

A double-blind, placebo-controlled comparison of venlafaxine and fluoxetine treatment in depressed outpatients [☆]

Charles B. Nemeroff ^{a,*}, Michael E. Thase ^b, On behalf of the EPIC 014 Study Group

^a Department of Psychiatry and Behavioral Science, Emory University School of Medicine, 101 Woodruff Circle, Room 4115, Atlanta, GA 30322-4990, United States

^b Department of Psychiatry, University of Pittsburgh Medical Center, Pittsburgh, PA, United States

Received 5 April 2005; received in revised form 20 July 2005; accepted 20 July 2005

Abstract

This double-blind, placebo-controlled study compared venlafaxine (immediate release), the first modern serotonin-norepinephrine reuptake inhibitor, with the selective serotonin reuptake inhibitor fluoxetine. Outpatients were randomly assigned to 6 weeks of treatment with venlafaxine (75–225 mg/day; $n = 102$), fluoxetine (20–60 mg/day; $n = 104$), or placebo ($n = 102$). Efficacy was assessed using the 21-item Hamilton Depression Rating Scale (HAM-D₂₁), the Montgomery-Åsberg Depression Rating Scale (MADRS), the Clinical Global Impression-Severity of Illness (CGI-S) scale, response and remission rates, and several other measures. Intent-to-treat analyses utilized both the last observation carried forward and ETRANK methods to account for missing data. At week 6 or study endpoint, venlafaxine (mean dose: 142 mg/day) was superior to placebo on most outcomes measures, whereas the differences between fluoxetine (mean dose: 41 mg/day) and placebo were less consistent. Final remission (defined as HAM-D ≤ 7) rates were 32%, 28%, and 22% for venlafaxine, fluoxetine, and placebo, respectively. Few differences between the active treatments attained statistical significance. Both active therapies were generally well tolerated; however, attrition due to adverse events, incidence of selected side effects, and increases in pulse and blood pressure favored fluoxetine over venlafaxine. This study provides further evidence that venlafaxine is effective after 6 weeks of treatment compared with placebo. The efficacy profile of fluoxetine was somewhat less consistent. It is strongly recommended that future studies of comparative antidepressant efficacy be adequately powered to detect modest between-drug differences in efficacy.

© 2005 Published by Elsevier Ltd.

Keywords: Major depressive disorder; Antidepressants; Fluoxetine; Venlafaxine; Norepinephrine; Serotonin reuptake inhibitors

1. Introduction

Venlafaxine hydrochloride is the first member of a newer class of antidepressants now referred to as seroto-

nin-norepinephrine reuptake inhibitors (SNRIs) (Muth et al., 1986; Holliday and Benfield, 1995). Similar to fluoxetine, the first member of the selective serotonin reuptake inhibitor (SSRI) class of antidepressants to be widely used, venlafaxine has little or no affinity for cholinergic, adrenergic, or histaminergic receptors, which may account for the drugs' different tolerability profiles compared to tricyclic antidepressants (Ellingrod and Perry, 1994). Moreover, there is evidence that venlafaxine may be a more effective antidepressant than fluoxetine (Thase et al., 2001; Smith et al., 2002); this difference is presumed in part to be mediated by "dual" reuptake

[☆] Poster presentation at: 2001 American College of Neuropsychopharmacology, 2001 American Psychiatric Association; 2000 American Psychiatric Association, 2000 College of International Neuropsychopharmacology, 2000 New Clinical Drug Evaluation Unit. This study was funded by a series of grants to the participating research sites from Wyeth Research, Collegeville, PA.

* Corresponding author. Tel.: +1 404 727 8382; fax: +1 404 727 3233.

E-mail address: cnemer@emory.edu (C.B. Nemeroff).

inhibition (Thase et al., 2001). Not all studies have found a difference between these therapies, however, and, as only 2 published studies have included a placebo control group (Rudolph and Feiger, 1999; Silverstone and Ravindran, 1999), the assay sensitivity of the individual trials may be questioned (Thase, 2002). The current double-blind, placebo-controlled study therefore was undertaken to further assess the relative efficacy and safety of venlafaxine and fluoxetine.

2. Methods

The study was conducted in 13 centers (both university-affiliated and private research clinics) in the United States. The protocols were reviewed and approved by the institutional review boards of each study site, and the study was conducted according to the Declaration of Helsinki and its amendments. Participants signed a written informed consent at the time of their enrollment.

