Journal of Affective Disorders 149 (2013) 342–349

Contents lists available at SciVerse ScienceDirect

Journal of Affective Disorders

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

Research report

Reference values for major depression questionnaires: The Leiden Routine Outcome Monitoring Study

Yvonne W.M. Schulte-van Maaren [a,*], Ingrid V.E. Carlier [a], Frans G. Zitman [a], Albert M. van Hemert [a], Margot W.M. de Waal [b], A.J. Willem van der Does [a], Martijn S. van Noorden [a], Erik J. Giltay [a]

[a] Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands
[b] Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands

ARTICLE INFO

Article history:
Received 10 September 2012
Received in revised form
8 November 2012
Accepted 4 February 2013
Available online 29 March 2013

Keywords:
Reference values
Routine outcome monitoring
Questionnaires
Mood disorders
Major depressive disorder
Depression

ABSTRACT

Background: The Beck Depression Inventory-II (BDI-II), the Inventory of Depressive Symptoms (self-report) (IDS-SR) and the Montgomery-Äsberg Depression Rating Scale (MADRS) are questionnaires that assess symptom severity in patients with a depressive disorder, often part of Routine Outcome Monitoring (ROM). We aimed to generate reference values for both "healthy" and "clinically depressed" populations.

Methods: 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 dysthymia (ROM patient-group). The outermost 5% of observations were used to define limits for one-sided reference intervals (95th percentiles; P_{95}). Receiver Operating Characteristics (ROC) analyses were used to yield alternative cut-off values. Internal consistency was assessed.

Results: The mean age was 40.3 yr (SD = 12.6) and 39.3 (SD = 12.3) for the ROM reference and patient-groups, respectively, and 62.8% versus 61.0% were female. Cut-off (P_{95}) values differed for women and men, being respectively 15 and 12 for the BDI-II, 23 and 18 for the IDS-SR, and 12.5 and 9 for the MADRS. ROC analyses yielded almost equal reference values. The discriminative power of the BDI-II, IDS-SR and MADRS scores was very high. Internal consistency was excellent for total scores and satisfactory for all subscales, except for the IDS-SR subscale Atypical Characteristics.

Limitations: Substantial non-response and limited generalizability.

Conclusions: For the BDI-II, IDS-SR and MADRS a comprehensive set of reference values were provided. Reference values were higher in women than in men, implying the use of sex-specific cut-off values. Either instrument can be offered to every patient with MAS disorders to make responsible decisions about continuing, changing or terminating therapy.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Routine Outcome Monitoring (ROM) is the periodically repeated assessment of the condition of patients using diagnostic instruments and severity scales that may aid the evaluation of treatment. Both generic and disorder-specific measurement instruments are used. Generic instruments are completed by all patients and assess a broad range of psychopathological symptoms irrespective of their psychiatric disorder(s). Disorder-specific instruments are administered to patients who meet the criteria for a particular disorder (De Beurs et al., 2011; Schulte-van Maaren et al., 2013).

Reliable ratings from reference populations are essential to correctly interpret ROM results for clinical decisions on continuing, altering or terminating treatment (Kazdin, 2008). Furthermore, they can facilitate referral from specialized mental health care back to primary care. When establishing and interpreting reference values, several aspects need to be considered. First, reference values (Solberg and Grasbeck, 1989) are often established in healthy populations (Solberg, 2008) with health clearly defined (a priori) by inclusion and exclusion criteria (Geffre et al., 2009; Katayev et al., 2010; Sasse et al., 2000). As a consequence "super-normal" (i.e., too healthy) participants are selected (Kendall et al., 1999) resulting in unreasonable (often 10% narrower) reference intervals (Horn et al., 2001). Second, (sub)sample sizes of at least 120 are needed to reduce the amount of uncertainty and error

* Correspondence to: Department of Psychiatry, Leiden University Medical Center, P.O. Box 9600, 2300 RC Leiden, Leiden, The Netherlands.
Tel.: +31 71 5261175/3785; fax: +31 71 5248156.
E-mail addresses: Y.W.M.Schulte-van_Maaren@lumc.nl, yvonne_schulte@hotmail.com (Y.W.M. Schulte-van Maaren).

0165-0327/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jad.2013.02.009

caused by potential outliers (Reed et al., 1971). Third, when data tend toward a non-Gaussian distribution non-parametric percentile scores (Reed et al., 1971; Solberg, 2008) are more appropriate reference values than parametric mean (and standard deviation (SD) of confidence interval) values or weighted cut-off scores calculated by the Jacobson & Truax method (Jacobson and Truax, 1991). In that case, the 95th percentile commonly serves as the reference value (Solberg, 2008). 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 (Fan et al., 2006).

The self-report Beck Depression Inventory-II (BDI-II) (Beck et al., 2002; Beck and Steer, 1987), the Inventory of Depressive Symptoms-Self-Report (IDS-SR) (Nolen and Dingemans, 2004; Rush et al., 1996) and the observer-rated Montgomery Äsberg Depression Rating Scale (MADRS) (Montgomery and Asberg, 1979) are three frequently used ROM instruments that assess symptom severity of Major Depressive Disorder (MDD). The BDI-II, rather than the IDS-SR and MADRS, is also used as a diagnostic screening instrument for MDD (Nolen and Dingemans, 2004). Previous BDI-II studies reported cut-off and reference values for MDD outpatients (Beck et al., 1996b; Steer et al., 1999) and inpatients (Kumar et al., 2002; Steer et al., 2000). A study in 376 undergraduates (17–29 yr of age) and older adults (55–90 yr of age) (Segal et al., 2008a) reported a mean of 8.6 (SD=7.7). IDS-SR reference values were reported for depressed outpatients (Biggs et al., 2000; Rush et al., 1986; Rush et al., 1996; Trivedi et al. 2004) and inpatients (Corruble et al. 1999). For 118 normal controls a cut-off value of ≥18 was reported (Rush et al., 1996) and a mean of 2.1 (SD=2.2) for 23 normal controls. Many studies have reported means with SDs or cut-off values for the MADRS studies but were done with outpatients with MDD (Bondolfi et al., 2010; Hawley et al., 2002; Khan et al., 2002; Zimmerman et al., 2004b), inpatients with MDD (Hawley et al., 2002; Müller et al., 2000,2003), stroke patients (Sagen et al., 2009) and old age pensioners (Engedal et al., 2012). When all studies in healthy controls (total $n=569$) were reviewed means with SDs and the optimal cut-off scores were reported (Zimmerman et al., 2004a). However, because of the strongly positively skewed distribution of these total scores in healthy populations, the assumption of a normal distribution does not seem to be satisfied in these studies. Therefore, reference values should preferably be based on a distribution-free percentile or ROC methodologies.

