Annals of General Psychiatry

An analysis of correlations among four outcome scales employed in clinical trials of patients with major depressive disorder
Qin Jiang* and Saeeduddin Ahmed

Address: Wyeth Research, Collegeville, Pennsylvania, PA, USA

Email: Qin Jiang* - jiangq2@wyeth.com; Saeeduddin Ahmed - ahmeds@wyeth.com

* Corresponding author

Abstract

Background: The 17-item Hamilton Depression Rating Scale (HAM-D$_{17}$) remains the 'gold standard' for measuring treatment outcomes in clinical trials of depressed patients. The Montgomery sberg Depression Rating Scale (MADRS), Clinical Global Impressions-Severity (CGI-S) and -Improvement (CGI-I) scales are also widely used.

Objective: This analysis of data from 22 double-blind, placebo-controlled clinical studies of venlafaxine in adult patients with major depressive disorder was aimed at assessing correlations among these 4 scales.

Methods: Changes from baseline for MADRS, HAM-D$_{17}$ and CGI-S, and end point CGI-I scores and response ($\geq$50% decrease from baseline HAM-D$_{17}$ or MADRS, or CGI-S or CGI-I score $\leq$2) were analysed. Pearson correlation coefficients were calculated for all pairs of the four scales (HAM-D$_{17}$/MADRS, HAM-D$_{17}$/CGI-S, HAM-D$_{17}$/CGI-I, MADRS/CGI-S, MADRS/CGI-I, CGI-S/CGI-I) at different time points. Effect sizes were calculated using the Cohen *d*.

Results: Correlations were significant at all time points (p < 0.0001), increased over the course of treatment, and were similar across treatment groups. Effect sizes ranged from 0.31 to 0.42; MADRS and CGI-I effect sizes were slightly greater compared with HAM-D$_{17}$ or CGI-S for continuous measures and response.

Conclusion: Although MADRS and CGI-I were more sensitive to treatment effects, HAM-D$_{17}$, MADRS, CGI-S and CGI-I scores present a consistent picture of response to venlafaxine treatment.

Background

Many instruments have been developed to measure outcomes in studies of patients with major depressive disorder (MDD). Among them, the Hamilton Depression Rating Scale (HAM-D) [1], the Montgomery sberg Depression Rating Scale (MADRS) [2], and the Clinical Global Impressions-Severity scale (CGI-S) and -Improvement scale (CGI-I) [3], are investigator-rated instruments; the CGI-I differs from the other three scales in that it assesses the degree of symptom improvement rather than absolute severity of symptoms or specific pathology [3]. The HAM-D and the MADRS scales measure depressive symptoms, whereas the CGI-S and CGI-I assess global outcome.

The HAM-D was developed in the 1950s to evaluate efficacy of first-generation antidepressants; the 17-item HAM-D (HAM-D$_{17}$) has been accepted by many as the standard for measuring therapeutic efficacy in clinical tri-

als [1]. However, one problem with the HAM-D is that individual items are often multidimensional, with poor inter-rater and retest reliability. As a result, the HAM-D total score can be ambiguous [4]. The MADRS was designed to address some of the limitations of the HAM-D. Specifically, the MADRS may be more sensitive to treatment-related changes in depression and may better distinguish responders from non-responders [2,5]. Recent analyses have confirmed the correlation between HAM-D, MADRS, and CGI-S in a systematic literature review and two retrospective chart reviews [4-6].

The present analysis was undertaken in a large dataset of 22 double-blind, placebo-controlled, clinical studies of venlafaxine in patients with MDD to identify and assess correlations among these 4 widely-used, rating scales: the HAM-D$_{17}$, MADRS, CGI-S, and CGI-I.

Methods
Studies and patients
Data were pooled from 22 multicenter, double-blind, placebo-controlled studies of venlafaxine (Table 1). All studies included adult patients with MDD, defined according to the diagnostic criteria from the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III [7], DSM-III-R [8], or DSM-IV [9] depending on when the study was designed). Outpatients were enrolled in 19 studies [10-22] and inpatients were enrolled in the other 3 studies [23] [Wyeth

Research: Data on File. Collegeville, PA, USA: Wyeth Research; 2006. unpublished data]. Two studies (016 and 206) enrolled patients with melancholia [10,23], and one study (360) enrolled patients with concomitant anxiety[21]. Study durations ranged from 4 weeks to 52 weeks.