2.1. Patients

2.1.1. Inclusion criteria

Participants were outpatients 18 years or older and met *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV) criteria for major depressive disorder (American Psychiatric Association, 1994). All patients had symptoms present for at least 1 month before study entry and scored at least 20 on the 21-item Hamilton Rating Scale for Depression (HAM-D₂₁).

2.1.2. Exclusion criteria

Patients were excluded if they had a history or presence of bipolar disorder or any psychotic disorder. Patients with a history of alcohol or substance abuse within the past year were excluded from the study, as were those who had any clinically significant medical disorders or abnormalities detected during the prestudy physical screening that might compromise study participation. Additionally, patients were excluded if they were acutely suicidal to the degree that precautions against suicide were needed. Another cause for exclusion was a history of nonresponse to venlafaxine or fluoxetine. Further, any patient who had received either study drug within 6 months prior to starting the double-blind treatment period was excluded. Patients were excluded if they had received any of the following treatments before entering the trial: electroconvulsive therapy within 3 months; any investigational drug or antipsychotic drug within 30 days; astemizole, cisapride, sumatriptan, terfenadine, any monoamine oxidase inhibitor, paroxetine, or sertraline within 14 days; any other antidepressant, anxiolytic, sedative-hypnotic drug (except chloral hydrate), or any other psychotropic drug within 7 days of the start of double-blind treatment; or any other drug

with psychotropic effects within 7 days of the start of the double-blind treatment period unless a stable dose of the drug had been maintained for at least 1 month (3 months for thyroid or hormonal medications) before study day 1. Pregnant or lactating women were excluded from the study, as were women capable of childbearing who were unwilling to use a medically acceptable form of contraception.

Study candidates underwent a complete evaluation, including a psychiatric history and examination, HAM-D₂₁, medical history and physical examination, laboratory assessments, and electrocardiogram (ECG). Potentially eligible patients entered a single-blind, placebo phase, which typically lasted 7 days (± 3 days). At study day 0, the HAM-D₂₁ was repeated and other efficacy measurements described below were performed. Eligibility for randomization required no more than a 20% decrease in HAM-D₂₁ during the single-blind, placebo lead-in. Patients meeting the entry criteria at both the screening and baseline visits were then randomly assigned to double-blind therapy with venlafaxine immediate release (IR), fluoxetine, or placebo.

2.2. Study medication

Study medications were dispensed in identically appearing capsules. Double-blind therapy was initiated as follows: venlafaxine IR 37.5 mg twice each day, fluoxetine 20 mg in the morning (with one placebo capsule in the evening), or placebo twice daily. If clinically indicated, patients could receive up to 2 dose increases. The venlafaxine dosage could be increased on day 15–75 mg twice daily (150 mg/day), and again on day 29–112.5 mg twice daily (maximum permitted dose: 225 mg/day). Likewise, fluoxetine could be increased on day 15–40 mg/morning (with 2 placebo capsules in the evening) and again on day 29 to the maximum permitted dose, 60 mg/morning (with 3 placebo capsules in the evening). At the end of the study, patients receiving more than 2 capsules daily entered a medication taper period of up to 2 weeks, decreasing their dosage in steps before stopping medication to avoid discontinuation effects.

2.3. Efficacy and safety assessments

In addition to the HAM-D₂₁ total score, item 1 of the HAM-D (depressed mood), the Montgomery-Åsberg Depression Rating Scale (MADRS) total score, and the Clinical Global Impression-Severity of Illness (CGI-S) and -Improvement (CGI-I) scores were considered to be primary efficacy variables in the original data analytic plan. Efficacy assessments were repeated at weeks 1, 2, 3, 4, and 6 or study endpoint. Four definitions of response were compared: (1) $\geq 50\%$ decrease (from week 0) in the HAM-D₂₁ score, (2) $\geq 50\%$ decrease in the MADRS, (3) a final CGI-I score of 1 or

2 (very much or much improved), and (4) a final Patient Global Impressions (PGI) score of 1 or 2 (very much or much improved). Remission was defined in the original data analytic plan as a HAM-D₂₁ total score ≤ 8 . As remission was defined in subsequent work as a score ≤ 7 on the first 17 items of the HAM-D (Thase et al., 2001), we report rates according to both definitions.

