

# The structure of the Montgomery–Åsberg depression rating scale over the course of treatment for depression

LENA C. QUILTY,<sup>1,2</sup> JENNIFER J. ROBINSON,<sup>1,2</sup> JEAN-PIERRE ROLLAND,<sup>3</sup> FILIP DE FRUYT,<sup>4</sup>  
FRÉDÉRIC ROUILLON<sup>5</sup> & R. MICHAEL BAGBY<sup>1,2</sup>

1 Centre for Addiction and Mental Health, Toronto, Canada

2 University of Toronto, Toronto, Canada

3 Université Paris Ouest Nanterre La Défense, France

4 Ghent University, Belgium

5 CMME Hôpital Sainte Anne - University Paris 5 - INSERM U669, France

---

## Key words

major depression, assessment,  
confirmatory factor analysis,  
gender invariance

## Correspondence

Lena C. Quilty, Centre for  
Addiction and Mental Health,  
250 College Street, Suite 648,  
Toronto, Ontario, M5T 1R8,  
Canada.  
Telephone (+1) 416-535-8501,  
ext. 34313 Fax (+1) 416-260-4125  
Email: lena.quilty@camh.ca

## Abstract

The Montgomery–Åsberg Depression Rating Scale (MADRS) is a widely used clinician-rated measure of depressive severity. Empirical support for the factor structure of the MADRS is mixed; further, the comparison of MADRS scores within and between patients requires the demonstration of consistent instrument properties. The objective of the current investigation was to evaluate MADRS factor structure as well as MADRS factorial invariance across time and gender. The MADRS was administered to 821 depressed outpatients participating in a large-scale effectiveness study of combined pharmacotherapy and psychotherapy for depression. Treatment outcome did not differ across treatment groups. Factor structure and invariance was evaluated via confirmatory factor analysis. A four-factor model consisting of Sadness, Negative Thoughts, Detachment and Neurovegetative symptoms demonstrated a good fit to the data. This four-factor structure was invariant across time and gender. A hierarchical model, in which these four factors served as indicators of a general depression factor, was also supported. A limitation of the current study is the lack of comprehensive characterization of patient clinical features; results need to be replicated in more severely depressed or treatment refractory patients. Overall, evidence supported the use of the MADRS total score as well as subscales focused on affective, cognitive, social and somatic aspects of depression in male and female outpatients.

Copyright © 2013 John Wiley & Sons, Ltd.

## Introduction

The Montgomery–Åsberg Depression Rating Scale (MADRS; Montgomery and Åsberg, 1979) was developed to provide an assay of depressive symptom severity superior to the widely used Hamilton Depression Rating Scale (Ham-D; Hamilton, 1960), for use in the investigation of treatment response to antidepressant medication. To this end, 10 items were selected from the Comprehensive Psychopathological Rating Scale (CPRS; Åsberg *et al.*, 1978) on the basis of their ability to detect depression change. The MADRS now follows the Ham-D as the most extensively used clinician-rated measure of depressive severity in clinical research (Santor *et al.*, 2006; Bagby *et al.*, 2004) and demonstrates psychometric properties equal or superior to other standardized measures of depressive severity (Carmody *et al.*, 2006b; Khan *et al.*, 2002; Mulder *et al.*, 2003; Uher *et al.*, 2008).

The use of the MADRS to monitor depression change over the course of treatment requires that the psychometric properties of this instrument are robust over time. Should the factor structure of MADRS items be variable over time – even in the context of acute symptom change – differences in scores over the course of time may in fact reflect changes in instrument properties rather than treatment response. Multiple investigations have provided evidence for a unifactorial structure of the MADRS (Carmody *et al.*, 2006a; Rocca *et al.*, 2002, Week 8; Uher *et al.*, 2008; Wolthaus *et al.*, 2000). Yet, a series of multifactorial models of MADRS items have been proposed (see Table 1). Moreover, factor structure stability over treatment has been called into question by evidence for a change in MADRS structure from multiple factors at treatment initiation to a single factor after one or more months of treatment (Galinowski and Lehert, 1995; Rocca *et al.*, 2002).

The use of the MADRS to assess depressive severity similarly in men and women further requires that the psychometric properties of this instrument are consistent across gender. Should the meaning of MADRS scores vary across gender, any differences in depressive severity across men and women may in fact reflect differential instrument functioning rather than true group differences (Millsap and Kwok, 2004). Evidence exists for the consistency of MADRS item associations across gender (e.g. Riedel *et al.*, 2010); however, a fulsome evaluation of the invariance of MADRS factor structure across gender has yet to be evaluated. Mixed evidence exists for such equivalence within other standardized measures of depressive severity across gender (Hunt-Shanks *et al.*, 2010; Rivera-Medina *et al.*, 2010).

Previous investigations of MADRS structure have differed in research design and analytic strategy. First, study samples have included patients with a range of diagnoses including major depressive disorder (Parker *et al.*, 2003),

dysthymic disorder (Rocca *et al.*, 2002), bipolar disorder (Benazzi, 2001), mild cognitive impairment (Gabryelewicz *et al.*, 2004) or a heterogeneous psychiatric sample (Craighead and Evans, 1996). Gabryelewicz *et al.* (2004) and Parker *et al.* (2003) evaluated elderly samples, whereas other studies included a greater age range (e.g. Benazzi, 2001). As noted by previous investigators, different populations may exhibit different factor structures of the MADRS (Suzuki *et al.*, 2005). Second, statistical analyses have frequently made use of suboptimal methods. Method of determining the number of factors to extract was often not reported, or utilized the unreliable Kaiser (eigenvalues > 1) or subjective Cattell (scree test) criteria. Unrotated or orthogonal rotations were common. The most sophisticated statistical analytic procedures have provided support for a one-factor model of the MADRS items (Carmody *et al.*, 2006a, 2006b; Uher *et al.*, 2008). These discrepancies in analytic procedures constrain comparisons across investigations.