Only data from patients receiving venlafaxine or placebo were included in this analysis, although 15 studies included an additional active-comparator arm [10-13,16-18,21] [unpublished data]. Venlafaxine extended release (ER) was used in 7 studies and venlafaxine immediate release (IR) in 14. In one trial, both formulations were used [14]. Venlafaxine IR was administered twice or three times daily in fixed or flexible doses ranging from 25 to 375 mg/d [11-14,16-20] [unpublished data]. Venlafaxine ER was administered once daily in fixed or flexible doses ranging from 37.5 to 375 mg/d [13-15,21,22] [unpublished data].

Statistical analysis
Continuous outcomes were defined as total change from baseline for MADRS and HAM-D$_{17}$, change in score from baseline for CGI-S, and end point scores for CGI-I. These scores were calculated using observed data for the total patient populations at weeks 1, 2, 3, 4, 6, and 8 (for studies less than 8 weeks in duration, data were included for the number of weeks available), and for the final on-therapy (FOT) visit. HAM-D$_{17}$, MADRS, CGI-S, and CGI-I

Table 1: Summary of 22 placebo-controlled clinical trials of venlafaxine for treatment of major depressive disorder[a]

Study no.	IR/ER	Fixed/flexible Dosing	Dose range (mg/day)	Practice setting	Duration (weeks)	Median baseline HAM-D$_{17}$
014 [11]	IR	Fixed	75, 150, 225	Outpatient	6	21
015 [12]	IR	Fixed	75, 150, 225	Outpatient	8	21
016 [10]	IR	Flexible	37.5 to 375	Inpatient	6	26
203 [16]	IR	Fixed ranges	75, 150 to 225, 300 to 375	Outpatient	6	22
206 [23]	IR	Flexible	150 to 375	Inpatient	4	27
208 [14]	IR and ER	Flexible	IR: 75 to 150; ER: 75 to 150	Outpatient	12	22
209 [15]	ER	Flexible	75 to 225	Outpatient	8	21
211 [13]	ER	Flexible	75 to 225	Outpatient	8	22.5
300	IR	Flexible	150 to 375	Inpatient	6	29
301	IR	Flexible	75 to 225	Outpatient	6	22
302 [17]	IR	Flexible	75 to 200	Outpatient	6	22
303 [18]	IR	Flexible	75 to 225	Outpatient	6	22
313 [19]	IR	Fixed ranges	25, 50 to 75, 150 to 200	Outpatient	6	23
341	IR	Flexible	100 to 200	Outpatient	52	22
342 [20]	IR	Fixed	75, 150, 200	Outpatient	12	22
343	IR	Fixed ranges	100 to 150, 175 to 225	Outpatient	14	20
360 [21]	ER	Flexible	75 to 225	Outpatient	12	25
367 [22]	ER	Fixed	75, 150	Outpatient	8	25
372	IR	Flexible	200 to 375	Outpatient	6	22
384	ER	Flexible	150 to 375	Outpatient	6	25
402	ER	Flexible	37.5 to 300	Outpatient	10	23
414	ER	Flexible	37.5 to 300	Outpatient	10	22

ER, extended release; HAM-D$_{17}$, 17-item Hamilton Depression Rating Scale; IR, immediate release.
[a]Data on File at Wyeth Research. 2006.

scores were stratified by treatment arm, and Pearson correlation coefficients were calculated for all possible pairs of the four scales (HAM-D_{17} vs MADRS, HAM-D vs CGI-S, HAM-D_{17} vs CGI-I, MADRS vs CGI-S, MADRS vs CGI-I, CGI-S vs CGI-I) for each of the data points.

The four scales also were used to determine binary outcomes (response or no response). For CGI-I and CGI-S, response was defined as scores ≤2, and for HAM-D_{17} and MADRS total scores, response was defined as a 50% or greater decrease from baseline. Pearson correlation coefficients were determined for all possible pairs of the four scales for binary outcomes at weeks 1 through 8. Correlations were calculated for the FOT scores for the total population, and separately for those in the venlafaxine and placebo arms.

Pearson product-moment correlation coefficient (*r*), a measure of the tendency of two variables to increase or decrease together, was used to measure the correlation of a pair of two efficacy variables measured on the same sub-

ject. Effect sizes (Cohen *d*) were calculated to measure the magnitude of the treatment effect at the FOT evaluation for the pooled data and individually for each study.