The aim of this study was to establish reference values, including percentiles, ROC-based cut-off points next to the more commonly reported means with SDs for the BDI-II, IDS-SR and MADRS. A sample of 1295 subjects from the general population, recruited through General Practitioners (GPs), was compared with a sample of 4627 outpatients diagnosed with MDD or dysthymia (with or without other psychiatric disorders). Particular aspects of the current study were that next to a well-defined psychiatric patient-group, a healthy, but not necessarily symptom free, reference-group was included, and that both sample sizes were large.

2. Methods

2.1. 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; mean age=40.3 yr; SD=12.6) 18–65 yr old were included in the ROM reference-group, as part of the "Leiden Routine Outcome Monitoring Study" (Schulte-van Maaren et al., 2013; Van Noorden et al., 2010). In order to recruit a representative general population sample participants were randomly selected from the registration systems of eight GPs in the Leiden region. In the Netherlands, 99.9% of the general population is registered with a GP (Poortvliet et al., 2005). Participants receiving treatment for psychiatric disorders and/or dependence on alcohol or drugs within six months prior to the assessment were excluded because the group was aimed to be a psychiatrically healthy reference-group. Further exclusion criteria were hearing impairment or limited cognitive or language abilities (i.e., aphasia, severe dyslexia or dementia; illiteracy or insufficient mastery of the Dutch language). To make the group demographically comparable to the ROM patient-group the reference-group was matched for gender, age and urbanization-level (62.3% urban). Participants in the reference-group completed the same assessments as the patient-group, except that the ROM reference-group completed *every* disorder-specific instrument. The response rate was 37.1%, as detailed previously (Schulte-van Maaren et al., 2013; Schulte-van Maaren et al., 2012). Invitations for this study were sent to 4840 persons; 1283 could not be contacted and 67 were not included because of time constraints. Of the remaining 3490 potential participants, we could generate complete datasets of 1295 individuals.

The ROM patient-group consisted of a baseline sample of 4627 psychiatric outpatients, aged between 18 and 65 years (61.0% females; mean age=39.3, SD=12.3), diagnosed with and treated for depressive disorders (MDD or dysthymia) in the Leiden University Medical Center (LUMC) Department of Psychiatry or the Rivierduinen specialized mental healthcare centers. Baseline assessment was part of the usual ROM procedure. On average, 80% of the referred patients with a tentative diagnosis of Mood-, Anxiety- and/or Somatoform (MAS) disorder were assessed with ROM in the study period (Van Noorden et al., 2010).

2.2. Procedures and instruments

Procedures of the web-based ROM program of the LUMC Department of Psychiatry are described in detail elsewhere (Carlier et al., 2012; Van Fenema et al., 2012; Van Noorden et al., 2010). For our study, we used baseline ROM assessments that comprised a standardized diagnostic interview (Dutch version of the Mini-International Neuropsychiatric Interview Plus, version 5.00-R: MINI-Plus) (Sheehan et al., 1998; Van Vliet et al., 2000; Van Vliet and De Beurs, 2007), collection of sociodemographic and socioeconomic data, observer-rated scales, and self-report instruments. The assessments were performed by trained research nurses in the outpatient clinics. The BDI-II was completed by 455 reference subjects and 4019 patients. Halfway through the study, the BDI-II was replaced by the IDS-SR. The IDS-SR was completed by another group of 769 reference subjects and 474 patients. The MADRS was completed by the majority of both groups: 1291 reference subjects and 4627 patients. The BDI-II measures the severity of self-reported depression in adolescents and adults according to Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (American Psychiatric Association, 1994) criteria for diagnosing MDD (Beck et al., 1996a; Beck et al., 2002 Dutch version). It was a substantially revised and upgraded version of the original Beck Depression Inventory (BDI) (Beck and Steer, 1987). The BDI-II is scored by summing the scores for each of the 21 items. Each item is rated on a 4-point scale ranging from 0 to 3, with higher scores indicating higher levels of depression. The total score can range from 0 to 63, where scores from 0 to 13 denote "Minimal" depression; total scores from 14 to 19 denote "Mild" depression; total scores from 20 to 28 denote "Moderate" depression; total scores from 29 to 63 denote "Severe" depression. In the international literature, next to the total score, two (Steer et al., 1999; Storch et al., 2004; Whisman et al., 2000), three (Beck

et al., 2002) or no (Segal et al., 2008b) subscales have been identified; the Cognitive (from now on noted as Cognitive[1]) and Somatic-Affective (from now on Somatic-Affective[1]) subscales are most commonly used (Kumar et al., 2002; Steer et al., 1999). They consist of 8 and 13 items and sum scores range from 0 to 24 and from 0 to 39, respectively (Beck et al., 1996a). Alternatively, the Cognitive (from now on noted as Cognitive[2]; 7 items; sum scores range from 0 to 21), Affective (Affective[2]; 5 items; sum scores range from 0 to 15) and Somatic (Somatic[2]; 9 items; sum scores range from 0 to 27) subscales can be used (Beck et al., 2002).

The IDS-SR self-report instrument (Nolen and Dingemans 2004; Rush et al., 1996 Dutch translation) is designed to measure overall depressive symptom severity. The IDS-SR consists of 30 items: 23 of which cover the diagnostic DSM-IV criteria for MDD (including its atypical and melancholic subtypes). Seven items are not diagnostic criteria but commonly associated symptoms for 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 (sum)score ranges from 0 to 84, because only decreased or increased weight and appetite are scored (i.e. 28 of 30 items are scored). Scores from 0 to 13 denote "Normal/no" depression; total scores from 14 to 21 denote "Possible/mild" depression; total scores from 22 to 30 denote "Moderate" depression; total scores from 31 to 38 denote "Severe" depression; and scores from 39 onward denote "Very severe" depression (Rush et al., 2000). Frequently used subscales are the 10-item Atypical Characteristics subscale (Novick et al., 2005) and 11-item Melancholic Characteristics subscale (van Reedt Dortland et al., 2010).

