

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/23044> holds various files of this Leiden University dissertation

Author: Schulte-van Maaren, Yvonne W.M.

Title: NormQuest : reference values for ROM instruments and questionnaires

Issue Date: 2014-01-21

NormQuest

Reference Values for ROM
Instruments and Questionnaires

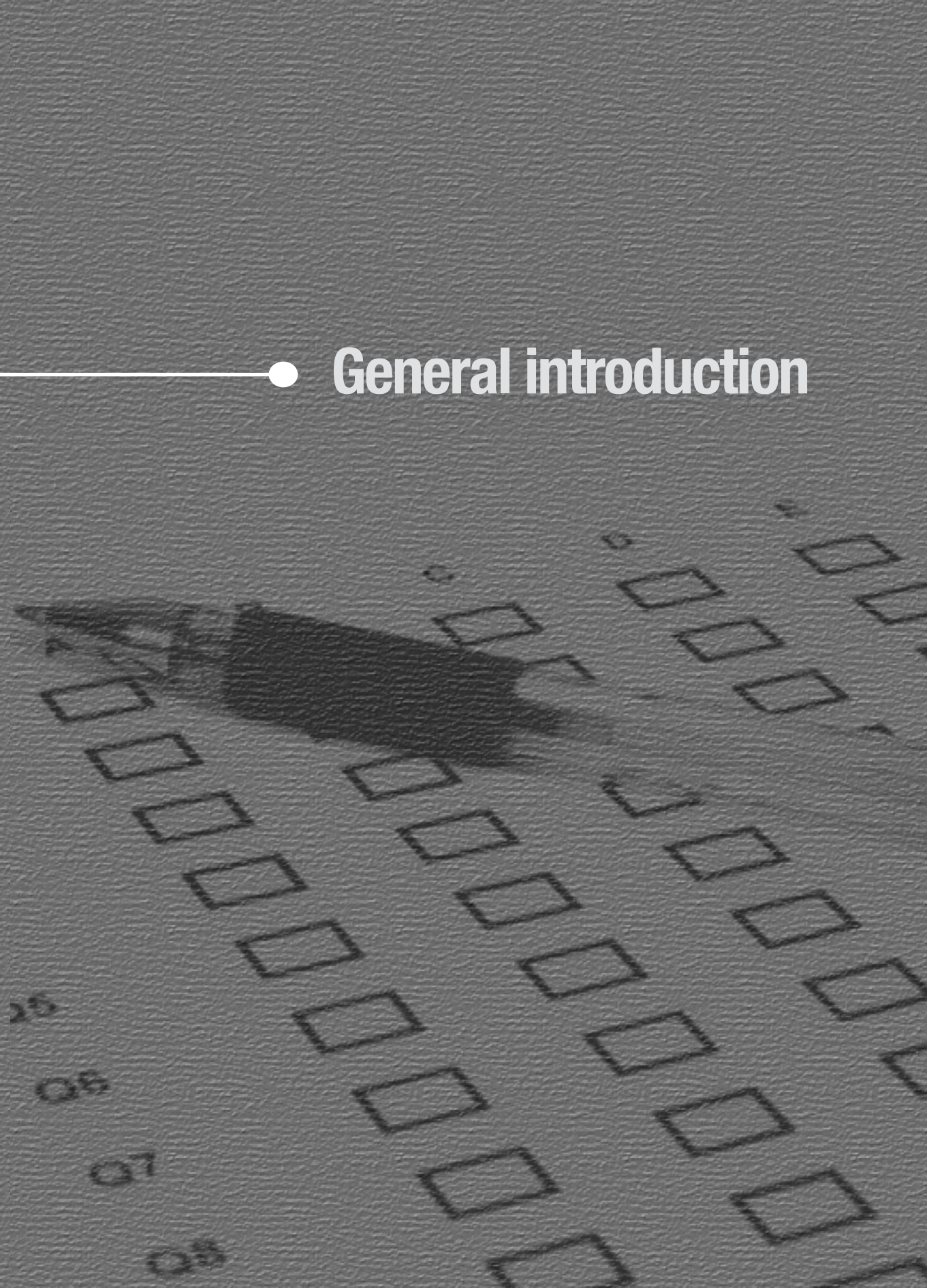
Q1

Q2

Q3

Q4

C

The background of the slide is a grayscale image of a document. It features a grid of many small, empty square checkboxes arranged in rows and columns. A dark, textured object, possibly a pen or a piece of paper, is placed diagonally across the middle of the grid. In the bottom left corner, there are some faint, partially visible labels: 'Q6', 'Q7', and 'Q8'.

● 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-Åsberg Depression Rating Scale (MADRS), where a higher score means more psychopathology (see Figure 1.1).

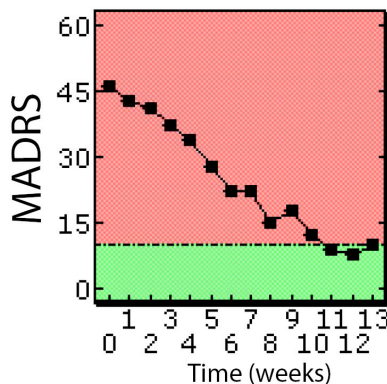


Figure 1.1. ROM graph of MADRS scores of 12 consecutive assessments of an ECT treated patient diagnosed with major depressive disorder.