The present investigation thus consisted of two objectives. First, we evaluated the factor structure of the MADRS in a large sample of depressed outpatients. Evidence has failed to converge upon a robust factor structure underlying MADRS items; yet, subscales based upon multifactorial models of the MADRS are increasingly used to characterize psychopathology and to predict treatment outcome (e.g. Higuchi *et al.*, 2008). Multifactorial models of MADRS items have generally been derived within psychiatric or medical samples at admission (Benazzi, 2001; Gabryelewicz *et al.*, 2004; Suzuki *et al.*, 2005) or within samples heterogeneous with respect to treatment status (Craighead and Evans, 1996; Parker *et al.*, 2003). Indeed, although psychotic symptoms have demonstrated a robust structure over the course of treatment (Harvey *et al.*, 2006), bipolar symptoms resolve from a multifactorial structure at pre-treatment to a unifactorial structure at post-treatment (Harvey *et al.*, 2008) – suggesting that measures of mood disorder symptoms may have a less stable symptom structure. Second, we evaluated the factor invariance of the MADRS items across time and gender. The MADRS is currently widely used in male and female inpatients and outpatients over the course of treatment; however, at the time of this study no evidence existed for the invariance of instrument structure. The present investigation thus provided a comprehensive evaluation of the factor structure of depressive symptoms as assessed by the MADRS over time and across male and female depressed outpatients.

## Methods

### Participants

Eight hundred and twenty-one outpatients [32% male, 68% female; age ranging from 18 to 66, mean (*M*) = 39.47 years,

**Table 1.** Previous multifactor models of MADRS items

| Study                             | Language  | N   | % Male | Participants             | Number         | Extraction | Rotation           | Factor (items)                                                                                                                                            |
|-----------------------------------|-----------|-----|--------|--------------------------|----------------|------------|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Two-factor models</i>          |           |     |        |                          |                |            |                    |                                                                                                                                                           |
| Rocca <i>et al.</i> (2002)        | Italian   | 100 | 33     | Dysthymic outpatients    | NR             | PCA        | NR                 | Global (1, 2, 3, 6, 7, 8, 9, 10)<br>Somatic (4, 5)                                                                                                        |
| <i>Three-factor models</i>        |           |     |        |                          |                |            |                    |                                                                                                                                                           |
| Galinowski and Leher (1995)       | French    | 137 | 64     | Depressed patients       | Kaiser         | PCA        | Unrotated          | Factor 1 (1, 2, 3, 6, 7, 8, 9, 10)<br>Factor 2 (4, 5)<br>Factor 3 (3, 6)                                                                                  |
| Andersson <i>et al.</i> (1999)    | Norwegian | 72  | 79     | Brain injury patients    | NR             | PCA        | Oblimin            | Depressed mood (1, 2, 3, 9)<br>Somatic symptoms (3, 4, 5)<br>Negative symptoms (6, 7, 8)<br>Factor 1 (1, 2, 8, 10)<br>Factor 2 (3, 4)                     |
| Benazzi (2001) (Unipolar)         | Italian   | 306 | 37     | Unipolar outpatients     | NR             | SMC        | Varimax            | Factor 3 (6, 7, 8, 9)<br>Factor 1 (1, 2)                                                                                                                  |
| Benazzi (2001) (Bipolar II)       | Italian   | 251 | 30     | Bipolar outpatients      | NR             | SMC        | Varimax            | Factor 2 (4, 5)<br>Factor 3 (6, 7, 8, 9, 10)<br>Dysphoric Apathy (1, 2, 6, 7, 8)<br>Psychic Anxiety (3, 9, 10)                                            |
| Parker <i>et al.</i> (2003)       | English   | 225 | NR     | Elderly patients         | Kaiser Cattell | EFA        | Promax             | Vegetative Symptoms (4, 5)<br>Anhedonia-Pessimism (1, 2, 8, 9)<br>Anxiety-Vegetative (3, 4, 5, 10)<br>Cognitive-Inhibition (6, 7)<br>Dysphoria (2, 9, 10) |
| Gabryelewicz <i>et al.</i> (2004) | Polish    | 102 | 29     | Neurological outpatients | Kaiser         | PCA        | Varimax            | Retardation (1, 6, 7, 8)<br>Vegetative (3, 4, 5)<br>Anhedonia (5, 7, 8, 10)<br>Sadness (1, 2, 9, 10)<br>Agitation (3, 6, 4)                               |
| Suzuki <i>et al.</i> (2005)       | Japanese  | 132 | 39     | Depressed participants   | Kaiser         | PCA        | Promax/<br>Varimax |                                                                                                                                                           |
| Famer <i>et al.</i> (2009)        | Norwegian | 163 | 50     | Elderly stroke patients  | Kaiser Scatter | PCA        | Oblimin            |                                                                                                                                                           |

(Continues)