Results

At baseline, 5,117 observations were available for the HAM-D_{17}, 4,871 for the MADRS, and 5,103 for the CGI-S, respectively. Mean baseline scores were 23.0, 29.1, and 4.4 for HAM-D_{17}, MADRS, and CGI-S, respectively. Pretreatment correlations were 0.52 (CGI-S and HAM-D_{17}), 0.53 (CGI-S and MADRS), and 0.62 (HAM-D_{17} and MADRS).

Correlations between scales were significant at all time points (p < 0.0001) and increased over the course of treatment. At week 8, correlations ranged from 0.82 (CGI-S and CGI-I) to 0.92 (HAM-D_{17} and MADRS) (Figure 1). Correlations for the FOT scores also were significant (p < 0.0001), ranging from 0.87 (CGI-S and CGI-I) to 0.93 (HAM-D_{17} and MADRS) (Figure 2). Comparisons were
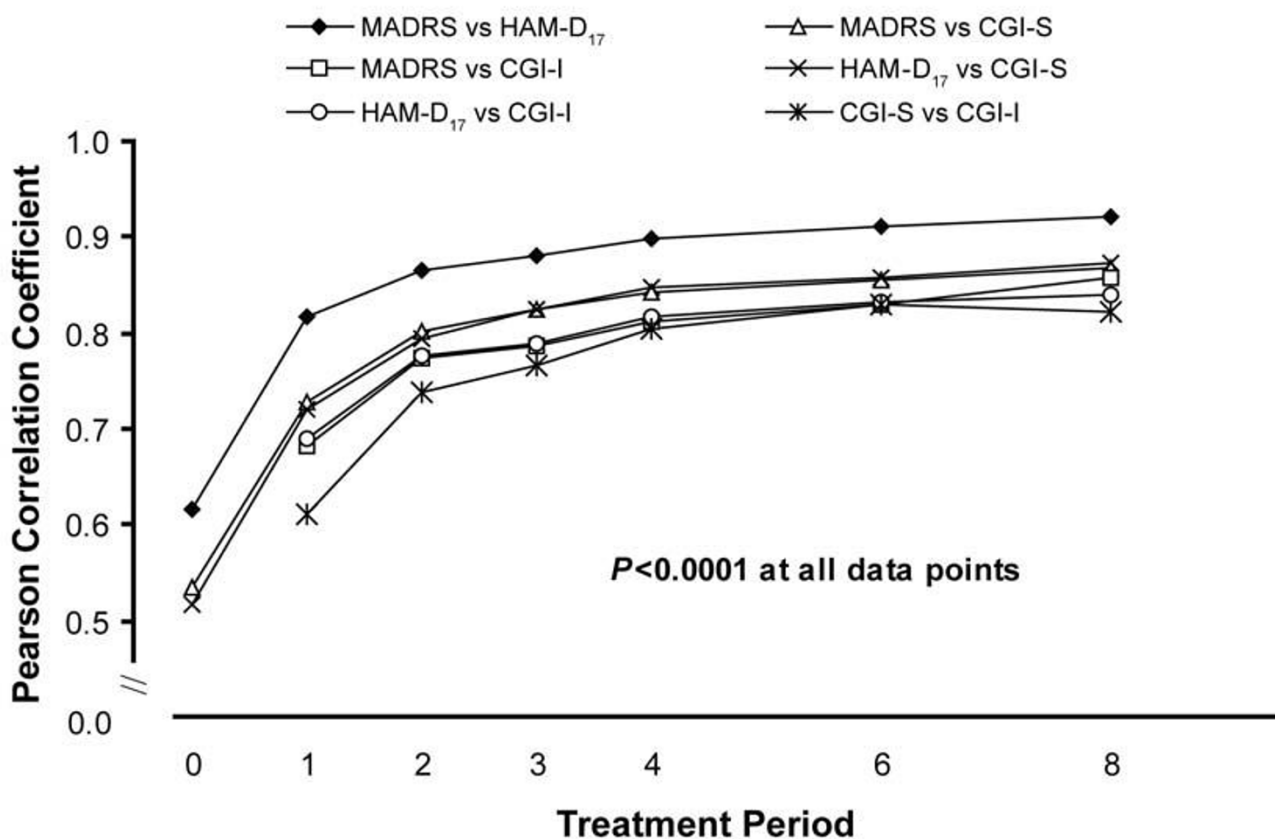
Figure 1
Correlation coefficients, changes from baseline (all patients). CGI-I, Clinical Global Impressions Improvement scale; CGI-S, Clinical Global Impressions Severity scale; HAM-D17, 17-item Hamilton Rating Scale for Depression; MADRS, Montgomery sberg Depression Rating Scale.
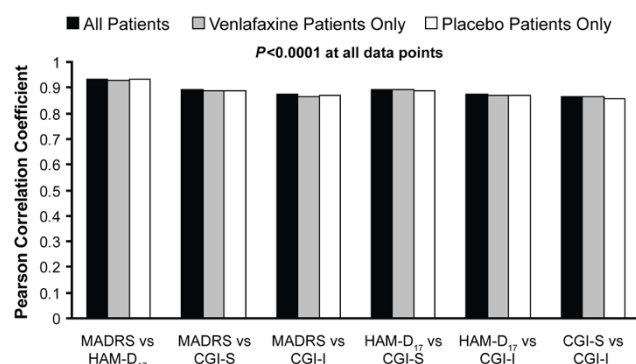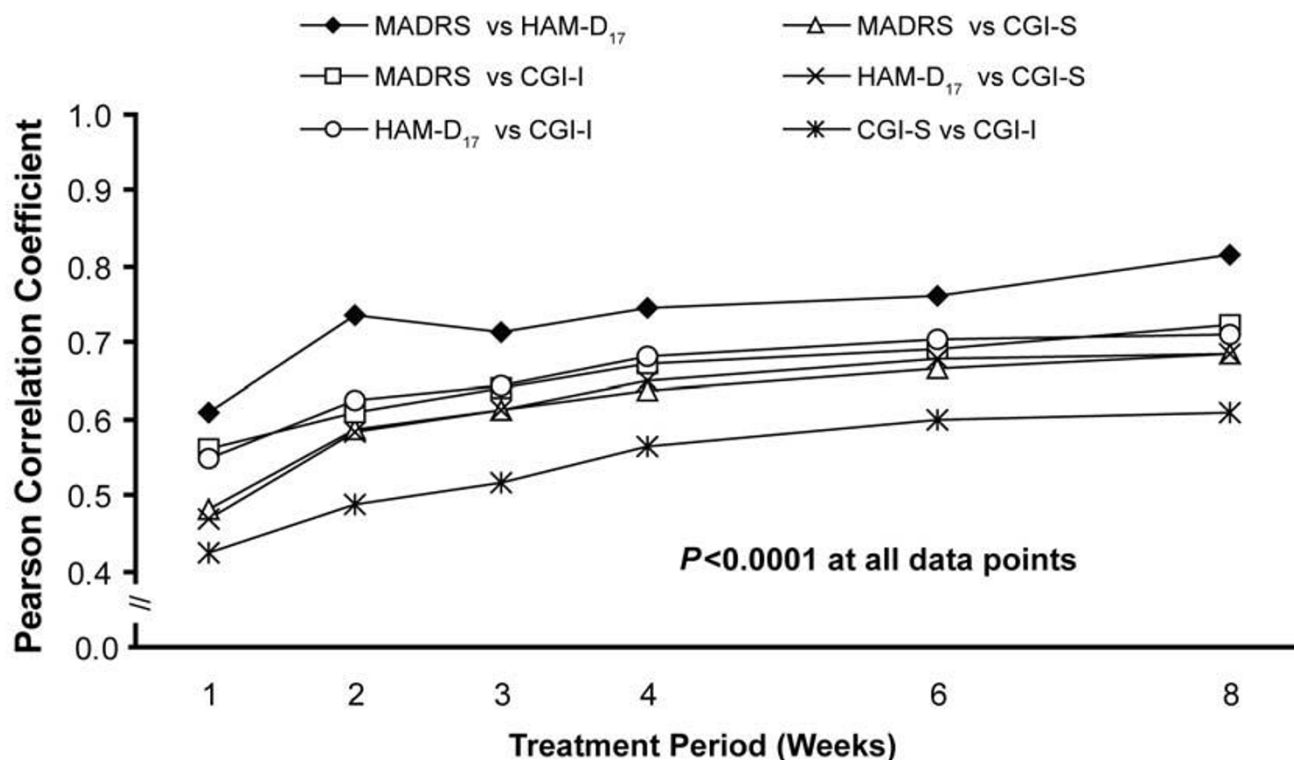
Figure 2
Pearson correlation coefficient, changes from baseline (final on therapy). CGI-I, Clinical Global Impressions Improvement scale; CGI-S, Clinical Global Impressions Severity scale; HAM-D17, 17-item Hamilton Rating Scale for Depression; MADRS, Montgomery sberg Depression Rating Scale.