A battery of brief questionnaires was administered at week 0 and repeated at week 6 or endpoint to assess: (1) general life functioning (General Life Functioning Scale) (Elkin et al., 1985; Lenderking et al., 1999), (2) social and task-related activities (Activities Questionnaire) (Lenderking et al., 1999), (3) cognitive functioning (Cognitive Functioning Scale from the Medical Outcomes Study [MOS]) (Ware and Sherbourne, 1992), (4) general health (General Health Scale from the MOS Short-Form Health Survey [SF-36]) (Ware and Sherbourne, 1992), and (5) vitality (Vitality Scale from the MOS SF-36) (Ware and Sherbourne, 1992).

Adverse events (AEs), efficacy, and vital signs were evaluated at each visit and physical examinations, laboratory determinations, and ECGs were repeated at the end of study. Study events were identified weekly from spontaneous patient reports, signs or symptoms detected during physical examination and clinical evaluation, and patient response to the nonspecific question “How have you been feeling since your last visit?” Safety was evaluated on the basis of reports of study events and the results of physical examinations, laboratory determinations, and ECGs. Blood pressure was taken in the same arm and performed by the same site personnel throughout the study. Patients were instructed to avoid smoking or consumption of beverages containing caffeine for at least 2 h prior to each blood pressure measurement. Patients rested for at least 2 min prior to the initial blood pressure reading at each visit. After supine pulse was measured, 2 consecutive supine blood pressure measurements were taken 2 min apart. Patients then rose to a standing position, and blood pressure was recorded after 1 min and again 2 min later.

2.4. Statistical analysis

The original data analytic plan followed standards used for submission of results to the FDA and, when viewed from more contemporary perspectives, is probably inefficient and redundant. Further, a deficiency existed in the study design in that the protocol defined 5 primary outcome measures but did not provide for any statistical adjustments. Nevertheless, to fully report the work that was done, a description of that plan follows. Efficacy analyses were performed on the modified intent-to-treat population, which was defined as all patients who took at least one dose of double-blind study medication and completed at least one efficacy evalua-

tion during treatment. Data were evaluated by both observed cases (OC) and last-observation-carried-forward (LOCF) methods. For the LOCF analysis, the last available on-therapy evaluation was entered for missing visits.

Pretreatment scores on the HAM-D₂₁, HAM-D item 1, MADRS, CGI-S, and the quality-of-life questionnaires were first compared with a 1-way analysis of variance (ANOVA) using treatment group as the factor. A 2-way analysis of covariance (ANCOVA), using treatment and site as the main factors and pretreatment scores as the covariate, was used to analyze each scheduled postbaseline visit (weeks 1, 2, 3, 4, and 6). Because the CGI-I has no pretreatment score, no covariate was used. Response and remission rates were compared at each scheduled visit using χ^2 tests.

With respect to the redundancy of the original analytic plan, we report here only results of the week 6 LOCF analyses. Results of interim assessments are illustrated graphically for the HAM-D, MADRS, and CGI-S. It is noted that there were few differences between the LOCF and OC analyses and, when differences did exist, they favored one or both active drugs over placebo.

The use of LOCF to handle missing data also can be problematic (Lavori et al., 1995). Although this method has been the standard accepted by the FDA for decades, it is quite conservative and distorts the variance structure of repeated measurements. To address this issue, a secondary analysis that directly addresses treatment-related withdrawals, ETRANK (Entsuah, 1990, 1996), was conducted using HAM-D assessments. Briefly, this nonparametric procedure assigns scores to patients depending on the degree of response and whether or not the patient completed the study. All non-missing data are assigned scores based on the degree of response. Missing data from premature withdrawals to AEs or lack of efficacy are assigned the worst scores. The model can be used for all the longitudinal observations (full data) or only the last observation (endpoint) for a patient. Several algorithms for scoring are available within ETRANK procedure: (1) scores that are equal to actual ranks, (2) scores that assign heavier weights to later observations, (3) scores that weigh middle observations more similarly to each other than extreme observations, or (4) the observed data. For this report, the test statistic, *F*, is presented for each of the scoring procedures. Although the protocol defined 5 primary outcome measures, the ETRANK analysis was limited to HAM-D assessments, which we considered the most clinically meaningful of the protocol defined primary measures.

Safety analyses were performed on all patients who took at least one dose of study medication. Adverse events were defined as treatment emergent if they appeared on or after the first dose of study medication,

or, if present prior to the first dose of study medication, they increased in severity during the study.

ECG findings, vital signs (means at baseline and mean changes from baseline to final on-therapy visit), and body weights were summarized for each treatment group. Laboratory and vital signs parameters were compared among treatment groups using 2-way ANCOVA with treatment and investigator as the main factors, and pretreatment value as the covariate.