Table 1388. (Continued)

| Study                           | Language | N   | % Male | Participants               | Number  | Extraction | Rotation | Factor (items)                                                                                                              |
|---------------------------------|----------|-----|--------|----------------------------|---------|------------|----------|-----------------------------------------------------------------------------------------------------------------------------|
| <i>Four-factor models</i>       |          |     |        |                            |         |            |          |                                                                                                                             |
| Craighead and Evans (1996)      | English  | 340 | 33     | Psychiatric inpatients     | Cattell | MLE        | Promax   | Cognitive-Pessimism (1, 2, 9, 10)<br>Affective (1, 7, 8)<br>Cognitive-Anxious (3, 6)<br>Vegetative (4, 5)<br>Sadness (1, 2) |
| Williamson <i>et al.</i> (2006) | English  | 788 | 37     | Bipolar depressed patients | NR      | PCA        | Equimax  | Negative Thoughts (9, 10)<br>Detachment (6, 7, 8)<br>Neurovegetative (3, 4, 5)                                              |

Note: NR = Not reported; PCA = principal components analysis; MLE = maximum likelihood estimation; SMC = squared-multiple correlation; Kaiser = Kaiser criterion (eigenvalue > 1 rule); Cattell = Cattell criteria (scree test).

standard deviation (SD) = 10.61] participated in a multi-centre effectiveness study of a combination of medication (tianeptine or fluoxetine) and psychotherapy (supportive, cognitive-behavioural, or psychodynamic therapy) for depression (see De Fruyt *et al.*, 2006). Eligibility was restricted to patients presenting with moderate or severe depression, as determined by a Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) diagnosis of major depressive disorder and a score of at least 20 on the MADRS (Montgomery and Åsberg, 1979; Snaith *et al.*, 1986). Patients were excluded if they exhibited psychosis or substance abuse; currently or recently received pharmacotherapy or psychotherapy; exhibited resistance or contraindications to study treatments; or were hospitalized for electroconvulsive therapy. The study was conducted in accordance with the Declaration of Helsinki and Good Clinical Practice Guidelines. All patients provided written informed consent after receiving a detailed description of the study and were free to withdraw at any time.

### Procedure

Outpatients consulting different psychiatric centres near Paris, France, were screened for inclusion and exclusion criteria and requested to participate in the longitudinal study. Screening was based on a non-standardized interview wherein each criterion of major depressive disorder was queried explicitly. Additional, detailed information regarding patient history and current presenting issues was collected to substantiate the determination of the presence or absence of each criterion. Details about the overall objectives, procedures, and samples can be obtained from the fifth author. Patients were randomized to receive either tianeptine (50 mg/day) or fluoxetine (20 mg/day). Within each of these cells, patients were non-randomly assigned to receive supportive, cognitive-behavioural, or psychodynamic psychotherapy.

Of the 821 patients entered in the present study, 821 completed the MADRS at pre-treatment, 817 at month one, 735 at month three, and 685 at month six. Patients who did versus did not complete the protocol differed on pre-treatment suicidality, as assessed by item 10 of the MADRS ( $t = 2.78$ ,  $p < 0.01$ ,  $d = 0.27$ ). Patients who did versus did not complete treatment did not differ on any remaining pre-treatment demographic or clinical characteristics (i.e. age, sex, MADRS items one through nine, MADRS total score). Further, patient attrition did not differ across medication or psychotherapy groups. No information was available regarding those not eligible to participate or regarding patient attrition (see Figure 1).

## Measurement

The MADRS is a clinician-rated measure of depression severity (Montgomery and Åsberg, 1979). The MADRS comprises the following 10 items: (1) apparent sadness; (2) reported sadness; (3) inner tension; (4) reduced sleep; (5) reduced appetite; (6) concentration difficulties; (7) lassitude; (8) inability to feel; (9) pessimistic thoughts; and (10) suicidal thoughts. These items are clinician-rated on a seven-point Likert scale and are summed to produce a total scale score ranging from 0 to 60, with higher scores reflecting greater depression severity. The MADRS was administered at pre-treatment, and one month ( $\pm 7$  days), three months ( $\pm 15$  days) and six months ( $\pm 15$  days) following treatment initiation (or post-treatment).

## Statistical analyses

To provide information regarding overall treatment efficacy, we performed a repeated measures analysis of variance (ANOVA), wherein the MADRS total scores served as the dependent variable, and medication condition and psychotherapy condition served as independent variables.