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.

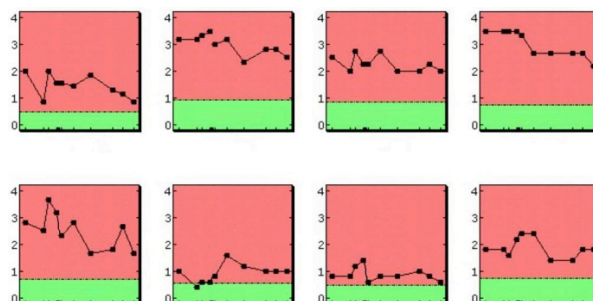
At the start, ROM was restricted to patients referred for treatment of mood, anxiety, and somatoform (MAS) disorders. These patients form a relatively homogenous group with substantial mutual comorbidity (Kessler et al., 1996) and they mainly receive outpatient care. To be eligible, patients had to have 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 (5th, 25th, 50th, 75th, and 95th) 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.

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

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

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

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. ROC 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).

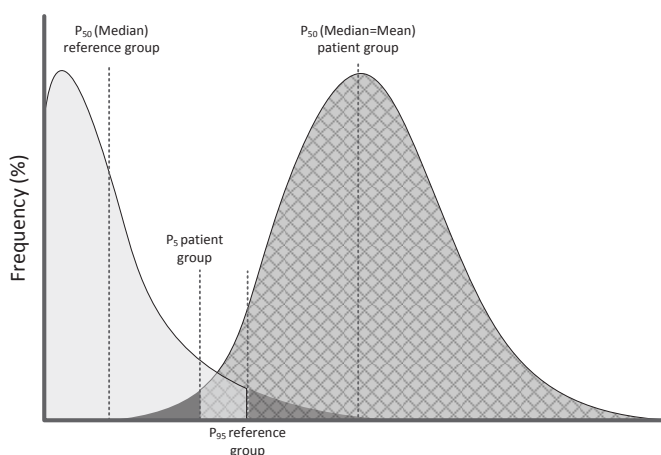


Figure 1.2. Hypothetical distribution of the scores of a questionnaire measuring psychopathology within the reference population and within the patient population. Two cut-off values are depicted: the 95th percentile score (P_{95}) of the reference group and the 5th percentile score of the patient group. 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).

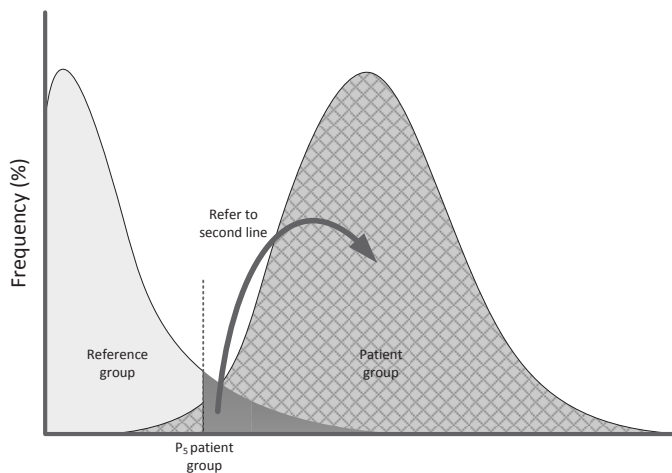


Figure 1.3. Cut-off values relevant for referral from primary care to secondary care. Patients enter treatment when they are no longer part of the reference population, but belong to the patient population instead, above the cut-off value P_5 of the patient group.

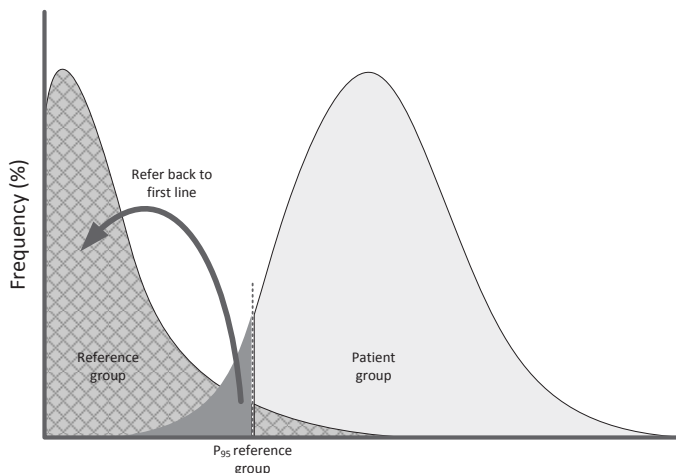


Figure 1.4. Cut-off values relevant for referral from secondary care to primary care. Patients depart from treatment when they no longer belong to the patient population, but belong to the reference population instead, below the cut-off value P_{95} of the reference group.

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 P₅ 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 P₉₅ 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-Åsberg 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.
7. 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, Buntinx 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, Schaefer 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.