ELSEVIER

Journal of Affective Disorders 102 (2007) 93–99

JOURNAL OF AFFECTIVE DISORDERS

www.elsevier.com/locate/jad

Research report

Sensitivity to change, discriminative performance, and cutoff criteria to define remission for embedded short scales of the Hamilton depression rating scale (HAMD)

Javier Ballesteros [a,f,*], Julio Bobes [b], Antonio Bulbena [c], Antonio Luque [d], Rafael Dal-Ré [e], Nora Ibarra [f], Itziar Güemes [a]

[a] *University of the Basque Country, UPV-EHU, Department of Neuroscience-Psychiatry, Leioa, Spain*
[b] *University of Oviedo, Department of Psychiatry, Oviedo, Spain*
[c] *Autonomous University of Barcelona, Department of Psychiatry, Barcelona, Spain*
[d] *AstraZeneca Farmacéutica Spain S.A., Madrid, Spain*
[e] *GlaxoSmithKline S.A., Madrid, Spain*
[f] *Maria Josefa Recio Foundation- Institute for Psychiatric Research, Bilbao, Spain*

Received 12 July 2006; received in revised form 14 December 2006; accepted 15 December 2006
Available online 26 January 2007

Abstract

Background: The Hamilton depression rating scale (HAMD) has been criticised for its multidimensionality, sensitivity to change, and discriminative power to define remission. To overcome these limitations several short scales have been devised but they have had limited use. We compared the performance of five HAMD short scales and their parental 17-item HAMD on sensitivity to change and discriminative power.
Methods: A local multicenter study was conducted with depressed outpatients ($n=113$). Depression severity was appraised at baseline and at 6 weeks since inception with the HAMD-17 and the clinical global impression scale (CGI). Sensitivity to change was calculated by a within-group standardised effect size (dw). Discriminative power (against a clinical remission criterion [CGI=1]) was assessed by receiver operating characteristic (ROC) analysis and the areas under the ROC curves (AUC).
Results: There were no differences among the five short scales on sensitivity to change (HAMD-17 dw: 1.6, 95% CI: 1.3–2.0; subscales range: 1.5–1.7), and discriminative power (HAMD-17 AUC: 0.93, 95% CI: 0.86–0.99; subscales range: 0.86–0.99). Appropriate cutoff points to define remission with short scales are suggested.
Limitations: The non-independence of the scales may have overestimated their performance. Nevertheless their comparisons seem fair as we do not expect a differential bias among them.
Conclusions: The short scales showed similar performance when compared with the parental HAMD. Since some were devised as unidimensional depression severity measures, and others to be sensitive to change, their use could circumvent previous criticisms raised to the canonical HAMD.

Keywords: Hamilton depression rating scale; HAMD; HDRS; Unidimensional versions; Short scales; Sensitivity to change; Remission cutoff points; Depression

* Corresponding author. University of the Basque Country, UPV-EHU, Department of Neuroscience-Psychiatry, Barrio Sarriena S/N, E-48940 Leioa; Spain.
E-mail address: javier.ballesteros@ehu.es (J. Ballesteros).

0165-0327/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jad.2006.12.015

1. Introduction

The Hamilton depression rating scale (HAMD) has played a leading role as a main outcome measure in randomised clinical trials (RCTs) designed to test the efficacy of antidepressant drugs. Originally devised to assess the severity of a representative set of depressive symptoms within the framework of a clinical interview with depressed patients (Hamilton, 1960; Hamilton, 1967), the HAMD was quickly incorporated at face value as an outcome measure into the armamentarium of trialists (Guy, 1976). However, several recent reviews have strongly challenged such leading role (Bagby et al., 2004; Zimmerman et al., 2005). The main criticisms to the HAMD as a depression outcome measure concern to its construct validity (it has been repeatedly shown that it represents a multidimensional measure reflecting more than the core symptoms of depression) (Bagby et al., 2004; Maier et al., 1985b), its sensitivity to adequately characterize the clinical change of depressive patients through their treatment (Montgomery and Asberg, 1979), and its problems to define remission with appropriate and empirically derived cutoff points (Zimmerman et al., 2005).

To overcome some of these criticisms, several short scales, based on item reduction of the HAMD by using techniques of Rasch analysis (Bech et al., 1975, 1981), factor analysis (Maier and Philipp, 1985a), and item response analysis (Gibbons et al., 1993), have been proposed in the literature. All these short scales have shown similar psychometric properties regarding sensitivity to change in RCTs (Entsuah et al., 2002; Faries et al., 2000; O'Sullivan et al., 1997), whereas the HAMD does not seem to perform worse than other competitor scales (Khan et al., 2002; Mulder et al., 2003). Recently, two new HAMD short scales have been added to the psychometric resources of clinical researchers (Evans et al., 2004; McIntyre et al., 2002, 2005). One of those scales has been devised by using item response analysis from the secondary analysis of several RCTs (Evans et al., 2004). The other has been empirically derived according to the frequency of response and sensitivity to change of the individual items of the parent HAMD-17 in depressed populations on treatment as usual (McIntyre et al., 2002, 2005).

Since we have recently published a Spanish validation of the HAMD canonical version (Bobes et al., 2003), and all the short scales so far reported are embedded in it, we thought it could be appropriate to use our data to perform new analyses comparing the sensitivity to change and the discriminative power (against the criteria of remission of depressive symptoms) of the HAMD short scales. Additionally, we tried to obtain appropriate cut-off points to define the criteria of remission for the HAMD-17 (as an internal control) and its related short scales. This paper updates previous reviews on the comparative assessment of the HAMD short scales (Entsuah et al., 2002; Faries et al., 2000; O'Sullivan et al., 1997) and gives also information about their performance in a population of mild to severe depressed outpatients treated as usual in a local multicenter study (Bobes et al., 2003).

2. Methods

2.1. Patients sample

The original characteristics and design of the psychometric study on which the analyses described in this paper relied on have been previously reported (Bobes et al., 2003). In short, it was a classical validation study for the canonical versions of the HAMD (21 and 17 items) and the Bech short version (Bech-6) (Bech et al., 1981) conducted on a sample of depressive outpatients on treatment by standard clinical management (treatment as usual) across 15 psychiatric centers in Spain. The study was designed as a psychometric study on its own and thus it was not embedded into or secondary to a RCT. The study protocol was approved by the corresponding ethical and research committees and all patients gave their written informed consent before being included in the study. Inclusion criteria were an age $\geq$18 years; to attend treatment as outpatient; and to be diagnosed (by DSM-IV criteria) of a major depressive disorder, a dysthymic disorder, or an adjustment disorder with depressed mood with a score $\geq$4 in the clinical global impression (CGI) (Guy, 1976) at baseline. At study inception patients were judged to be on stable or unstable clinical conditions by their psychiatrists on charge. Patients considered to be in stable condition were those who were not expected to change significantly in their clinical severity in one week time, whereas unstable patients were those others in whom a clinical significant change in severity was expected (either by a change of treatment or because of presenting new or recurrent episodes of depression). Stable patients underwent a test–retest reliability design at 1 week after inception and thus are of no further importance regarding the aims of this study. Unstable patients underwent a sensitivity to change design at 6 weeks after inception, and they account for the depressed patients included in this study.

2.2. Measures

All patients were assessed by their psychiatrists with the HAMD-21 (Hamilton, 1960) and the CGI (Guy, 1976) at baseline and at the end of the trial. We selected the HAMD-