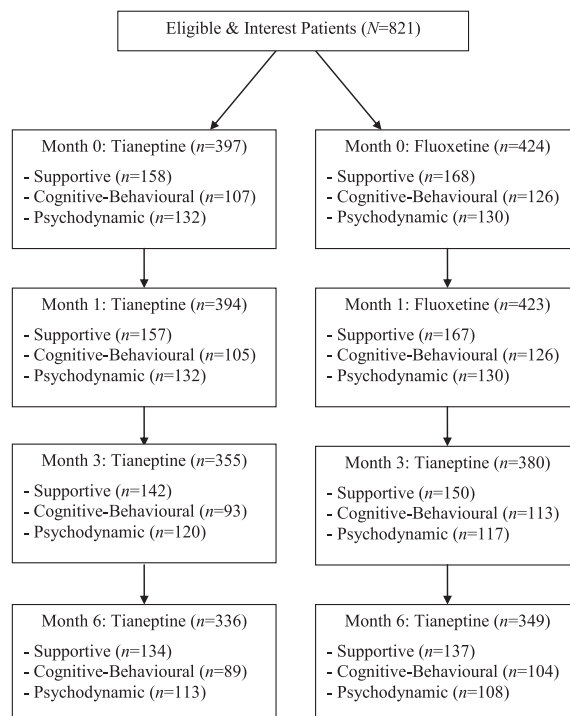


**Figure 1** Patient participation.

## Factor structure of the MADRS

Confirmatory factor analysis (CFA) was carried out on the MADRS data of the total sample prior to treatment. All previously proposed one, two, three, and four factor models of all MADRS items were evaluated, using pre-treatment MADRS scores. Models proposed in non-English journals as well as models omitting more than one MADRS item were not evaluated (Corruble *et al.*, 1999; Hammond, 1998; Serretti *et al.*, 1999). Items were assigned to factors as specified by original investigators; in the absence of these specifications, items loading  $> 0.40$  on a factor were modelled as indicators of that factor. In factors subsumed by only two items, item loadings were restricted to be equal. Factors were allowed to correlate unless otherwise specified in the original model. All error co-variances and item cross-loadings described in the original models were included in analyses.

CFA was conducted with Amos 16.0, using maximum likelihood method of estimation. Due to missing data, multivariate indices of normality could not be computed. Some forms of multivariate non-normality can be indicated by univariate non-normality, however, with acceptable levels indicated by univariate skew  $< 3$  and kurtosis  $< 8$  (Kline, 2005). We report the  $\chi^2$  statistic; due to the sensitivity of this index to sample size, goodness-of-fit was evaluated using the comparative fit index (CFI; good fit  $\geq 0.9$ ) and the root mean square error of approximation (RMSEA; good fit  $< 0.05$ ;  $0.05 \leq$  reasonable fit  $\leq 0.08$ ). The RMSEA significance test for close fit is similar to that of the  $\chi^2$  in that  $p < 0.05$  signifies that a hypothesis of close fit is rejected. All models with acceptable fit according to the CFI and RMSEA were subject to invariance testing.

## Temporal and gender invariance of the MADRS

Temporal invariance was evaluated via a series of CFAs, for each factor within the models demonstrating acceptable fit. Separate models were evaluated as convergence problems arose while modeling multifactor models across four assessment points, similar to past research (e.g. Conroy *et al.*, 2003). In each model, four latent factors associated with each assessment point were modelled, with later factors regressed onto earlier ones. Item uniquenesses were permitted to covary across waves. Gender invariance was evaluated via a series of “stacked” CFAs, in which models for both men and women are estimated simultaneously, with increasingly strict constraints of measurement invariance placed across groups in each model.

To determine whether the MADRS items decomposed into the same factor structure across time/gender (configural invariance), the model by Williamson *et al.*

(2006) (see later) was initially evaluated with no equality constraints on any parameters. To determine whether the strength of the association between items and factors were equivalent across time/gender (metric invariance), factor loadings were constrained to be equal across assessment points/gender groups. To determine whether item intervals and zero points were equivalent across time/gender (scalar or strong invariance), both factor loadings and item intercepts were constrained to be equal across assessment points/gender groups. To determine whether items measured depressive severity with the same degree of measurement error across time/gender (strict invariance), factor loadings, item intercepts, and variances of the error terms were constrained to be equal across assessment points/gender groups. Evidence for each form of invariance was demonstrated by a non-significant  $\Delta\chi^2$  and a  $\Delta CFI \leq 0.01$  with the inclusion of each set of constraints (Cheung and Rensvold, 2002).

## Results

There was a significant effect for time,  $F(3, 1995) = 2299.33$ ,  $p < 0.01$ , but no significant main effects of medication or psychotherapy, or their interaction. These findings indicate that the treatment provided was effective in reducing depression severity, but treatment groups did not differ in effectiveness.

### Factor structure of the MADRS

Descriptive statistics for the MADRS item and total scores at each assessment point are displayed in Table 2. Items nine and 10 demonstrated increasing skew and kurtosis at each successive assessment point; however, all indices were below recommended cutoffs.

### Factor structure of the MADRS over treatment

Goodness-of-fit indices for all previously proposed one-, two-, three-, and four-factor models are displayed in Table 3. Two models (Farner *et al.*, 2009; Craighead and Evans, 1996) produced non-positive definite matrices and therefore no valid solution was able to be estimated. Only the model proposed by Williamson *et al.* (2006) demonstrated acceptable fit according to the CFI and RMSEA; indeed, this model demonstrated a close fit according to the RMSEA. The model identified by Williamson *et al.* (2006) consists of four correlated factors – Sadness, Negative Thoughts, Detachment and Neurovegetative. All items loaded significantly onto their associated factors at each time point, although items four (reduced sleep) and five (reduced appetite) demonstrated lower loadings (e.g.  $< 0.25$ ).