The MADRS (Montgomery and Asberg, 1979) is a clinician-rated instrument assessing severity and range of depressive symptoms. The 10 items were designed to be particularly sensitive to treatment effects and overall change. They are relevant to the illness and occur in the majority of cases although they do not cover all 9 DSM-IV MDD criteria. The items emphasize psychological symptoms (Heo et al., 2007). 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 that ranges from 0 to 60. Scores from 0 to 8 denote "Normal/no" depression; total scores from 9 to 18 denote "Possible/mild" depression; total scores from 19 to 26 denote "Moderate" depression; total scores from 27 to 34 denote "Severe" depression; and scores from 35 onward denote "Very severe" depression (Müller et al., 2000).

The MINI-Plus (Sheehan et al., 1998; Van Vliet et al., 2000; Van Vliet and De Beurs, 2007) was used to establish the presence of Axis I diagnosis according to the DSM-IV. This standardized diagnostic interview comprises 23 modules for mood, anxiety, psychotic, somatoform and eating disorders.

2.3. Statistical analyses

Analyses were performed in the two groups separately, while ROC and internal consistency analyses were done using data from both groups combined. In both groups, subjects who had more than 1 missing value per subscale were excluded. 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 "clinically depressed" were obtained by ROC analyses. Sensitivity and specificity were chosen to be equal, taking into account the trade-off between the two (Fan et al., 2006). The discriminatory power of the instrument (sub)scales was assessed with the associated areas under the ROC curve (AUCs), where AUC's over 0.75 were considered clinical useful with 0.85 showing moderate discriminatory power and 0.95 very high power (Barnabei et al., 2007). The 5th, 25th, 50th, 75th and 95th percentiles were calculated. Use of the central 95% of the distribution in a reference-group is commonly used in case of non-Gaussian distributions (Geffre et al., 2009; Solberg and Grasbeck, 1989), the remaining 5% categorized as being "abnormal" (Campbell and Robinson, 1993). We chose the top 5% of the reference-group (95th percentiles, P_{95}) as being "abnormal" because the lowest 2.5% (functioning "abnormally" good) is not identifiable in general population samples: 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 (5th 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 all participants combined, as well as for 4 strata: young women (aged 18–40 yr), older women (aged 41–65 yr), young men (aged 18-40 yr), and older men (aged 41–65 yr). The internal consistency of the instruments was tested using Cronbach's alpha for the total and any subscales (with >0.70 indicating adequate internal consistency). For all analyses, SPSS version 17.0 was used.

3. Results

3.1. Sociodemographic and psychopathological characteristics

The sociodemographic and psychopathological characteristics of the ROM reference-group and patient-group are shown in Table 1.

Mean age (40.3 yr (SD=12.6) versus 39.3 (SD=12.3), $p=0.14$) and gender distribution (62.8% versus 61.0% females, $p=0.25$) were comparable between the ROM reference and patient-groups. Participants from the ROM reference-group were more often married than those from the ROM patient-group (68.7% versus 43.5%, $p<0.001$), less often living alone (15.5% versus 22.0%, $p<0.001$). The ROM reference-group showed higher levels of education than the ROM patient-group (77.2% versus 49.5% higher education, $p<0.001$). Level of education was defined as lower for general basic education only, or lower vocational education, and as higher for 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 self or both parents not being born in the Netherlands). Of the ROM reference-group 1.3% had at least one depressive disorder, compared to 100% of the subjects from the ROM patient-group (inclusion criterion). Of the ROM reference-group 8.1% met criteria of another psychiatric disorder on the MINI-Plus. There was a considerable amount of psychopathological co-morbidity with MDD in the ROM patient-group (53.5%).

3.2. Internal consistency

The internal consistencies of the instruments' total- and subscales for all subjects combined are shown in Table 3. The total scales of all three instruments showed excellent internal consistencies. Except for the IDS-SR Atypical Characteristics subscale (with a questionable alpha) none of the subscales had Cronbach's alphas below the critical cut-off of 0.70, indicating adequate internal consistency.

3.3. Reference values

3.3.1. Percentiles, means and SDs

Percentiles and other reference values of the BDI-II, IDS-SR and MADRS (sub-)scales in the ROM reference and patient-groups are shown in Table 2. In the reference-group, the distribution of each

Table 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)
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 (%)[a]				($p<0.001$)
Married/cohabitating	890	(68.7)	2027	(43.8)
Divorced/separated/widow	78	(6.0)	689	(14.9)
Single	327	(25.3)	1382	(39.9)
Housing situation (%)[a]				($p<0.001$)
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 (%)[a,d]				($p<0.001$)
Lower	295	(22.8)	1843	(39.8)
Higher	1000	(77.2)	2253	(48.7)
Employment status (%)[a]				($p<0.001$)
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 (%)[a–b]				($p<0.001$)
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[c]	
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[c]	

ROM: Routine Outcome Monitoring.

[a] No data from 70 (1.5%)–570 (12.3%) patients.
[b] No data from 11 reference subjects.
[c] Selection criterion.
[d] Lower education: general basic education only, or lower vocational education higher education: middle or higher vocational education, college or university.

total and subscale score was positively skewed. Apparent health was also demonstrated by the substantial percentage of participants having the lowest possible scores.

