European Neuropsychopharmacology (2006) **16**, 601–611

www.elsevier.com/locate/euroneuro

The Montgomery Äsberg and the Hamilton ratings of depression: A comparison of measures

Thomas J. Carmody [a,*], A. John Rush [a], Ira Bernstein [b], Diane Warden [a], Stephen Brannan [c], Daniel Burnham [d], Ada Woo [b], Madhukar H. Trivedi [a]

[a] *Department of Psychiatry, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd. Dallas, TX 75390-9086, United States*
[b] *University of Texas at Arlington, United States*
[c] *Cyberonics, Inc., United States*
[d] *GlaxoSmithKline, King of Prussia, PA, United States*

Received 16 September 2005; received in revised form 5 April 2006; accepted 9 April 2006

KEYWORDS
MADRS;
HRSD;
Item response theory;
Classical test theory;
Psychometrics

Abstract The 17-item Hamilton Rating Scale for Depression ($HRSD_{17}$) and the Montgomery Äsberg Depression Rating Scale (MADRS) are two widely used clinician-rated symptom scales. A 6-item version of the HRSD ($HRSD_6$) was created by Bech to address the psychometric limitations of the $HRSD_{17}$. The psychometric properties of these measures were compared using classical test theory (CTT) and item response theory (IRT) methods. IRT methods were used to equate total scores on any two scales. Data from two distinctly different outpatient studies of nonpsychotic major depression: a 12-month study of highly treatment-resistant patients ($n=233$) and an 8-week acute phase drug treatment trial ($n=985$) were used for robustness of results.

MADRS and $HRSD_6$ items generally contributed more to the measurement of depression than $HRSD_{17}$ items as shown by higher item-total correlations and higher IRT slope parameters. The MADRS and $HRSD_6$ were unifactorial while the $HRSD_{17}$ contained 2 factors. The MADRS showed about twice the precision in estimating depression as either the $HRSD_{17}$ or $HRSD_6$ for average severity of depression. An $HRSD_{17}$ of 7 corresponded to an 8 or 9 on the MADRS and 4 on the $HRSD_6$.

The MADRS would be superior to the $HRSD_{17}$ in the conduct of clinical trials.

1. Introduction

The measurement of depressive symptom severity is important not only for the conduct of efficacy and effectiveness trials but increasingly also for the proper implementation of treatment guideline recommendations for major depressive and other mood disorders (Crismon et al., 1999; Depression Guideline Panel, 1993; Rush et al., 2003; Trivedi et al.,

* Corresponding author. Tel.: +1 214 648 4610; fax: +1 214 648 4612.
E-mail address: Thomas.Carmody@UTSouthwestern.edu (T.J. Carmody).

0924-977X/$ - see front matter © 2006 Elsevier B.V. and ECNP. All rights reserved.
doi:10.1016/j.euroneuro.2006.04.008

2004). A number of self-reports (e.g., Carroll Rating Scale, Beck Depression Inventory, Zung Self-Rating Scale, and Inventory of Depressive Symptomatology — Self Report) and clinician ratings (e.g., Hamilton Rating Scale for Depression, Montgomery Äsberg Depression Rating Scale, and Inventory of Depressive Symptomatology — Clinician-rated) are available. Perhaps the two most popular clinical ratings are the Hamilton Rating Scale for Depression (HRSD), which comes in several versions (e.g., 17, 21, 24, 28, and 31 items) (Hamilton, 1960, 1967) and the 10-item Montgomery Äsberg Depression Rating Scale (MADRS) (Montgomery and Äsberg, 1979). The MADRS is used frequently in European registration and other clinical trials, while the Hamilton continues to be more widely used in the United States, though recent reports (Bagby et al., 2004; Zimmerman et al., 2005) have highlighted significant shortcomings in the HRSD.