**Table 2.** Descriptive characteristics for the MADRS item and total scores at each assessment point

| Item  | Month 0 (N = 821) |       |          |  | Month 1 (N = 817) |       |          |  | Month 3 (N = 735) |       |          |  | Month 6 (N = 685) |      |          |  |
|-------|-------------------|-------|----------|--|-------------------|-------|----------|--|-------------------|-------|----------|--|-------------------|------|----------|--|
|       | M (SD)            | Skew  | Kurtosis |  | M (SD)            | Skew  | Kurtosis |  | M (SD)            | Skew  | Kurtosis |  | M (SD)            | Skew | Kurtosis |  |
| 1     | 3.64 (0.95)       | -0.22 | 0.96     |  | 2.17 (1.12)       | 0.29  | 0.39     |  | 1.36 (1.06)       | 0.63  | 0.43     |  | 0.91 (1.05)       | 1.20 | 1.22     |  |
| 2     | 3.76 (0.83)       | -0.23 | 0.81     |  | 2.35 (1.10)       | -0.05 | -0.19    |  | 1.51 (1.10)       | 0.44  | -0.17    |  | 1.00 (1.08)       | 1.09 | 0.74     |  |
| 3     | 3.45 (0.84)       | -0.18 | 0.61     |  | 2.50 (0.98)       | -0.10 | 0.32     |  | 1.84 (1.02)       | -0.03 | -0.33    |  | 1.45 (1.00)       | 0.44 | 0.31     |  |
| 4     | 2.98 (1.40)       | -0.57 | -0.21    |  | 1.85 (1.27)       | 0.20  | -0.58    |  | 1.24 (1.18)       | 0.72  | -0.10    |  | 0.93 (1.07)       | 1.01 | 0.31     |  |
| 5     | 1.90 (1.52)       | 0.20  | -1.00    |  | 1.17 (1.26)       | 0.77  | -0.33    |  | 0.67 (1.00)       | 1.53  | 2.04     |  | 0.45 (0.92)       | 2.47 | 6.50     |  |
| 6     | 3.42 (0.96)       | -0.50 | 0.81     |  | 2.32 (1.15)       | -0.05 | 0.04     |  | 1.65 (1.08)       | 0.34  | -0.09    |  | 1.16 (1.06)       | 0.77 | 0.27     |  |
| 7     | 3.55 (0.98)       | -0.59 | 1.00     |  | 2.31 (1.18)       | -0.01 | -0.30    |  | 1.56 (1.10)       | 0.39  | -0.16    |  | 1.15 (1.09)       | 1.07 | 1.14     |  |
| 8     | 3.17 (0.99)       | -0.04 | 0.48     |  | 2.02 (1.11)       | -0.01 | -0.39    |  | 1.25 (1.05)       | 0.56  | -0.27    |  | 0.86 (0.97)       | 1.07 | 0.73     |  |
| 9     | 2.74 (0.97)       | 0.02  | -0.03    |  | 1.82 (1.02)       | 0.06  | -0.09    |  | 1.26 (1.02)       | 0.44  | -0.24    |  | 0.90 (0.96)       | 0.87 | 0.16     |  |
| 10    | 1.83 (1.21)       | 0.69  | 0.42     |  | 0.90 (0.97)       | 1.13  | 1.28     |  | 0.47 (0.74)       | 1.67  | 2.83     |  | 0.33 (0.70)       | 2.69 | 8.71     |  |
| Total | 30.44 (5.37)      | 0.54  | 0.18     |  | 19.40 (7.56)      | 0.10  | 0.21     |  | 12.83 (7.37)      | 0.54  | 0.19     |  | 9.13 (7.46)       | 1.28 | 1.57     |  |

**Table 3.** Goodness-of-fit of MADRS factor models

| Study                             | $\chi^2$ (df) | <i>p</i> | CFI  | RMSEA ( <i>p</i> ) |
|-----------------------------------|---------------|----------|------|--------------------|
| <i>One-factor model</i>           | 325.06 (35)   | <0.01    | 0.75 | 0.10 (<0.01)       |
| <i>Two-factor model</i>           |               |          |      |                    |
| Rocca <i>et al.</i>               | 295.09 (33)   | <0.01    | 0.78 | 0.10 (<0.01)       |
| <i>Three-factor models</i>        |               |          |      |                    |
| Galinowski and Leher              | 324.58 (35)   | <0.01    | 0.75 | 0.10 (<0.01)       |
| Andersson <i>et al.</i>           | 173.47 (29)   | < 0.01   | 0.88 | 0.08 (<0.01)       |
| Benazzi (Unipolar)                | 351.16 (27)   | <0.01    | 0.72 | 0.12 (<0.01)       |
| Benazzi (Bipolar II)              | 351.59 (29)   | <0.01    | 0.72 | 0.12 (<0.01)       |
| Parker <i>et al.</i>              | 242.00 (33)   | <0.01    | 0.82 | 0.09 (<0.01)       |
| Gabryelewicz <i>et al.</i>        | 486.80 (36)   | <0.01    | 0.61 | 0.12 (<0.01)       |
| Suzuki <i>et al.</i>              | 323.52 (32)   | <0.01    | 0.75 | 0.10 (<0.01)       |
| Farner <i>et al.</i> <sup>1</sup> | —             | —        | —    | —                  |
| <i>Four-factor models</i>         |               |          |      |                    |
| Craighead and Evans <sup>1</sup>  | —             | —        | —    | —                  |
| Williamson <i>et al.</i>          | 117.01 (31)   | <0.01    | 0.93 | 0.06 (0.11)        |

<sup>1</sup>This model demonstrated a non-positive covariance and did not provide a valid solution.