Two sets of BDI-II subscales were analyzed, the internationally most commonly used Cognitive[1] and Somatic-Affective[1] BDI-II subscales (Steer et al., 1999), and the Cognitive[2], Affective[2] and Somatic[2] subscales (Beck et al., 2002). For the ROM reference group, Table 2 shows the following P_{95} cut-off values for the BDI-II total- and subscales: for the total score 13; for the Cognitive[1] subscale 5; for the Somatic-Affective[1] subscale 10; for the Cognitive[2] subscale 4; for the Affective[2] subscale 3; and for the Somatic[2] subscale 8. The mean BDI-II total score was 3.7 (SD=4.7) for the ROM reference group, indicating that the overall 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 P_5 value for the ROM patient group was 14. For the IDS-SR total- and subscales the P_{95} cut-off values for the ROM reference group were for the total score 20; for the Atypical Characteristics subscale 11; and for the Melancholical Characteristics subscale 10. The mean IDS-SR total score was 6.7 (SD=6.9) for the ROM reference group, compared to 38.1 (SD=12.1) in the MDD patient group. The P_5 value for the ROM patient group was 18. For the MADRS, the P_{95} reference score was 11. The mean MADRS score was 2.8 (SD=3.8) for the ROM reference group, contrasting with 23.4 (SD=7.8) for the MDD patient group. The P_5 value for the ROM patient group was 11.

Supplementary Tables 1 through 3 (see Appendix, available with the full text of this article), show the reference values according to gender- and age-strata. For the self-report BDI-II and IDS-SR higher age was associated with a higher severity of MDD symptoms in women and men from the ROM reference group, which was not the case for the observer-rated MADRS. Women from the ROM reference group scored on average higher on all three MDD severity scales as compared to their male counterparts, illustrated by the higher P_{95} cut-off values. For the BDI-II the P_{95} values for women and men were, respectively, 15.0 and 12.0 for the total score; 4.5 and 4.5 for the Cognitive subscale[1];10.5 and 8.0 for the Somatic-Affective subscale[1]; 5.0 and 2.5 for the Cognitive[2] subscale; 3.0 and 2.5 for the Affective[2] subscale; and 9.0 and 7.0 for the Somatic[2] subscale. For the IDS-SR P_{95} cut-off values for women and men were, respectively, 23.0 and 18.0 for the total score; 9.5 and 9.0 for the Atypical Characteristics subscale; and 12.0 and 10.5 for the Melancholical Characteristics subscale. For the MADRS, again, the P_{95} cut-off value was higher for women than for men: 12.5 compared to 9.0.

3.3.2. Receiver operating characteristic (ROC) curves

Cut-off points, defined by equal sensitivity and specificity, were calculated with ROC analyses (see Table 3).

The cut-off point of the BDI-II total score discriminating health from disease was 13.5 with a sensitivity and specificity of 96%. Therefore, 4% of the reference participants who had a total score of 13.5 or higher would be incorrectly classified as depressed, and also 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. Although all subscales showed excellent discriminative power, the best performing was the Somatic-Affective subscale[1]. See Fig. 1 for the discriminative powers of the three MDD scales.

The cut-off point of the IDS-SR total score discriminating health versus disease was 18.5 with a sensitivity and specificity of 94%. The AUC value was 0.98 for the total score, with the best performing subscale being Melancholical Characteristics. The cut-off point of the MADRS total score discriminating health versus disease was 10.5 with a sensitivity and specificity of 95%. The AUC value was 0.99.

4. Discussion and conclusion

We reported reference values for the generic instruments BDI-II, IDS-SR and MADRS in two large samples from a "healthy" and "psychiatrically ill" population. 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 sex-specific pattern in reference values was observed, with women showing higher values than men.

The mean BDI-II scores for the ROM reference group were lower than the mean BDI-II scores reported by Segal et al. (2008a) (mean=3.8, SD=4.7 versus mean=8.6, SD=7.7), suggesting that

Table 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					
	P5	P25	P50 (median)	P75	P95	Mean±SD	P5	P25	P50 (median)	P75	P95	Mean ± SD
Beck Depression Inventory-II (BDI-II)*			(*n*=**455**)						(*n*=**4019**)			
Total score	0	0	2	5	13	3.74±4.74	14	24	30	38	49	30.80±10.52
Cognitive[1]	0	0	0	1	5	0.98±1.67	2	7	10	14	19	10.44±5.19
Somatic-Affective[1]	0	0	2	4	10	2.76±3.53	9	16	20	25	31	20.36±6.64
Cognitive[2]	0	0	0	1	4	0.95±1.72	2	6	10	13	18	9.78±4.79
Affective[2]	0	0	0	1	3	0.52±1.03	3	5	7	9	12	7.09±3.06
Somatic[2]	0	0	1	4	8	2.27±2.74	6	11	14	17	21	13.94±4.64
Inventory of Depressive Symptomatology - Self-Report*			(*n*=**769**)						(*n*=**474**)			
Total score	0	2	5	9	20	6.74±6.88	18	30	38	46	58	38.05±12.07
			(*n*=**196**)						(*n*=**208**)			
Atypical characteristics[3]	1	3	4	6	11	4.71±3.01	8	11	14	17	21	14.12±4.13
			(*n*=**165**)						(*n*=**115**)			
Melancholic characteristics°	1	3	4	6	10	4.62±2.31	6	8	10	12	17	10.24±3.5
Montgomery Åsberg Depression Rating Scale**			(*n*=**1291**)						(*n*=**4627**)			
Total score	0	0	2	4	11	2.79±3.84	11	18	23	28	36	23.44±7.75

ROM: Routine outcome monitoring.

* 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.

[2] Beck et al., 2002; Cognitive subscale comprises items 3, 5, 6, 7, 8, 13 and 14; Affective subscale comprises items 1, 2, 4, 9 and 12; Somatic subscale comprises items 10, 11, 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.

° van Reedt Dortland et al;., 2010: Melancholic characteristics subscale comprise items 3, 6, 9, 10, 11, 12, 17, 20, 25, 27 and 28.

Table 3
Internal consistency and cut-off scores in the ROM reference (*n*=1295) and patient (*n*=4627) groups for Routine Outcome Monitoring depressive disorder instruments.

	Cronbach's alpha	Number of items	N	ROC cut-off	AUC	Sensitivity/specificity
Beck Depression Inventory-II (**BDI-II**),			**4474**[a]			
Total score	0.93	21		13.5	0.99	0.96
Cognitive[c]	0.87	8		3.5	0.97	0.91/0.093
Somatic-Affective[c]	0.89	13		9.5	0.99	0.95
Cognitive[d]	0.86	7		3.5	0.97	0.91/0.93
Affective[d]	0.83	5		2.5	0.98	0.95
Somatic[d]	0.83	9		7.5	0.98	0.92/0.94
Inventory of Depressive Symptomatology-Self-Report (**IDS-SR**)			**1243**[a]			
Total score	0.94	32		18.5	0.98	0.94
Atypical characteristics[e]	0.68	10		7.5	0.92	0.84
Melancholic characteristics[f]	0.78	11		9.5	0.97	0.94/0.91
Montgomery Åsberg Depression Rating Scale (**MADRS**)			**5918**[b]			
Total score	0.90	10		10.5	0.99	0.95

ROM: Routine Outcome Monitoring.

