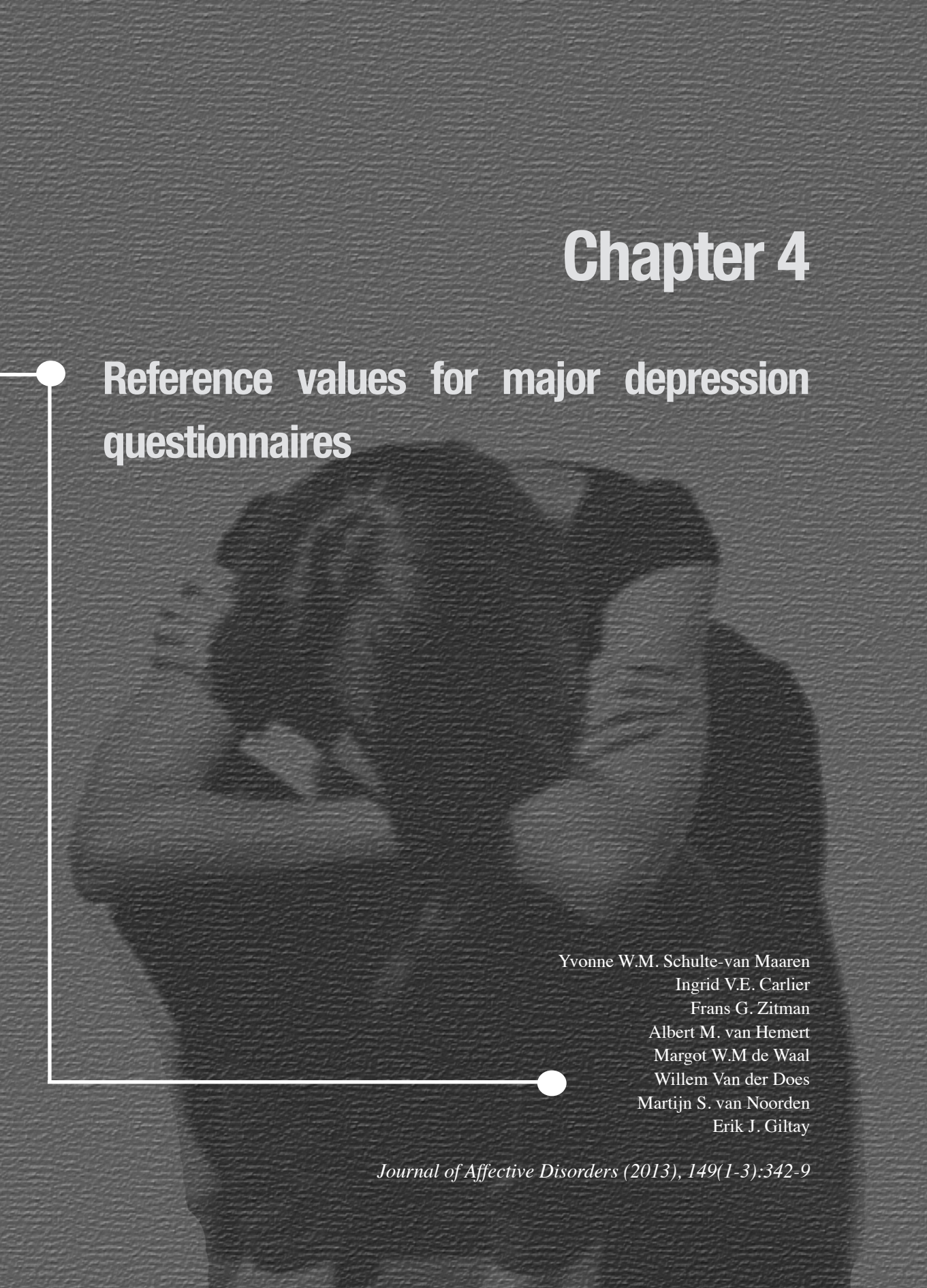
NormQuest

Reference Values for ROM Instruments and Questionnaires

Chapter 4

Reference values for major depression questionnaires



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

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

ABSTRACT

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

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

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

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

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

INTRODUCTION

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

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

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

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

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

METHODS

Participants

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

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

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

Procedures and instruments

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

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

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

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

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

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

Statistical analyses

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

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

RESULTS

Statistical analyses

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

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

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

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

Table 4.1: continued.

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

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

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

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

REFERENCE VALUES

Percentiles, means and SDs

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

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

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

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

Receiver operating characteristic (ROC) curves

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

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

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

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

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

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

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

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

Internal consistency reliability

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

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

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

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

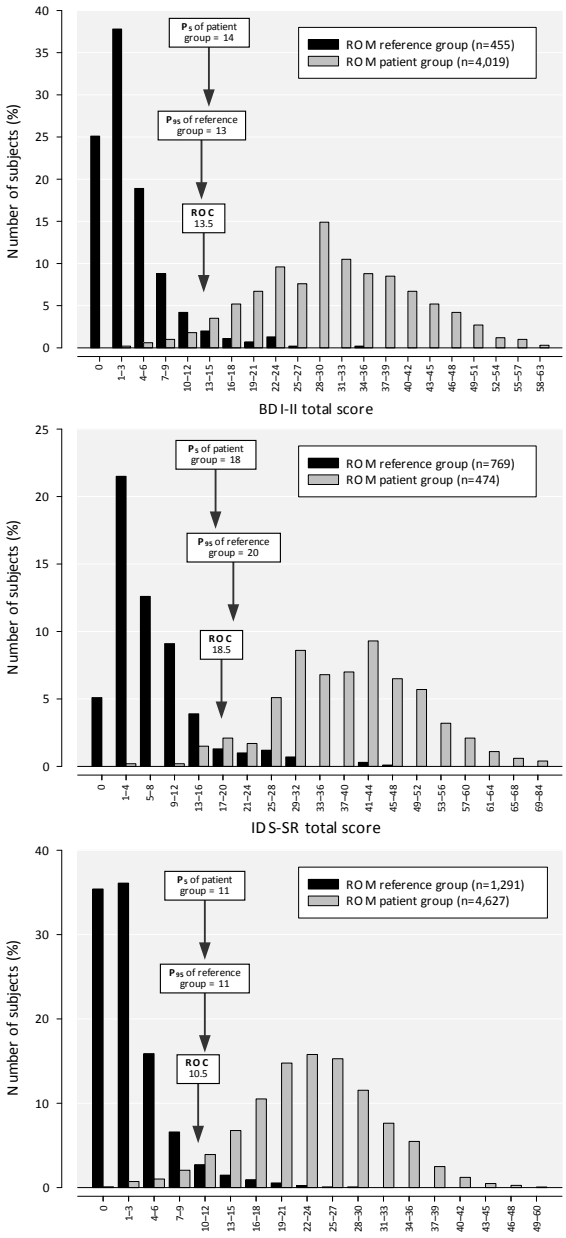


Figure 4.1: Distribution of the scores of Beck Depression Inventory-II (BDI-II) total scale, the Inventory of Depressive Symptoms (self-report) (IDS-SR) total scale and the Montgomery Äsberg Depression Rating Scale (MADRS). Three types of cut-off points are depicted: the 75th percentile score (P_{75}),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

DISCUSSION AND CONCLUSION

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

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

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

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

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

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

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

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

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

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

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

Reference List

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Supplementary Material

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

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

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

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

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

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

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