A hierarchical model in which these four factors loaded on a second-order Depression factor also provided a good fit to the data (see Figure 2). This analysis revealed that all factors loaded strongly and positively on a higher-order factor – supporting a general depression factor underlying all MADRS items, subsumed by more specific symptom subsets. This hierarchical model compared favourably to the original model. The Akaike's Information Criterion (AIC) can be used to compare fit

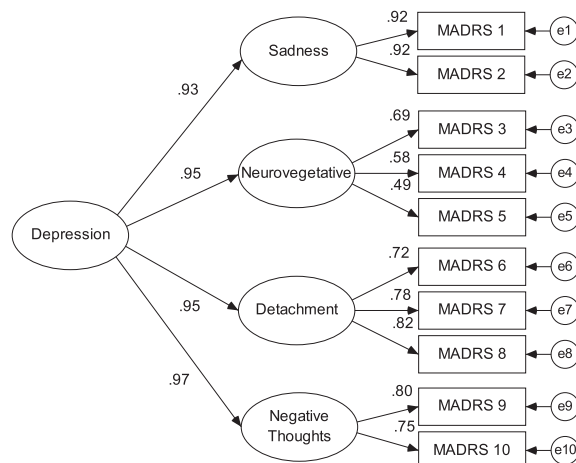


**Figure 2** Hierarchical model of the MADRS. Goodness of fit:  $\chi^2(33) = 117.03$ ,  $p < 0.01$ , CFI = 0.93; RMSEA = 0.05,  $p = 0.18$ .

across such non-nested models, with the model with the lower score to be preferred. The AIC associated with the Williamson *et al.* (2006) and hierarchical models were 185.01 and 181.03, respectively.

### Temporal and gender invariance of MADRS factor structure

Goodness-of-fit indices for the configural, metric, scalar, and strict models of the Williamson *et al.* (2006) structure are presented in Table 4. For temporal invariance analyses, the models associated with configural and metric invariance for Sadness and Negative Thoughts factors were equivalent, as these factors consisted of only two indicators. Configural invariance was supported for all factors; metric invariance was supported for Sadness, Negative Thoughts, and Detachment. There was mixed evidence for the metric invariance of the Neurovegetative factor. Scalar or strict invariance across time were not supported for any factor of the Williamson *et al.* (2006) model. For gender invariance analyses, all forms of invariance were supported. The same pattern of results was found at each time point and for the hierarchical model at each time point as well.

### Discussion

The results of the current study provide support for the structure of the MADRS items proposed by Williamson

**Table 4.** Temporal and Gender Invariance of MADRS Williamson *et al.* (2006) Model

| Factor            | Model      | $\chi^2$ | df | $\Delta\chi^2$ | $\Delta df$ | CFI   | $\Delta CFI$ | RMSEA ( <i>p</i> ) |
|-------------------|------------|----------|----|----------------|-------------|-------|--------------|--------------------|
| <i>Temporal</i>   |            |          |    |                |             |       |              |                    |
| Sadness           | Configural | —        | —  | —              | —           | —     | —            | —                  |
|                   | Metric     | 8.30     | 6  | —              | —           | 0.999 | —            | 0.02 (0.871)       |
|                   | Scalar     | 1290.91  | 12 | 1282.61**      | 6           | 0.582 | 0.417        | 0.40 (<0.01)       |
|                   | Strict     | 1368.66  | 18 | 77.75**        | 6           | 0.559 | 0.023        | 0.34 (<0.01)       |
| Negative Thoughts | Configural | —        | —  | —              | —           | —     | —            | —                  |
|                   | Metric     | 8.92     | 6  | —              | —           | 0.999 | —            | 0.03 (0.84)        |
|                   | Scalar     | 973.58   | 12 | 964.66**       | 6           | 0.525 | 0.474        | 0.35 (<0.01)       |
|                   | Strict     | 1228.98  | 18 | 255.40**       | 6           | 0.402 | 0.123        | 0.32 (<0.01)       |
| Detachment        | Configural | 25.94    | 30 | —              | —           | 1.00  | —            | 0.00 (1.00)        |
|                   | Metric     | 34.66    | 36 | 8.72           | 6           | 1.00  | 0.000        | 0.00 (1.00)        |
|                   | Scalar     | 1312.55  | 45 | 1277.89**      | 9           | 0.618 | 0.382        | 0.20 (<0.01)       |
|                   | Strict     | 1394.68  | 54 | 82.13**        | 9           | 0.596 | 0.022        | 0.19 (<0.01)       |
| Neurovegetative   | Configural | 67.02    | 30 | —              | —           | 0.982 | —            | 0.04 (0.78)        |
|                   | Metric     | 87.03    | 36 | 20.00*         | 6           | 0.975 | 0.007        | 0.05 (0.69)        |
|                   | Scalar     | 1188.95  | 45 | 1101.92**      | 9           | 0.437 | 0.538        | 0.20 (<0.01)       |
|                   | Strict     | 1577.79  | 54 | 388.84**       | 9           | 0.250 | 0.187        | 0.20 (<0.01)       |
| <i>Gender</i>     |            |          |    |                |             |       |              |                    |
| Gender            | Configural | 136.18   | 58 | —              | —           | 0.932 | —            | 0.04 (0.96)        |
|                   | Metric     | 145.84   | 68 | 9.66           | 10          | 0.933 | 0.001        | 0.04 (0.99)        |
|                   | Scalar     | 152.89   | 78 | 7.05           | 10          | 0.935 | 0.002        | 0.03 (1.00)        |
|                   | Strict     | 169.83   | 88 | 16.94          | 10          | 0.929 | 0.006        | 0.03 (1.00)        |

\**p* < 0.01.\*\**p* < 0.001.

*et al.* (2006): a four-factor model, including sadness, negative thoughts, detachment and neurovegetative symptoms. The Williamson *et al.* (2006) model was the only previously proposed factor model providing acceptable fit to the data at baseline. A hierarchical model in which the sadness, negative thoughts, detachment and neurovegetative domains of the depressive syndrome loaded onto a second order overall depression factor was also supported. These results support the use of the total scale score for the MADRS in addition to subscales to target particular symptoms of depression as is appropriate or needed. Subscales may provide more focused assessment of treatment course than does the total MADRS score; although these subscales demonstrated inadequate internal reliability at baseline ( $M = 0.55$ ), they were adequate for use at subsequent time points ( $M = 0.74$ , range 0.53–0.92).