3. Results

One hundred two (102) patients were randomized to venlafaxine, 104 to fluoxetine, and 102 to placebo. The mean daily dosages were 142 (SD 64) mg/day for venlafaxine and 41 (SD 17) mg/day for fluoxetine. Two patients each in the venlafaxine and fluoxetine groups were lost to follow-up almost immediately and could not be included in the safety analysis. In addition, 6 patients (4 in the venlafaxine group, 1 in the fluoxetine group, and 1 in the placebo group) could not be included in the efficacy analysis due to lack of efficacy assessments either at pretreatment or posttreatment, leaving 96 patients in the venlafaxine group, 100 in the

fluoxetine group, and 101 in the placebo group for efficacy analysis. Pretreatment clinical and demographic characteristics did not differ significantly among treatment groups (Table 1).

Nearly 80% of participants completed the trial and the proportions of completers were similar for the 3 treatment groups (Table 2). Significant differences among groups were observed for discontinuation due to adverse events (described later) and miscellaneous reasons (venlafaxine 0%, fluoxetine 1%, and placebo 5%, $\chi^2 = 7.22$; $df = 2$; $P = 0.027$). However, there were no significant differences in discontinuation due to lack of efficacy (venlafaxine 4%, fluoxetine 4%, placebo 6%; $\chi^2 = 0.63$; $df = 2$; $P = 0.730$).

3.1. Efficacy measures

The improvements in the 3 treatment groups in HAM-D₂₁, MADRS, and CGI-S scores are shown in Fig. 1. On the HAM-D, overall differences among treatment groups at week 6 did not quite reach statistical significance ($F(2,281) = 3.01$, $P = 0.051$), though the difference between the venlafaxine and placebo groups was statistically significant ($F(1,281) = 5.92$, $P = 0.016$). The differences between fluoxetine and placebo

Table 1
Pretreatment demographic and clinical characteristics

	Venlafaxine (n = 102)	Fluoxetine (n = 104)	Placebo (n = 102)	Test statistic; df	P Value
Ethnic origin (%)					
Caucasian	91	93	92	1.85; (8) ^a	.985
African-American	3	2	2		
Asian	1	1	1		
Hispanic	4	3	5		
Other	1	1	0		
Sex (%)					
Male	35	31	44	4.08; (2) ^a	.130
Female	65	69	56		
Age (y)					
Mean (SD)	40.1 (11.1)	37.9 (11.5)	40.4 (11.7)	1.50; (2, 305) ^b	.224
Range	19–70	19–75	18–72		
Weight (kg)					
Mean (SD)	78.5 (17.4)	80.2 (21.6)	80.4 (18.4)	0.28; (2, 300) ^b	.754
Range	48.5–119.8	48.5–204.1	46.3–127.0		
HAM-D ₂₁ total score, mean (SD)	23.5 (3.2)	23.7 (3.2)	23.7 (3.3)	0.18; (2, 300) ^b	.873
HAM-D ₂₁ depressed mood item score, mean (SD)	2.7 (0.6)	2.8 (0.5)	2.8 (0.6)		
MADRS total score, mean (SD)	26.9 (5.3)	26.9 (4.5)	27.8 (5.1)		
Duration of current depressive episode (wk)					
Median	55	71	54.5	2.08 (2) ^c	.354
Range	3–776	4–1560	4–2444		
Prior antidepressant medications (%)	49	41	38	2.57; (2) ^a	.277

HAM-D₂₁, 21-item Hamilton Depression Rating Scale.

MADRS, Montgomery-Åsberg Depression Rating Scale; and SD, standard deviation.

^a χ^2 test; (df).

^b F test; (n, d).

^c Kruskal–Wallis χ^2 .

Table 2
Primary reasons for discontinuation from study (%)

Population	Venlafaxine (n = 102)	Fluoxetine (n = 104)	Placebo (n = 102)	χ^2	P Value
All causes	24	18	24	1.12	.571
Adverse experience	12	7	3	6.03	.049
Failed to return	8	7	8	0.12	.940
Unsatisfactory response	4	4	6	0.63	.730
Protocol violation	0	0	2	4.07	.131
Other	0	1	5	7.22	.027

