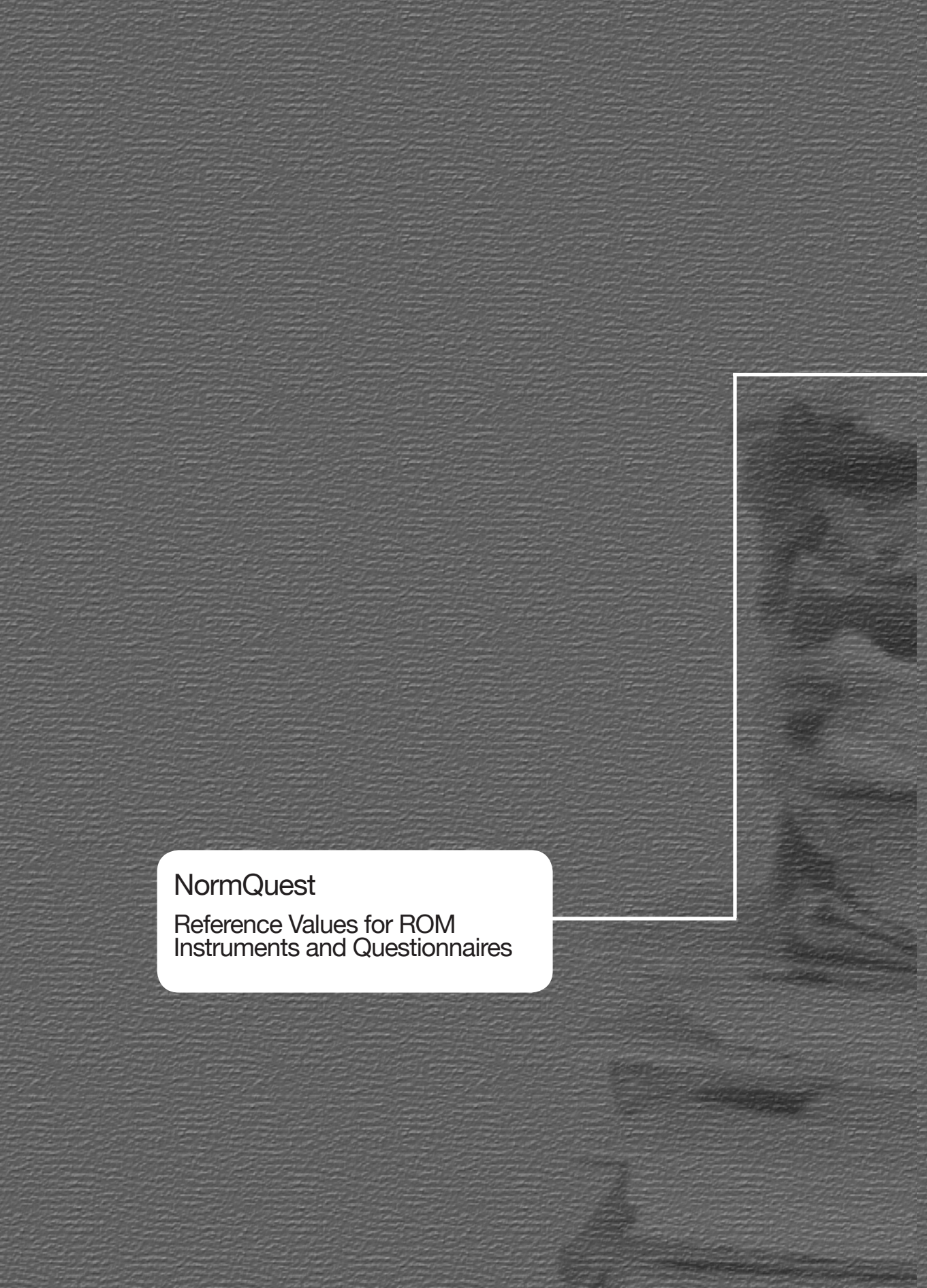
NormQuest : reference values for ROM instruments and questionnaires

Schulte-van Maaren, Y.W.M.



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.

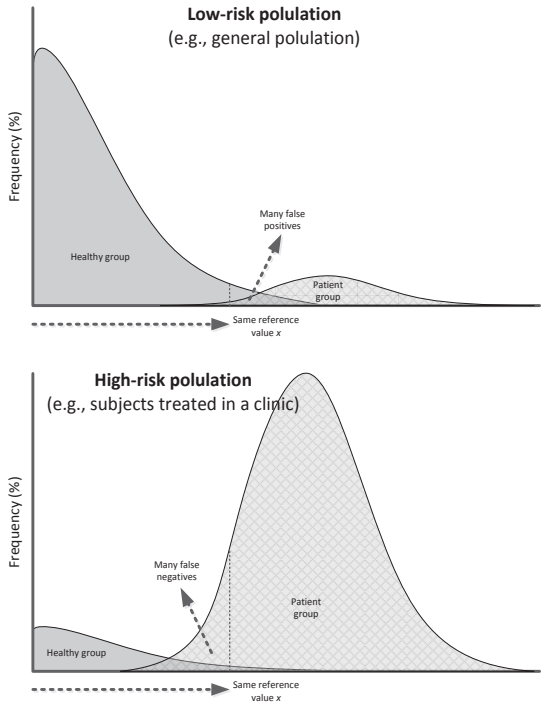


Figure 8.1. Prevalence-dependent cut-off values. Top: low prevalence of the psychiatric disorder: a fixed reference value x results in many false positives. Bottom: high prevalence of the psychiatric disorder in a high-risk population: the same fixed reference value x now results in many false negatives.

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.