Previous investigations of MADRS structure may not have converged upon a consistent solution due to differences in design and analysis. As described earlier, samples used in these studies differed across both demographic and clinical characteristics. Further, methods for determining the number of factors and the conditions under which to rotate extracted factors have varied. Statistical methods to evaluate psychometric properties have become increasingly advanced and empirically-based over time. The current study synthesizes previous empirical work, utilizing a large clinical sample and a comprehensive, confirmatory analytic approach.

The current study provided the first demonstration of temporal and gender invariance for the MADRS to our knowledge. Only configural and metric invariance was supported for the temporal invariance of MADRS factors. In contrast, the MADRS factor structure demonstrated a robust strict factor structure across gender. More precisely, the factors of the MADRS were manifested similarly and the items of the MADRS had the same operational definition across gender. The former invariance is an essential for the meaningful interpretation of cross-group comparisons.

Previously identified factors underlying the MADRS items have been utilized to characterize the phenomenology within and between diagnostic classes, as well as the differential course of specific symptom sets. For example, Rocca *et al.* (2002) reported that dysthymic disorder was principally associated with cognitive, social and motivational disruption as compared to somatic symptoms, and that both study medications significantly improved this class of difficulty. Parker *et al.* (2003) suggested that dysphoric apathy/retardation and psychic anxiety may most particularly reflect geriatric depression, and demonstrate distinct neurochemical underpinnings and treatment course. The limitations of the current investigation include an inability to assess the invariance of the MADRS across other important demographic characteristics, as well as the prognostic utility of the subscales identified here. Yet, this investigation contributes to the increasing support of the MADRS as one of the most psychometrically strong clinician-rated measures of depressive severity (Uher *et al.*, 2008). Both the total score and the subscales identified may be used justifiably within both male and female depressed outpatients to provide an overall as well as a more nuanced assay of treatment response. Future research may evaluate the differential prognostic utility of subscales, to empirically demonstrate their incremental validity.

### Acknowledgements

The authors are indebted to all patients and clinicians for their time and effort to complete all assessments and to the Institut de Recherches Internationales SERVIER (IRIS) for making this data available for analyses.

### Declaration of interest statement

All authors declare that they have no conflicts of interest nor do any of the authors have a financial interest in the results of the study.

### References

- Andersson S., Krogstad J.M., Finset A. (1999) Apathy and depressed mood in acquired brain damage: relationship to lesion localization and psychophysiological reactivity. *Psychological Medicine*, **29**, 447–456.
- Åsberg M., Perris C., Schalling D., Sedvall G. (1978) CPRS: development and applications of a psychiatric rating scale. *Acta Psychiatrica Scandinavica*, **271**, 69.
- Bagby R.M., Ryder A.G., Schuller D.R., Marshall M.B. (2004) The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *American Journal of Psychiatry*, **161**, 2163–2177.
- Benazzi F. (2001) Factor Analysis of the Montgomery Åsberg Depression Rating Scale in 251 bipolar II and 306 unipolar depressed outpatients. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, **25**, 1369–1376.
- Carmody T.J., Rush A.J., Bernstein I.H., Brannan S., Husain M.M., Trivedi M.H. (2006a) Making clinicians lives easier: guidance on use of the QIDS self-report in place of the MADRS. *Journal of Affective Disorders*, **95**, 115–118.
- Carmody T.J., Rush A.J., Bernstein I., Warden D., Brannan S., Burnham D., Woo A., Trivedi M.H. (2006b) The Montgomery-Åsberg and the Hamilton ratings of depression: a comparison of