[a] BDI-II samples and IDS-SR samples were non-overlapping.

[b] MADRS samples were partly overlapping with BDI-II or IDS-SR samples.

[c] 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.

[d] Beck et al. (2002); Cognitive subscale comprises items 3, 5, 6, 7, 8, 13 and 14; Affective subscale comprises items 1, 2, 4, 9 and 12; Somatic subscale comprises items 10, 11, 15, 16, 17, 18, 19, 20 and 21.

[e] Novick et al. (2005): Atypical characteristics subscale comprises items 4, 8, 9, 10, 11, 15, 18, 24, 33 and 34.

[f] van Reedt Dortland et al. (2010): Melancholic characteristics subscale comprise items 3, 6, 9, 10, 11, 12, 17, 20, 25, 27 and 28.

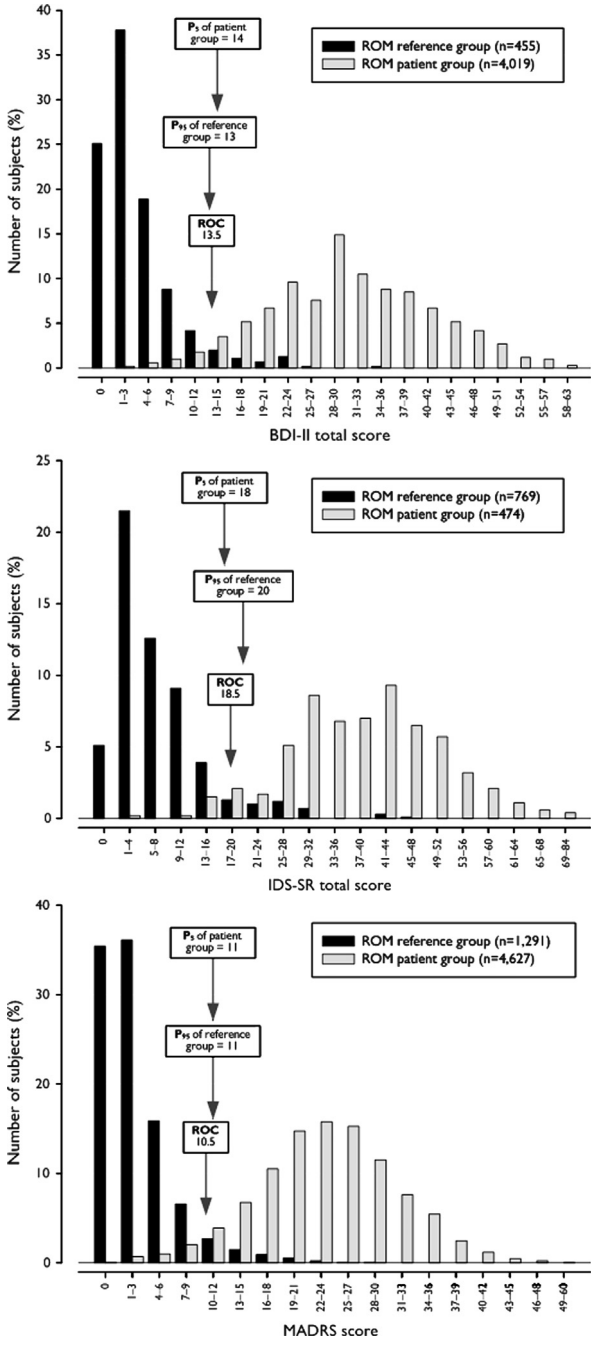


Fig. 1. Distribution of the scores of Beck Depression Inventory-II (BDI-II) total scale, the Inventory of Depressive Symptoms (self-report) (IDS-SR) total scale and the Montgomery Åsberg Depression Rating Scale (MADRS). Three types of cut-off points are depicted: the 75th percentile score (P_{75}), the 95th percentile score (P_{95}) and the Receiver Operating Characteristics (ROC) cut-off point defined by equal sensitivity and specificity.

our reference group was relatively healthy. The mean IDS-SR score from our ROM reference group, however, was slightly higher than the mean IDS-SR scores reported by Rush et al. (1986) (mean=6.7, SD=6.9 versus mean=2.1, SD=2.2). It should be noted however that their skewed distributions preclude an accurate comparison of these parametric estimates. The ROC cut-off value for the ROM reference group was similar to the one reported later by Rush et al. (1996) (18.5 versus 18.0). The mean MADRS score for the ROM reference group was slightly lower than the weighted mean MADRS score reported by Zimmerman et al. (2004a) (mean=2.8, SD=3.8 versus mean=4.0, SD=5.8). 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 consistency of the BDI-II, IDS-SR and MADRS are in accordance with previous studies (De Beurs et al., 2009; De Beurs and Zitman, 2006; DeRogatis and Melisaratos 1983; Gandek and Ware, Jr., 1998; Van Kampen et al., 2008; Wardenaar et al., 2010).

The results of our study have the following clinical implications. The excellent performance of the instruments indicates that our reference values are suitable for different purposes: (1) diagnostics; (2) decisions about treatment termination and referral back to primary care; (3) identification of people who may benefit from referral by primary care to specialized mental health care. Regarding diagnostics, these cutoff values can aid in the screening for MDD, although clinical judgment and validated diagnostic tools (e.g., MINI (Sheehan et al., 1998; Van Vliet et al., 2000), Composite International Diagnostic Interview (CIDI) (World Health Organization, 1990), the Structured Clinical Interview for Diagnostic and Statistical Manual (DSM)-III-R (SCID) (Spitzer et al., 1988) remain the "gold standard". Moreover, cut-off scores may be used to classify depression severity (Beck et al., 1996a). When making decisions about treatment termination or referral to primary care, specificity has to be high. The 95th percentile scores of the ROM reference group results 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 ROC-based cut-offs or 5th percentile scores from the ROM patient group for that purpose.