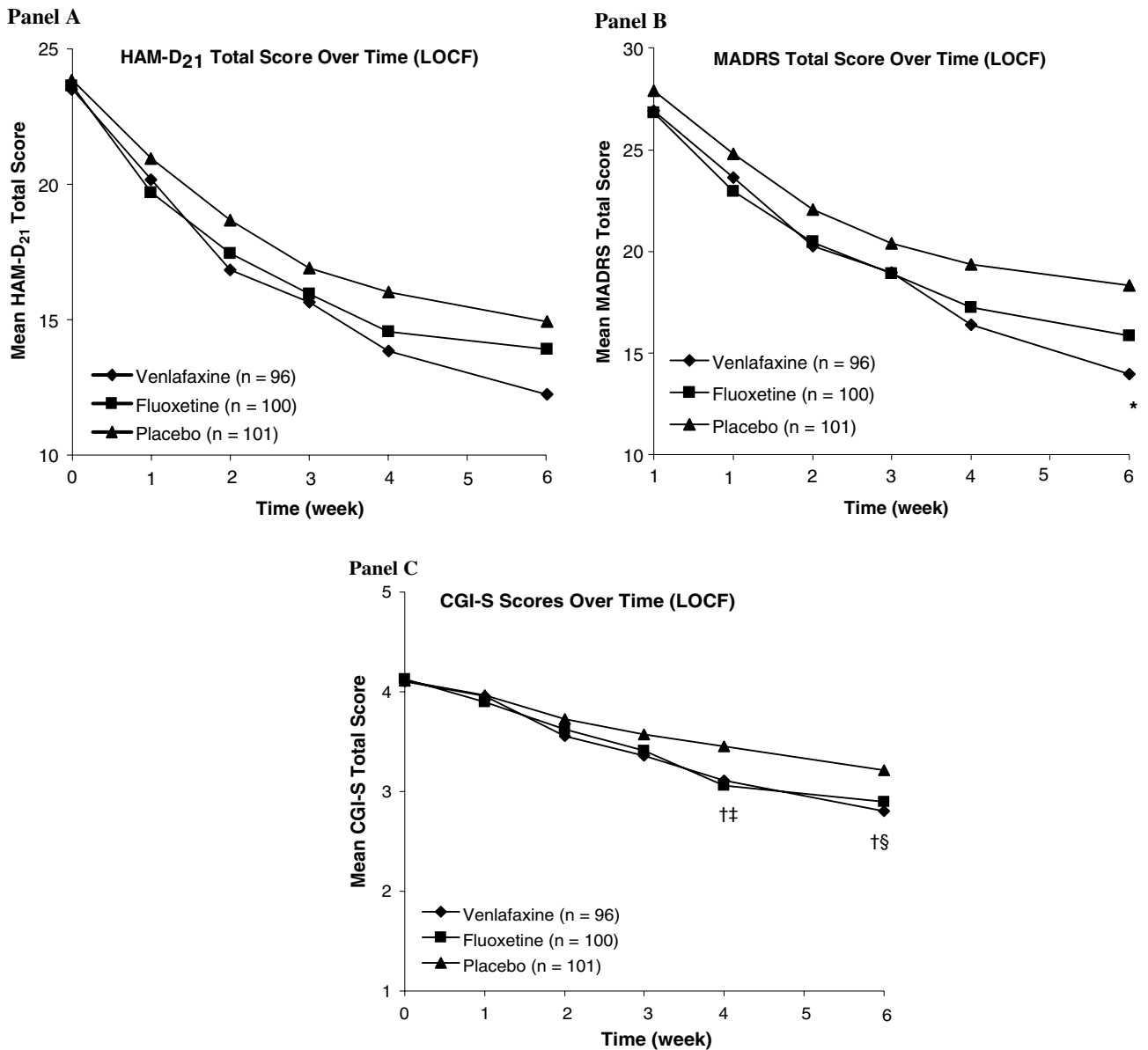


Fig. 1. Mean scores over time (LOCF): (A) HAM-D₂₁ total score; (B) MADRS total score; (C) CGI-S total score; * $P < 0.01$ venlafaxine vs. placebo; † $P < 0.05$ venlafaxine vs. placebo; ‡ $P < 0.01$ fluoxetine vs. placebo; § $P < 0.05$ fluoxetine vs. placebo. P values are from a one-way ANOVA at baseline and two-way ANCOVA with baseline as covariate and treatment and investigator as factors at postbaseline visits.

($F(1,281) = 0.85, P = 0.358$) and between venlafaxine and fluoxetine ($F(1,281) = 2.31, P = 0.130$) were not statistically significant.