- measures. *European Neuropsychopharmacology*, **16**, 601–611.
- Cheung G.W., Rensvold R.B. (2002) Evaluating goodness-of-fit indices for testing measurement invariance. *Structural Equation Modeling*, **9**, 233–255.
- Conroy D.E., Metzler J.N., Hofer S.M. (2003) Factor invariance and latent mean stability of performance failure appraisals. *Structural Equation Modeling*, **10**, 401–422.
- Corruble E., Legrand J.M., Duret C., Charles G., Guelfi J.D. (1999) IDS-C and IDS-SR: psychometric properties in depressed in-patients. *Journal of Affective Disorders*, **56**, 95–101.
- Craighead W.E., Evans D.D. (1996) Factor analysis of the Montgomery–Åsberg Depression Rating Scale. *Depression*, **4**, 31–33.
- De Fruyt F., Van Leeuwen K.G., Bagby R.M., Roland J.P., Rouillon F. (2006) Assessing and interpreting personality change and continuity in patients treated for major depression. *Psychological Assessment*, **18**, 71–80.
- Farner L., Wagle J., Flekkøy K., Wyller T.B., Fure B., Stensrød B., Engedal K. (2009) Factor analysis of the Montgomery Åsberg Depression Rating Scale in an elderly stroke population. *International Journal of Geriatric Psychiatry*, **24**, 1209–1216.
- Gabryelewicz T., Styczynska M., Pfeffer A., Wasiak B., Barczak A., Luczywek E., Androsiuk W., Barcikowska M. (2004) Prevalence of major and minor depression in elderly persons with mild cognitive impairment – MADRS factor analysis. *International Journal of Geriatric Psychiatry*, **19**, 1168–1172.
- Galinowski A., Leher P. (1995) Structural validity of MADRS during antidepressant treatment. *International Clinical Psychopharmacology*, **10**, 157–161.
- Hamilton M. (1960) A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*, **23**, 56–62.
- Hammond M.F. (1998) Rating depression severity in the elderly physically ill patient: reliability and factor structure of the Hamilton and the Montgomery–Åsberg Depression Rating Scales. *International Journal of Geriatric Psychiatry*, **13**, 257–261.
- Harvey P.D., Endicott J.M., Loebel A.D. (2008) The factor structure of clinical symptoms in mixed and manic episodes prior to and after antipsychotic treatment. *Bipolar Disorders*, **10**, 900–906.
- Harvey P.D., Green M.F., Bowie C., Loebel A. (2006) The dimensions of clinical and cognitive change in schizophrenia: evidence for independence of improvements. *Psychopharmacology*, **187**, 356–363.
- Higuchi H., Sato K., Yoshida K., Takahashi H., Kamata M., Otani K., Yamaguchi N. (2008) Predictors of antidepressant response to fluvoxamine obtained using the three-factor structure of the Montgomery and Åsberg Depression Rating Scale for major depressive disorders in Japanese patients. *Psychiatry and Clinical Neurosciences*, **62**, 301–306.
- Hunt-Shanks T., Blanchard C., Reid R., Fortier M., Cappelli M. (2010) A psychometric evaluation of the Hospital Anxiety and Depression Scale in cardiac patients: Addressing factor structure and gender invariance. *British Journal of Health Psychology*, **15**, 97–114.
- Khan A., Khan S.R., Shankles E.B., Polissar N.L. (2002) Relative sensitivity of the Montgomery Åsberg Depression Rating Scale, the Hamilton Depression Rating Scale and the Clinical Global Impressions Ratings Scale in antidepressant clinical trials. *International Clinical Psychopharmacology*, **17**, 281–285.
- Kline R.B. (2005) *Principles and Practice of Structural Equation Modeling*, 2nd edition. New York, Guilford.
- Millsap R.E., Kwok O. (2004) Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, **9**, 93–115.
- Montgomery S.A., Åsberg M. (1979) A new depression scale designed to be sensitive to change. *British Journal of Psychiatry*, **134**, 382–389.
- Mulder R.T., Joyce P.R., Frampton C. (2003) Relationships among measures of treatment outcome in depressed patients. *Journal of Affective Disorders*, **76**, 127–135.
- Parker R.D., Flint E.P., Bosworth H.B., Pieper C.F., Steffens D.C. (2003) A three-factor analytic model of the MADRS in geriatric depression. *International Journal of Geriatric Psychiatry*, **18**, 73–77.
- Riedel O., Heuser I., Klotzsche J., Dodel R., Wittchen H. (2010) Occurrence risk and structure of depression in Parkinson Disease with and without dementia: results from the GEPAD study. *Journal of Geriatric Psychiatry and Neurology*, **23**, 27–34.
- Rivera-Medina C.L., Caraballo J.N., Rodriguez-Cordero E.R., Bernal G., Dávila-Marrero E. (2010) Factor structure of the CES-D and measurement invariance across gender for low-income Puerto Ricans in a probability sample. *Journal of Consulting and Clinical Psychology*, **78**, 398–408.
- Rocca P., Fonzo P., Ravizza L., Rocca G., Scotta M., Zanalda E., Bogetto F. (2002) A comparison of paroxetine and amisulpride in the treatment of dysthymic disorder. *Journal of Affective Disorders*, **70**, 313–317.
- Santor D.A., Gregus M., Welch A. (2006) Eight decades of measurement in depression. *Measurement: Interdisciplinary Research and Perspectives*, **4**, 135–155.
- Serretti A., Jori M.C., Casadei G., Ravizza L., Smeraldi E., Akiskal H. (1999) Delineating psychopathologic clusters within dysthymia: a study of 512 out-patients without major depression. *Journal of Affective Disorders*, **56**, 17–25.
- Snaith R.P., Harrop F.M., Newby D.A., Telae C. (1986) Grade scores of the Montgomery–Åsberg Depression and the Clinical Anxiety Scales. *British Journal of Psychiatry*, **148**, 599–601.
- Suzuki A., Aoshima T., Fukasawa T., Yoshida K., Higuchi H., Shimizu T., Otani K. (2005) The three-factor model of the MADRS in major depressive disorder. *Depression and Anxiety*, **21**, 95–97.
- Uher R., Farmer A., Maier W., Rietschel M., Hauser J., Marusic A., Mors O., Elkin A., Williamson R.J., Schmael C., Henigsberg N., Perez J., Mendlewicz J., Janzing J.G., Zobel A., Skibinski M., Kozel D., Stamp A.S., Bajs M., Placentino A., Barreto M., McGuffin P., Aitchison K.J. (2008) Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychological Medicine*, **38**, 289–300.
- Williamson D., Brown E., Perlis R.H., Ahl J., Baker R.W., Tohen M. (2006) Clinical relevance of depressive symptom improvement in bipolar I depressed patients. *Journal of Affective Disorders*, **92**, 261–266.
- Wolthaus J.E.D., Dingemans P.M.A.J., Schene A.H., Linszen D.H., Knegtering H., Holthausen E.A.E., Cahn W., Hijman R. (2000) Component structure of the Positive and Negative Syndrome Scale (PANSS) in patients with recent-onset schizophrenia and spectrum disorders. *Psychopharmacology*, **150**, 399–403.