It is of note that the 95th percentile of the reference group, the 5th percentile of the patient group and the ROC cut-off values largely overlapped. They were also largely consistent with the cut-off values used for the BDI-II, IDS-SR and MADRS to distinguish individuals without depressive symptoms from those with mild symptoms (being 14, 18 and 12, respectively) (Beck et al., 1996b; Montgomery et al., 1993; Rush et al., 1996). Furthermore, women from the ROM reference group scored on average higher on all three MDD severity scales than males. For the BDI-II P_{95} values were 15 and 12, for the IDS-SR 23 and 18, and for the MADRS 12.5 and 9, in women and men, respectively. It may be too early to recommend gender specific reference values as more research is needed in reference populations. Nevertheless it was striking that reference values from non-depressed populations showed clinically important differences. Most previous studies did not stratify for gender (Rush et al., 1986; Segal et al., 2008a; Zimmerman et al., 2004a) but the one who did (Rush et al., 1996) found no gender differences. For the self-report BDI-II and IDS-SR higher age was associated with higher P_{95} reference scores in women and men from the ROM reference group, which 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 (ascertained by training and supervision). Second, the ROM reference group was large, clearly defined and resembled the patient group in all relevant respects (age, gender, level of urbanization) other than those under investigation (level of psychopathology). Furthermore, it likely represents the general population quite well, because of the high GP registration rate in the Netherlands. Finally, stratification of the ROM reference group into more homogeneous gender- and age-subgroups may have reduced variation among subgroups leading to gender-specific reference values to be used in clinical practice.

Limitations of the present study include the non-response in the ROM reference group. This was substantial at 63.2%, implying potential selection bias. Participation 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 and of ROM reference subjects that completed the BDI-II varied among the three scales. Additionally, the generalizability of this study is limited by the nature of our ROM reference group that was Dutch-speaking, mainly Caucasian and aged between 18 and 65. Reference values may not automatically be applicable to other ethnic or cultural groups, children, and the elderly. Finally, it is important to recognize that population-based reference values should not be applied rigidly. The choice of cut-off scores remains arbitrary and dependent one's goal.

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

Role of funding source
Inapplicable.

Conflict of interest
None.

Acknowledgments
This study was supported by Rivierduinen, a regional mental health care provider in the Netherlands. We are indebted to the interviewers for their help in collecting the data. In addition, the essential contribution made by the participants of this study and their general practitioners is very gratefully acknowledged.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jad.2013.02.009.

References

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

Barnabei, L., Marazia, S., De Caterina, R., 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. Journal of Cardiovascular Medicine (Hagerstown) 8 (11), 873–881, Available from: PM:17906471.

Beck, A.T., Steer, R., Brown, G.K., 1996a. Manual for the Beck Depression Inventory-II. Psychological Corporation, San Antonio,TX.

Beck, A.T., Steer, R., Brown, G.K., 2002. Beck Depression Inventory-II-NL. Handleiding. De Nederlandse versie van de Beck Depression Inventory. In: vert. en bew, A.J.W.van der Does (Ed.), 2nd edition Swets Test Publishers, Lisse, NL.

Beck, A.T., Steer, R.A., 1987. Manual for the Revised Beck Depression Inventory. Psychological Corporation, San Antonio, TX.

Beck, A.T., Steer, R.A., Ball, R., Ranieri, W., 1996b. Comparison of Beck Depression Inventories-IA and -II in psychiatric outpatients. Journal of Personality Assessment 67 (3), 588–597, Available from: PM:8991972.

Biggs, M.M., Shores-Wilson, K., Rush, A.J., Carmody, T.J., Trivedi, M.H., Crismon, M.L., Toprac, M.G., Mason, M., 2000. A comparison of alternative assessments of depressive symptom severity: a pilot study. Psychiatry Research 96 (3), 269–-279, Available from: PM:11084222.

Bondolfi, G., Jermann, F., Rouget, B.W., Gex-Fabry, M., McQuillan, A., Dupont-Willemin, A., Aubry, J.M., Nguyen, C., 2010. Self- and clinician-rated Montgomery-Asberg Depression Rating Scale: evaluation in clinical practice. Journal of Affective Disorders 121 (3), 268–272, Available from: PM:19660815.

Campbell, W.W., Robinson, L.R., 1993. Deriving reference values in electrodiagnostic medicine. Muscle and Nerve 16 (4), 424–428, Available from: PM:8455658.

Carlier, I.V., Meuldijk, D., Van Vliet, I.M., Van Fenema, E., Van der Wee, N.J., Zitman, F.G., 2012. Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. Journal of Evaluation in Clinical Practice 18 (1), 104–110, Available from: PM:20846319.

Corruble, E., Legrand, J.M., Zvenigorowski, H., Duret, C., Guelfi, J.D., 1999. Concordance between self-report and clinician's assessment of depression. Journal of Psychiatric Research 33 (5), 457–465, Available from: PM:10504014.

De Beurs, E., Den Hollander-Gijsman, M.E., Van Rood, Y.R., Van der Wee, N.J., Giltay, E.J., Van Noorden, M.S., Van der Lem, R., Van Fenema, E., Zitman, F.G., 2011. Routine outcome monitoring in the Netherlands: practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. Clinical Psychology and Psychotherapy 18, 1–12, Available from: PM:20238371.

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. Journal of Personality Disorders 23 (3), 308–326, Available from: PM:19538084.

De Beurs, E., Zitman, F.G., 2006. De Brief Symptom Inventory (BSI): De betrouwbaarheid en validiteit van een handzaam alternatief voor de SCL-90., 61 ed.

DeRogatis, L.R., Melisaratos, N., 1983. The Brief Symptom Inventory: an introductory report. Psychological Medicine 13 (3), 595–605, Available from: PM:6622612.

Engedal, K., Kvaal, K., Korsnes, M., Barca, M.L., Borza, T., Selbaek, G., Aakhus, E., 2012. The validity of the Montgomery-Aasberg depression rating scale as a screening tool for depression in later life. Journal of Affective Disorders 141 (2-3), 227–232.