The difference on the HAM-D depressed mood item was statistically significant among treatment groups at week 6 ($F(2,281) = 7.38, P \leq 0.001$); both active treat-

ments were significantly more effective than placebo (venlafaxine, $F(1, 281) = 14.56$, $P \leq 0.001$; fluoxetine, $F(1, 281) = 5.15$, $P = 0.024$). The difference between the active treatments was not statistically significant ($F(1, 281) = 2.47$, $P = 0.117$).

There was also a statistically significant difference among treatment groups at week 6 on the MADRS ($F(2, 281) = 4.48$, $P = 0.012$). Whereas venlafaxine therapy was significantly more effective than placebo ($F(1, 281) = 8.95$, $P = 0.003$), fluoxetine therapy was not ($F(1, 281) = 2.49$, $P = 0.116$). The difference between the fluoxetine and venlafaxine groups was not statistically significant ($F(1, 281) = 2.06$, $P = 0.152$).

A significant between-groups difference in CGI-S scores at week 6 also was observed ($F(2, 281) = 3.65$, $P = 0.027$); both active treatments were significantly more effective than placebo (venlafaxine: $F(1, 281) = 6.26$, $P = 0.013$; fluoxetine: $F(1, 281) = 4.49$, $P = 0.035$).

With respect to CGI-I scores, there was a significant effect for treatment ($F(2, 282) = 4.25$, $P = 0.015$) at week 6. Both venlafaxine ($F(1, 282) = 7.74$, $P = 0.006$) and fluoxetine ($F(1, 282) = 4.52$, $P = 0.035$) were more effective than placebo. There were no statistically significant differences between the venlafaxine and fluoxetine therapy groups on either CGI measure (CGI-S:

$F(1, 281) = 0.16$, $P = 0.689$; CGI-I: $F(1, 282) = 0.46$, $P = 0.499$).

3.2. Response and remission rates

The differences between the venlafaxine- and placebo-treated groups were statistically significant on all 4 definitions of response (Table 3). Fluoxetine therapy was significantly more effective than placebo according to the CGI and PGI definitions of response only. Neither active therapy separated significantly from placebo on the remission definitions. Among the 6 pair-wise comparisons of response/remission rates, a significant difference favoring venlafaxine over fluoxetine was found only on the PGI.

3.3. ETRANK analyses

Longitudinal nonparametric analysis on the HAM-D demonstrated statistically significant differences using all 4 ranking methods, both in the full-data analysis and the end point analysis. As shown in Table 4, statistically significant differences were observed for venlafaxine-placebo (full data and end point) and venlafaxine-fluoxetine (end point only) comparisons. No statistically

Table 3
Response and remission rates at week 6 or endpoint

	Venlafaxine	Fluoxetine	Placebo	Test statistic	P Value
HAM-D response ^a	53 (51/96)	45 (45/100)	37 (37/101)	5.42	0.067 ^c
MADRS response ^a	52 (50/96)	44 (44/100)	34 (34/101)	6.86	0.032 ^f
CGI response ^a	61 (59/96)	53 (54/101)	38 (38/101)	11.66	0.003 ^d
PGI Response ^b	67 (63/94)	52 (51/98)	38 (37/98)	16.58	0.001 ^e
Remission ≤ 8 ^a	32 (31/96)	32 (32/101)	22 (22/101)	3.41	0.181
Remission ≤ 7 ^{ag}	32 (31/96)	28 (28/101)	22 (22/101)	2.77	0.250

^a P values are from the χ^2 test.

^b P values are from the Cochran–Mantel–Haenszel row mean score test, stratified by site and using modified ridits.

^c Venlafaxine vs. placebo $P < 0.025$.

^d Venlafaxine vs. placebo $P < 0.001$; fluoxetine vs. placebo $P = 0.024$.

^e Venlafaxine vs. placebo $P < 0.001$; fluoxetine vs. placebo $P = 0.046$; venlafaxine vs. fluoxetine $P = 0.034$.

^f Venlafaxine vs. placebo $P = 0.009$.

^g Remission based HAM-D₁₇ ≤ 7 .

Table 4
ETRANK[®] test statistic data summary

	Venlafaxine vs. placebo		Venlafaxine vs. fluoxetine		Fluoxetine vs. placebo	
	Full data	Endpoint	Full data	Endpoint	Full data	Endpoint
ScF1	7.29*	9.27**	3.29	4.23***	1.28	0.82
Equal	8.03**	5.97***	2.93	2.45	2.24	0.45
Observed	8.38**	9.27**	3.32	4.27***	1.65	0.78
Entsuh	5.14***	7.44**	2.87	4.12***	1.14	0.60

Definitions: ScF1 (score function 1), scores that assign heavier weights to later observations; equal, scores that are equal to actual ranks; observed, the observed data; Entsuh, scores that weigh middle observations more similarly to each other than extreme observations.