statistically similar for the total population, the venlafaxine group, and the placebo group.

Correlation coefficients between binary outcomes (that is, response) were lower, ranging from 0.42 (CGI-I and CGI-S) to 0.61 (HAM-D_{17} and MADRS) at week 1 and from 0.61 (CGI-I and CGI-S) to 0.81 (HAM-D_{17} and MADRS) at week 8 (Figure 3). The correlations between binary outcomes at the FOT visit ranged from 0.68 (CGI-I and CGI-S) to 0.82 (MADRS and HAM-D_{17}) (Figure 4). All correlation coefficients were significant at all data points (p < 0.0001).

Pooled effect sizes for the continuous outcomes ranged from 0.39 on the CGI-I to 0.42 on the CGI-S (Figure 5). Effect sizes for the binary outcomes were lower, ranging from 0.31 (CGI-I response) to 0.41 (CGI-S response). Although differences were small, MADRS and CGI-I were better able to detect differences between venlafaxine and placebo than HAM-D_{17} or CGI-S for both sets of outcomes. Effect sizes across the individual studies varied considerably, but the pattern of results was largely consistent with that of the pooled data. In the majority of studies,
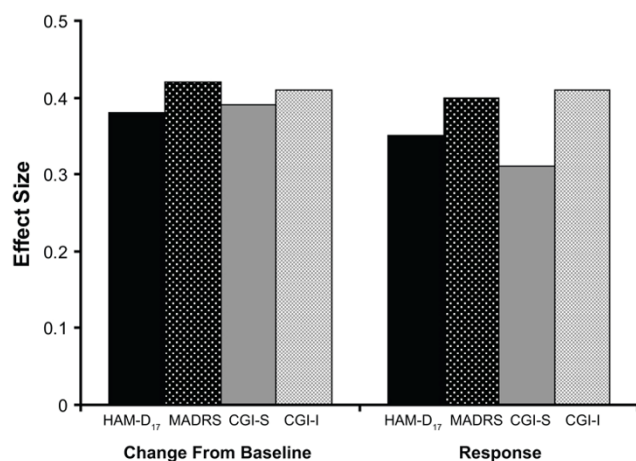

Figure 3
Correlation between definitions of response (all patients). CGI-I, Clinical Global Impressions Improvement scale; CGI-S, Clinical Global Impressions Severity scale; HAM-D17, 17-item Hamilton Rating Scale for Depression; MADRS, Montgomery sberg Depression Rating Scale.
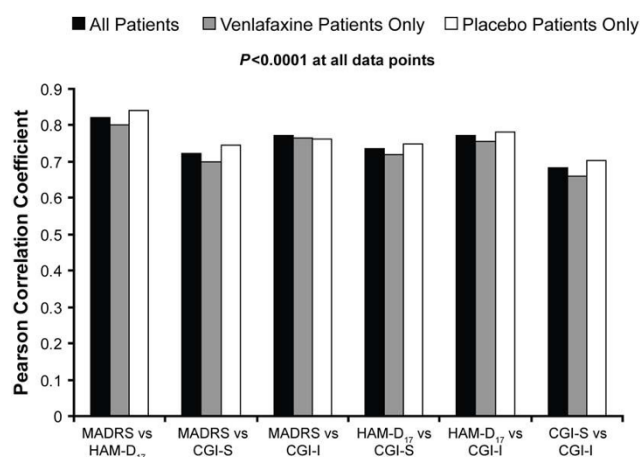
Figure 4
Correlation between definitions of response (final on therapy). CGI-I, Clinical Global Impressions Improvement scale; CGI-S, Clinical Global Impressions Severity scale; HAM-D17, 17-item Hamilton Rating Scale for Depression; MADRS, Montgomery sberg Depression Rating Scale.

effect sizes were greater on the CGI-I than CGI-S (continuous outcomes: 12 of 22 studies; response: 15 of 22 studies) and were greater on the MADRS compared with the HAM-D (continuous: 12 of 21 studies; response: 14 of 21 studies) (data not shown).