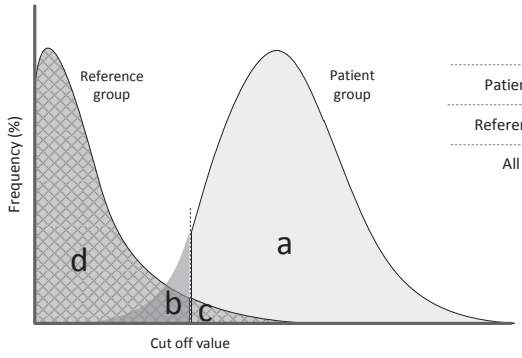
Generalizability of the reference values was another aim in this study. As noted before, the NormQuest sample was representative for the gender and age distributions of the ROM patient group. Therefore, its reference values can be validly used as a comparison against this patient group. However, several 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.

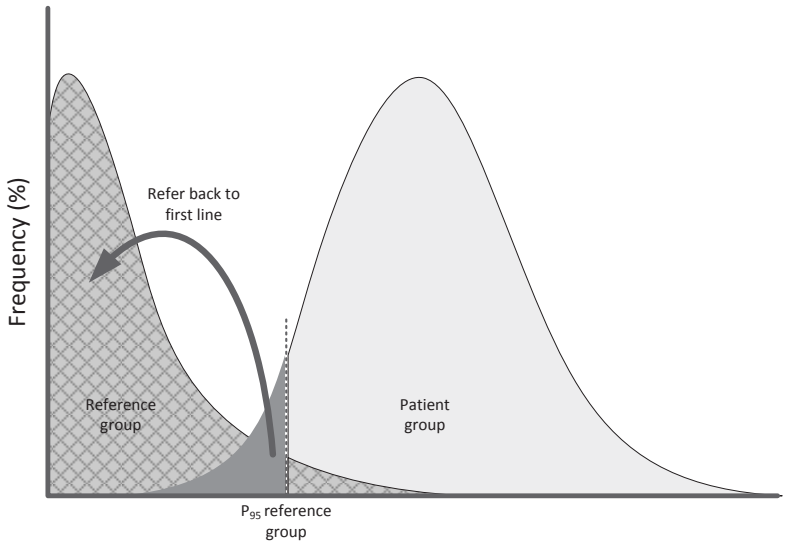


Figure 8.3: Cut-off values relevant for referral back to primary care. Patients depart from treatment when they no longer belong to the patient population, but belong to the reference population instead, below the cut-off value P_{95} ROM reference group.

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.

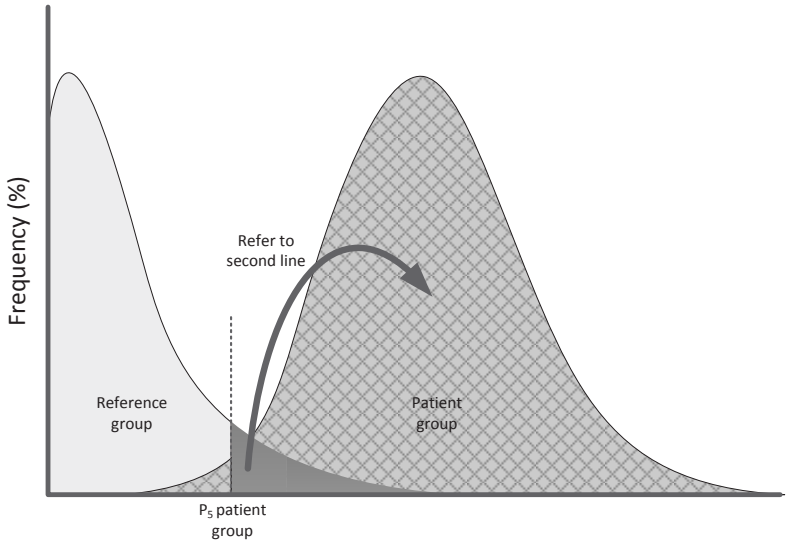


Figure 8.4: Cut-off value relevant for referral to specialized mental health care. Patients enter secondary treatment when they are no longer part of the reference population, but belong to the patient population instead, above the cut-off value P_5 ROM patient group.

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.

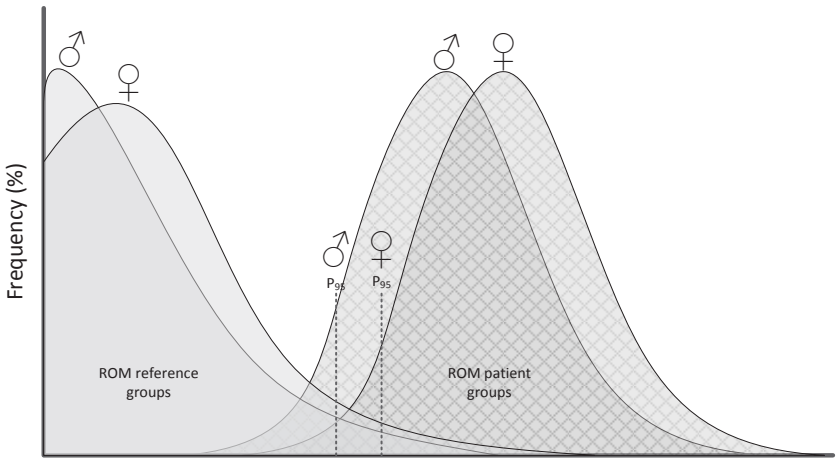


Figure 8.5. Hypothetical distribution of the scores of a questionnaire measuring psychopathology within the ROM reference group and within the ROM patient group. Scores are gender-specific: for women and men separately.

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