* $P < 0.001$.

** $P < 0.01$.

*** $P < 0.05$.

significant differences were observed with the fluoxetine-placebo comparison.

3.4. Quality-of-life measures

A statistically significant difference was observed on only 1 of the 5 quality-of-life measures (general life functioning; Table 5). On that variable, there was greater improvement in the venlafaxine group compared with the fluoxetine and placebo groups (venlafaxine vs. fluoxetine: $F(1, 236) = 4.61$, $P = 0.033$; venlafaxine vs. placebo: $F(1, 236) = 9.99$, $P = 0.002$).

3.5. Safety and tolerability assessments

Rates of discontinuation due to AEs significantly differed among treatment groups, with rates of 12%, 7%, and 3% in the venlafaxine, fluoxetine, and placebo groups, respectively ($\chi^2 = 6.03$; $P = 0.049$) (Table 2). There were significant overall differences in the incidence of 6 side effects – anxiety, constipation dizziness, nausea, sweating, and vomiting (Table 6). Among these, the prevalence of nausea ($\chi^2 = 8.06$; $P = 0.005$), sweating ($\chi^2 = 6.32$; $P = 0.012$), and constipation ($\chi^2 = 5.84$; $P = 0.016$) was significantly higher during venlafaxine

therapy compared to fluoxetine therapy. Nausea was the most common AE and occurred in 40% of the venlafaxine-treated patients, as compared with 22% and 8% of the patients treated with fluoxetine and placebo, respectively. Nausea tended to be transient, however, and less than 10% of those affected still reported this AE at week 6. Serious AEs were reported for 3 patients, 1 in each treatment group. None of these events was considered related to the study medication.

Mean changes in vital signs are presented in Table 7. Statistically significant differences were observed for supine pulse, supine diastolic blood pressure, and weight. Specifically, the increase in mean supine pulse rate associated with venlafaxine treatment and the decrease associated with fluoxetine treatment were statistically significant compared with placebo; the increase in mean supine diastolic blood pressure with venlafaxine treatment was significant compared to the decrease associated with placebo; fluoxetine therapy resulted in significantly greater weight loss than venlafaxine therapy, which in turn resulted in significantly greater weight loss than placebo. No differences were observed among the three treatments with respect to significant changes in ECG rhythms or onset of clinically significant abnormalities.

Table 5
Scores for the patient-reported outcome measures at end of study

	Venlafaxine ($n = 84$)	Fluoxetine ($n = 87$)	Placebo ($n = 81$)	F ; (n , df)	Overall P value*
General Life Functioning Total Score	55.7 (11.0)	52.8 (9.8)	50.9 (11.5)	5.21; (2, 236)	0.006**
Activities Questionnaire Total Score	53.0 (11.5)	52.3 (9.7)	50.4 (11.3)	1.55; (2, 236)	0.214
Cognitive Functioning	27.7 (6.5)	27.2 (5.5)	27.1 (6.1)	0.61; (2, 236)	0.543
General Health	18.8 (4.5)	19.1 (3.8)	18.5 (4.1)	0.41; (2, 236)	0.666
Vitality	13.3 (4.8)	12.6 (4.2)	12.2 (5.0)	1.34; (2, 236)	0.263

Questionnaire scores are mean (SD).

Higher scores correspond to a higher level of functioning.

* P values are from a two-way ANCOVA with baseline score as covariate and with treatment and investigator as factors at the last on-therapy visit.

** Venlafaxine vs fluoxetine: ($F(1, 236) = 4.61$, $P = 0.033$); venlafaxine vs. placebo ($F(1, 236) = 9.99$, $P = 0.002$); fluoxetine vs placebo ($F(1, 236) = 1.11$, $P = 0.293$).