Figure 5
Effect size for venlafaxine vs placebo (all patients, final on therapy). CGI-I, Clinical Global Impressions Improvement scale; CGI-S, Clinical Global Impressions Severity scale; HAM-D17, 17-item Hamilton Rating Scale for Depression; MADRS, Montgomery sberg Depression Rating Scale.

Discussion
The data presented here, which are derived from a large pooled dataset from 22 clinical trials, confirm and expand results of earlier comparisons of these 4 commonly used depression rating scales [4-6]. Previous analyses have included data from samples that were smaller and rather homogeneous in terms of baseline depression severity and duration of treatment; these analyses evaluated treatment effects with a variety of antidepressants, including tricyclic antidepressants, selective serotonin reuptake inhibitors, and serotonin-norepinephrine reuptake inhibitors [5,6]. The trials in this analysis all included patients with MDD. However, the diagnostic criteria differed according to the DSM criteria accepted at the time individual studies were designed. All studies in this analysis used venlafaxine; however, they differed in the venlafaxine formulation used, dosing regimens (fixed or flexible), and duration of study treatment. The variability among the studies analysed here did not appear to confound the results, as the observations made using the HAM-D$_{17}$, MADRS, CGI-S, and CGI-I were highly correlated. Furthermore, despite the differences between this and other analyses, the findings are consistent [6]. As might be expected, the highest correlations were between the HAM-D$_{17}$ and the MADRS rating scales, which share several items, have similar modes of administration and rating, and are generally performed by the same clinician. However, in some clinical trials, depression rating assessments and assessments of global illness severity or improvement may be performed by different clinicians; this may have contributed to the lower correlations between the HAM-D$_{17}$ or MADRS scales and the CGI scales observed in this analysis. The consistently and modestly lower correlations between the CGI-S and CGI-I scales were unexpected as these scales are sometimes considered interchangeable. However, this may be explained by the relatively narrow distribution of the score range (1 to 7) compared with the ranges for the HAM-D$_{17}$ and MADRS total scores.

Although they were significant, correlation coefficients among binary outcomes based on the scales were lower than those for the change from baseline or FOT scores. Moreover, effect sizes were smaller for all scales in measuring the binary outcomes. These differences may be related to the definitions of response or no response that were used for the different scales. Some patients may have experienced significant improvement, which would be reflected in the change from baseline, although the scores did not meet the threshold for response.

Conclusion
Overall, these results suggest that HAM-D$_{17}$, MADRS, CGI-S, and CGI-I scores present a consistent picture of response to antidepressant therapy with venlafaxine.

List of abbreviations

CGI-I/S: Clinical Global Impressions-Improvement/-Severity scale; DSM: Diagnostic and Statistical Manual of Mental Disorders; ER: extended release; FOT: final on-therapy; HAM-D$_{17}$: 17-item Hamilton Depression Rating Scale; IR: immediate release; MADRS: Montgomery sberg Depression Rating Scale; MDD: major depressive disorder.

Competing interests

QJ is an employee of Wyeth; SA is a former employee of Wyeth.

Authors' contributions

Both authors contributed to the research and writing of this manuscript and were involved in the development of the statistical analysis plan. QJ performed the statistical analyses, both QJ and SA contributed to manuscript development and read and approved the final manuscript draft