The MADRS has been reported as equivalent to or more sensitive to change in symptoms over time than the $HRSD_{17}$ (Mulder et al., 2003; Rivera et al., 2000; Senra, 1996) and equivalent to the $HRSD_{17}$ in detecting drug/placebo differences (Khan et al., 2002). The MADRS has been reported to be unifactorial after treatment (Galinowski and Lehert, 1995; Rocca et al., 2002), although more than one factor has been found using ratings with more limited ranges in total score (i.e., prior to treatment) (Corruble et al., 1999; Craighead and Evans, 1996; Galinowski and Lehert, 1995; Hammond, 1998; Rocca et al., 2002). A meta-analysis (Faries et al., 2000), however, found that the superiority of the MADRS or the HRSD in detecting differences between drug and placebo depended on the class of the medication, and the specific effects and side effects of the medication.

The $HRSD_{17}$ has been found consistently to be multidimensional (Bech et al., 1981; Gibbons et al., 1993; Hamilton, 1967; Maier et al., 1988), which may reduce its sensitivity to detecting changes in depression severity or in differentiating between two treatments. Prior analyses of the $HRSD_{17}$ have identified specific problematic items in terms of response characteristics (Bagby et al., 2004; Santor and Coyne, 2001). Several briefer versions of the HRSD have been developed to improve upon the $HRSD_{17}$ by creating a more unifactorial measure of depression that, consequently, should be more sensitive to detecting changes in depression or to detecting drug/placebo differences than the $HRSD_{17}$. The most commonly used brief HRSD may be that developed by Bech (Bech et al., 1975) — a 6-item scale that includes the following items: depressed mood, guilt, work and activities, retardation, psychic anxiety, and somatic symptoms general. In fact, the $HRSD_6$ has been found to be more clearly unidimensional (Bagby et al., 2004; Bech et al., 1992, 1997, 1984, 1975), more sensitive to change than the $HRSD_{17}$ (de Montigny et al., 1981; O'Sullivan et al., 1997), and equivalent to (Hooper and Bakish, 2000) or more sensitive to detecting drug/placebo or drug/drug differences than the $HRSD_{17}$ (Bech et al., 2000; Faries et al., 2000). This briefer version appears to have less psychometric bias as a result of side effects (Moller, 2001).

Item response theory (IRT) models (Embretson and Reise, 2000; Hambleton and Swaminathan, 1985; Hulin et al., 1983; Nunnally and Bernstein, 1994) represent an important and increasingly sophisticated framework for examination of the psychometric properties of rating scale total scores and individual items. The Rasch model has been used to examine

Table 1 Item total correlations for each measure in each study

Item	$HRSD_{17}$		MADRS		$HRSD_6$	
	(Study 1)	(Study 2)	(Study 1)	(Study 2)	(Study 1)	(Study 2)
1.	0.79	0.85	0.87	0.88	0.86	0.88
2.	0.50	0.66	0.89	0.90	0.62	0.72
3.	0.61	0.57	0.65	0.77	0.81	0.87
4.	0.51	0.56	0.47	0.68	0.52	0.60
5.	0.48	0.58	0.54	0.44	0.64	0.77
6.	0.42	0.55	0.68	0.79	0.67	0.75
7.	0.75	0.82	0.78	0.84		
8.	0.41	0.52	0.83	0.87		
9.	0.24	0.45	0.76	0.77		
10.	0.58	0.73	0.73	0.66		
11.	0.50	0.64				
12.	0.51	0.41				
13.	0.64	0.72				
14.	0.46	0.53				
15.	0.27	0.48				
16.	0.05	0.06				
17.	0.24	0.17				
Median	0.50	0.56	0.75	0.78	0.66	0.76

For the $HRSD_{17}$, the items are: 1. depressed mood; 2. guilt; 3. suicide; 4. initial insomnia; 5. middle insomnia; 6. delayed insomnia; 7. work and activities; 8. retardation; 9. agitation; 10. psychic anxiety; 11. somatic anxiety; 12. reduced appetite GI symptoms; 13. somatic symptoms general; 14. libido/genital symptoms; 15. hypochondriasis; 16. loss of insight; 17. weight loss. For the MADRS, the items are: 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; 10. suicidal thoughts. For the $HRSD_6$, the items are: 1. depressed mood; 2. guilt; 3. work and activities; 4. retardation; 5. psychic anxiety; and 6. somatic symptoms general.