Fan, J., Upadhye, S., Worster, A., 2006. Understanding receiver operating characteristic (ROC) curves. Canadian Journal of Emergency Medical Care 8 (1), 19–20, Available from: PM:17175625.

Gandek, B., Ware Jr., J.E., 1998. Methods for validating and norming translations of health status questionnaires: the IQOLA Project approach. International Quality of Life Assessment. Journal of Clinical Epidemiology 51 (11), 953–959, Available from: PM:9817112.

Geffre, A., Friedrichs, K., Harr, K., Concordet, D., Trumel, C., Braun, J.P., 2009. Reference values: a review. Veterinary Clinical Pathology 38 (3), 288–298, Available from: PM:19737162.

Hawley, C.J., Gale, T.M., Sivakumaran, T., 2002. Defining remission by cut off score on the MADRS: selecting the optimal value. Journal of Affective Disorders 72 (2), 177–184, Available from: PM:12200208.

Heo, M., Murphy, C.F., Meyers, B.S., 2007. Relationship between the Hamilton Depression Rating Scale and the Montgomery-Asberg Depression Rating Scale in depressed elderly: a meta-analysis. American Journal of Geriatric Psychiatry 15 (10), 899–905, Available from: PM:17911366.

Horn, P.S., Feng, L., Li, Y., Pesce, A.J., 2001. Effect of outliers and nonhealthy individuals on reference interval estimation. Clinical Chemistry 47 (12), 2137–-2145, Available from: PM:11719478.

Jacobson, N.S., Truax, P., 1991. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology 59 (1), 12–19, Available from: PM:2002127.

Katayev, A., Balciza, C., Seccombe, D.W., 2010. Establishing reference intervals for clinical laboratory test results: is there a better way? American Journal of Clinical Pathology 133 (2), 180–186, Available from: PM:20093226.

Kazdin, A.E., 2008. Evidence-based treatment and practice: new opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. American Psychologist 63 (3), 146–159, Available from: PM:18377105.

Kendall, P.C., Marrs-Garcia, A., Nath, S.R., Sheldrick, R.C., 1999. Normative comparisons for the evaluation of clinical significance. Journal of Consulting and Clinical Psychology 67 (3), 285–299, Available from: PM:10369049.

Khan, A., Khan, S.R., Shankles, E.B., Polissar, N.L., 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. International Clinical Psychopharmacology 17 (6), 281–285, Available from: PM:12409681.

Kumar, G., Steer, R.A., Teitelman, K.B., 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, Available from: PM:12066831.

Montgomery, S.A., Asberg, M., 1979. A new depression scale designed to be sensitive to change. British Journal of Psychiatry 134, 382–389, Available from: PM:444788.

Montgomery, S.A., Rasmussen, J.G., Tanghoj, P., 1993. A 24-week study of 20 mg citalopram, 40 mg citalopram, and placebo in the prevention of relapse of major depression. International Clinical Psychopharmacology 8 (3), 181–188, available from: PM:8263316.

Müller, M.J., Himmerich, H., Kienzle, B., Szegedi, A., 2003. Differentiating moderate and severe depression using the Montgomery-Asberg depression rating scale (MADRS). Journal of Affective Disorders 77 (3), 255–260, Available from: PM:14612225.

Müller, M.J., Szegedi, A., Wetzel, H., Benkert, O., 2000. Moderate and severe depression. Gradations for the Montgomery-Asberg depression rating scale. Journal of Affective Disorders 60 (2), 137–140, Available from: PM:10967373.

Nolen, W.N., Dingemans, P.M.A.J., 2004. Meetinstrumenten bij stemmingsstoornissen (Instruments for measuring mood dissorders). Tijdschrift voor Psychiatrie 46, 681–686.

Novick, J.S., Stewart, J.W., Wisniewski, S.R., Cook, I.A., Manev, R., Nierenberg, A.A., Rosenbaum, J.F., Shores-Wilson, K., Balasubramani, G.K., Biggs, M.M., Zisook, S., Rush, A.J., 2005. Clinical and demographic features of atypical depression in