Acknowledgements

References

1. Hamilton M: **A rating scale for depression.** *J Neurol Neurosurg Psychiatry* 1960, **23**:56-62.
2. Montgomery SA, sberg M: **A new depression scale designed to be sensitive to change.** *Br J Psychiatry* 1979, **134**:382-389.
3. Guy W: **Clinical Global Impressions.** In *ECDEU Assessment Manual for Psychopharmacology* Rockville, MD, USA: US Department of Health, Education, and Welfare; 1976:217-222.
4. Bagby RM, Ryder AG, Schuller DR, Marshall MB: **The Hamilton Depression Rating Scale: has the gold standard become a lead weight?** *Am J Psychiatry* 2004, **161**:2163-2177.
5. Khan A, Khan SR, Shankles EB, Polissar NL: **Relative sensitivity of the Montgomery-sberg Depression Rating Scale, the Hamilton Depression Rating Scale and the Clinical Global Impressions Rating Scale in antidepressant clinical trials.** *Int Clin Psychopharmacol* 2002, **17**:281-285.
6. Khan A, Brodhead AE, Kolts RL: **Relative sensitivity of the Montgomery-sberg Depression Rating Scale, the Hamilton Depression Rating Scale and the Clinical Global Impressions Rating Scale in antidepressant clinical trials: a replication analysis.** *Int Clin Psychopharmacol* 2004, **19**:157-160.
7. American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders (DSM-III)* Washington, DC, USA: American Psychiatric Association; 1981.
8. American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R)* Washington, DC, USA: American Psychiatric Association; 1987.
9. American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)* Washington, DC, USA: American Psychiatric Association; 1994.
10. Cantillon M, Daley M: **Further superiority of SNRI venlafaxine over SSRI fluoxetine in major depression and melancholia: a double-blind, placebo-controlled study of both response and remission (wellness).** Poster presented at: International Forum on Mood and Anxiety Disorders: Monte Carlo, Monaco; 2000.
11. Nemeroff CB, Thase ME: **A double-blind, placebo-controlled comparison of venlafaxine and fluoxetine treatment in depressed outpatients.** *J Psychiatr Res* 2007, **41**:351-359.
12. Schatzberg A, Roose S: **A double-blind, placebo-controlled study of venlafaxine and fluoxetine in geriatric outpatients with major depression.** *Am J Geriatr Psychiatry* 2006, **14**:361-370.
13. Rudolph RL, Feiger AD: **A double-blind, randomized, placebo-controlled trial of once-daily venlafaxine extended release (XR) and fluoxetine for the treatment of depression.** *J Affect Disord* 1999, **56**:171-181.
14. Cunningham LA: **Once-daily venlafaxine extended release (XR) and venlafaxine immediate release (IR) in outpatients with major depression. Venlafaxine XR 208 Study Group.** *Ann Clin Psychiatry* 1997, **9**:157-164.
15. Thase ME: **Efficacy and tolerability of once-daily venlafaxine extended release (XR) in outpatients with major depression. The Venlafaxine XR 209 Study Group.** *J Clin Psychiatry* 1997, **58**:393-398.
16. Rudolph R, Entsuah R, Aguiar L, Derivan A: **Early onset of antidepressant activity of venlafaxine compared with placebo and fluoxetine in outpatients in a double-blind study.** *Eur Neuropsychopharmacol* 1998, **8(Suppl 2):**S142.
17. Cunningham LA, Borison RL, Carman JS, Chouinard G, Crowder JE, Diamond BI, Fischer DE, Hearst E: **A comparison of venlafaxine, trazodone, and placebo in major depression.** *J Clin Psychopharmacol* 1994, **14**:99-106.
18. Schweizer E, Feighner J, Mandos LA, Rickels K: **Comparison of venlafaxine and imipramine in the acute treatment of major depression in outpatients.** *J Clin Psychiatry* 1994, **55**:104-108.
19. Mendels J, Johnston R, Mattes J, Riesenberg R: **Efficacy and safety of b.i.d. doses of venlafaxine in a dose-response study.** *Psychopharmacol Bull* 1993, **29**:169-174.
20. Khan A, Upton GV, Rudolph RL, Entsuah R, Leventer SM: **The use of venlafaxine in the treatment of major depression and major depression associated with anxiety: a dose-response study. Venlafaxine Investigator Study Group.** *J Clin Psychopharmacol* 1998, **18**:19-25.
21. Silverstone PH, Ravindran A: **Once-daily venlafaxine extended release (XR) compared with fluoxetine in outpatients with depression and anxiety. Venlafaxine XR 360 Study Group.** *J Clin Psychiatry* 1999, **60**:22-28.
22. Salinas E: **Once-daily extended release (XR) venlafaxine versus paroxetine in outpatients with major depression.** *Biol Psychiatry* 1997, **42(Suppl 1):**244S.
23. Guelfi JD, White C, Hackett D, Guichoux JY, Magni G: **Effectiveness of venlafaxine in patients hospitalized for major depression and melancholia.** *J Clin Psychiatry* 1995, **56**:450-458.