outpatients with major depressive disorder: preliminary findings from STAR*D. Journal of Clinical Psychiatry 66 (8), 1002–1011, Available from: PM:16086615.
Poortvliet, M.C., Lamkadden, M., Deville, W., 2005, Niet op naam ingeschreven (NONI) bij de huisarts. Inventarisatie en gevolgen voor de ziekenfondsverzekerden., NIVEL.
Reed, A.H., Henry, R.J., Mason, W.B., 1971. Influence of statistical method used on the resulting estimate of normal range. Clinical Chemistry 17 (4), 275–284, available from: PM:5552364.
Rush, A.J., Giles, D.E., Schlesser, M.A., Fulton, C.L., Weissenburger, J., Burns, C., 1986. The Inventory for Depressive Symptomatology (IDS): preliminary findings. Psychiatry Research 18 (1), 65–87, Available from: PM:3737788.
Rush, A.J., Gullion, C.M., Basco, M.R., Jarrett, R.B., Trivedi, M.H., 1996. The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychological Medicine 26 (3), 477–486, Available from: PM:8733206.
Rush, A.J., Pincus, H.A., First, M.B., 2000. Handbook of Psychiatric Measures. American Psychiatric Association, Washington DC.
Sagen, U., Vik, T.G., 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. Journal of Psychosomatic Research 67 (4), 325–332, Available from: PM:19773025.
Sasse, E.A., Doumas, B.T., Miller, W.G., D'Orazio, P., Eckfeldt, J.H., Evans, S.A., Graham, G.A., Myers, G.L., Parsons, P.J., Stanton, N.V., 2000. How to define and determine reference intervals in the clinical laboratory; approved guideline-Second edition. NCCLS document C28-A2, 20 ed. Wayne, PA, NCCLS.
Schulte-van Maaren, Y.W.M., Carlier, I.V., Giltay, E.J., Van Noorden, M.S., De Waal, M. W., Van der Wee, N.J., Zitman, F.G., 2013. Reference values for mental health assessment instruments: objectives and methods of the Leiden Routine Outcome Monitoring Study. Journal of Evaluation in Clinical Practice 19 (2), 342–350.
Schulte-van Maaren, Y.W.M., Carlier, I.V.E., Zitman, F.G., Van Hemert, A.M., De Waal, M.W.M., Van Noorden, M.S., Giltay, E.J., 2012. Reference values for generic instruments used in Routine Outcome Monitoring: the Leiden Routine Outcome Monitoring Study. BMC Psychiatry, 12:203 (21 November 2012) http://www.biomedcentral.com/1471-244X/12/203.
Segal, D.L., Coolidge, F.L., Cahill, B.S., O'Riley, A.A., 2008a. Psychometric properties of the Beck Depression Inventory II (BDI-II) among community-dwelling older adults. Behavior Modification 32 (1), 3–20, Available from: PM:18096969.
Segal, D.L., Coolidge, F.L., Cahill, B.S., O'Riley, A.A., 2008b. Psychometric properties of the Beck Depression Inventory II (BDI-II) among community-dwelling older adults. Behavior Modification 32 (1), 3–20, Available from: PM:18096969.
Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 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. Journal of Clinical Psychiatry 59 (Suppl. 20), S22–S33, Available from: PM:9881538.
Solberg, H.E., 2008. Establishment and use of reference values. In: Burtis, C.A., Ashwood, E.R., Bruns, D.E. (Eds.), Fundamentals of clinical chemistry, 6 ed. Saunders Elsevier, St. Louis, Missouri, pp. 229–238.
Solberg, H.E., Grasbeck, R., 1989. Reference values. Advances in Clinical Chemistry 27, 1–79, Available from: PM:2667291.
Spitzer, R.L., Williams, J.B.W., Gibbon, M., First, M.B., 1988. Structured Clinical Interview for DSM-III-R—Patient Version (SCID-P, 4/I/88). Biometrics Research Department, New York State Psychiatric Institute, New York.
Steer, R.A., Ball, R., Ranieri, W.F., Beck, A.T., 1999. Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. Journal of Clinical Psychology 55 (1), 117–128, Available from: PM:10100838.
Steer, R.A., Beck, A.T., Brown, G., Berchick, R.J., 1987. Self-reported depressive symptoms that differentiate recurrent-episode major depression from dysthymic disorders. Journal of Clinical Psychology 43 (2), 246–250.
Steer, R.A., Rissmiller, D.J., Beck, A.T., 2000. Use of the Beck Depression Inventory-II with depressed geriatric inpatients. Behaviour Research and Therapy 38 (3), 311–318, Available from: PM:10665163.
Storch, E.A., Roberti, J.W., Roth, D.A., 2004. Factor structure, concurrent validity, and internal consistency of the Beck Depression Inventory-Second Edition in a sample of college students. Depression and Anxiety 19 (3), 187–189, Available from: PM:15129421.
Trivedi, M.H., Rush, A.J., Ibrahim, H.M., Carmody, T.J., Biggs, M.M., Suppes, T., Crismon, M.L., Shores-Wilson, K., Toprac, M.G., Dennehy, E.B., Witte, B., Kashner, T.M., 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. Psychological Medicine 34 (1), 73–82, Available from: PM:14971628.
Van Fenema, E.M., Van der Wee, N.J., Giltay, E.J., Den Hollander-Gijsman, M.E., Zitman, F.G., 2012. Vitality predicts level of guideline-concordant care in routine treatment of mood, anxiety and somatoform disorders. Journal of Evaluation in Clinical Practice 18 (2), 441–448.
Van Kampen, D., De Beurs, E., Andrea, H., 2008. A short form of the Dimensional Assessment of Personality Pathology-Basic Questionnaire (DAPP-BQ): the DAPP-SF. Psychiatry Research 160 (1), 115–128, Available from: PM:18514918.
Van Noorden, M.S., Giltay, E.J., Den Hollander-Gijsman, M.E., Van der Wee, N.J., Van Veen, T., Zitman, F.G., 2010. Gender differences in clinical characteristics in a naturalistic sample of depressive outpatients: the Leiden Routine Outcome Monitoring Study. Journal of Affective Disorders 125 (1–3), 116–123, Available from: PM:20051291.
van Reedt Dortland, A.K., Giltay, E.J., van, V.T., van, P.J., Zitman, F.G., Penninx, B.W., 2010. Associations between serum lipids and major depressive disorder: results from the Netherlands Study of Depression and Anxiety (NESDA). Journal of Clinical Psychiatry 71 (6), 729–736, Available from: PM:20021996.
Van Vliet, I.M., Leroy, H., Van Megen, H.J.G.M., 2000. M.I.N.I. Internationaal Neuropsychiatrisch Interview.
Van Vliet, I.M., De Beurs, E., 2007. The MINI-International Neuropsychiatric Interview. A brief structured diagnostic psychiatric interview for DSM-IV en ICD-10 psychiatric disorders. **Tijdschrift** voor Psychiatrie 49 (6), 393–397, Available from: PM:17614093.
Wardenaar, K.J., Van Veen, T., Giltay, E.J., De Beurs, E., Penninx, B.W., Zitman, F.G., 2010. Development and validation of a 30-item short adaptation of the Mood and Anxiety Symptoms Questionnaire (MASQ). Psychiatry Research 179 (1), 101–106, Available from: PM:20472297.
Whisman, M.A., Perez, J.E., Ramel, W., 2000. Factor structure of the Beck Depression Inventory-Second Edition (BDI-II) in a student sample. Journal of Clinical Psychology 56 (4), 545–551, Available from: PM:10775046.
World Health Organization, 1990. The Composite International Diagnostic Interview (CIDI). Authorized Core Version 1.0. The Composite International Diagnostic Interview (CIDI), Geneva, Authorized Core Version 1.0..
Zimmerman, M., Chelminski, I., Posternak, M., 2004a. A review of studies of the Hamilton depression rating scale in healthy controls: implications for the definition of remission in treatment studies of depression. Journal of Nervous and Mental Disease 192 (9), 595–601, Available from: PM:15348975.
Zimmerman, M., Posternak, M.A., Chelminski, I., 2004b. 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. Journal of Psychiatric Research 38 (6), 577–582, Available from: PM:15458853.