Table 6
Percentage of patients in each group reporting treatment-emergent adverse events^a

Adverse event	Venlafaxine ($n = 100$)	Fluoxetine ($n = 102$)	Placebo ($n = 102$)	χ^2	Overall P value
Nausea	40	22	8	29.64	<0.001
Headache	36	24	33	3.92	0.129
Dry mouth	24	16	15	3.54	0.170
Insomnia	22	15	14	2.95	0.229
Dyspepsia	9	19	16	3.96	0.138
Sweating	14	4	2	13.67	0.001
Diarrhea	9	13	9	1.09	0.580
Dizziness	13	8	3	7.03	0.030
Vomiting	11	5	2	7.69	0.021
Fatigue	10	10	5	2.25	0.325
Anxiety	10	7	1	7.62	0.022
Constipation	10	2	5	6.32	0.042

^a Includes only events reported by $\geq 10\%$ of patients in at least one treatment group.

Table 7
Mean (SD) changes in vital signs^a

	Venlafaxine (<i>n</i> = 96)	Fluoxetine (<i>n</i> = 101)	Placebo (<i>n</i> = 101)	<i>F</i> Test (df)	Overall <i>P</i> value ^b
Supine pulse rate (beats/min)	4.1 (10.3)	−1.6 (8.3)	1.3 (9.1)	12.42 (2,282)	<0.001 ^c
Supine systolic BP (mm Hg)	2.3 (9.5)	0.4 (10.2)	−1.2 (10.1)	3.01 (2,281)	0.051
Supine diastolic BP (mm Hg)	1.6 (6.9)	0.2 (7.7)	−1.3 (7.5)	3.93 (2,281)	0.021 ^d
Standing systolic BP (mm Hg)	0.3 (9.7)	1.1 (9.3)	−2.0 (9.1)	2.83 (2,281)	0.060
Standing diastolic BP (mm Hg)	1.1 (7.3)	−0.2 (7.7)	−0.4 (7.2)	1.42 (2,281)	0.242
Weight (kg)	−0.9 (1.9)	−1.3 (1.9)	−0.03 (1.5)	13.55 (2,278)	<0.001 ^e

^a Changes are from baseline to last on-therapy visit.

^b *P* values are from a two-way ANCOVA with baseline score as covariate and with treatment and investigator as factors at the last on-therapy visit.

^c Venlafaxine vs. fluoxetine: *P* < 0.0001, *F* = 24.37; venlafaxine vs. placebo: *P* = 0.0021, *F* = 9.68.

^d Venlafaxine vs. placebo: *P* = 0.0055, *F* = 7.83.

^e Venlafaxine vs. placebo: *P* = 0.0018, *F* = 9.93; fluoxetine vs. placebo: *P* < 0.0001, *F* = 26.61; venlafaxine vs. fluoxetine: *P* = 0.05, *F* = 3.88.

4. Discussion

In this randomized clinical trial (RCT), venlafaxine therapy was significantly more effective than placebo on most outcomes, whereas fluoxetine therapy was characterized by a less consistent pattern of efficacy. The outcomes of the 2 active therapies were generally not statistically significant, though venlafaxine therapy was significantly more effective than fluoxetine therapy on the ETRANK analyses of endpoint data. Although the general pattern of results tended to favor venlafaxine over fluoxetine, none of the differences on the original 5 primary efficacy variables were large enough to remain statistically significant if the critical alpha value was corrected to account for multiple comparisons. There were also differences between drugs on several different measures of tolerability, which (with the exception of weight loss) favored fluoxetine.

Overall, the results of this trial illustrate the challenges now facing clinical pharmacotherapy studies. For example, there is clearly considerable difficulty demonstrating that effective antidepressants actually are superior to placebo in contemporary RCTs. Specifically, it is now known that 50% of contemporary RCTs fail to demonstrate significant drug-placebo differences and the average drug-placebo difference is only about two HAM-D points, ie, an effect size of 0.2 (Khan et al., 2000, 2002; Kirsch et al., 1998). In the current trial, there was only about 50% power to detect a difference of this magnitude favoring either of the active antidepressants over placebo. It therefore is not surprising that the active antidepressants studied showed inconsistent efficacy.

A second related point is the relative futility of trying to compare the efficacy of active antidepressants in a conventional study. Although one of the goals of the current RCT was to compare venlafaxine and fluoxetine, this study was ultimately underpowered (based on the effect sizes reported by Thase et al. (2001) and Smith et al. (2002) the current study had < 20% power to detect the differences between venlafaxine and fluoxetine). In retrospect, this study would have had to enroll more than 600 additional patients to have achieved the desired

80% power to detect the expected between-group differences. Thus, the risk of type 2 error was so high that the study must be considered to have failed *vis a vis* questions of relative efficacy.

Several more study-specific limitations also deserve comment. First, the dosing protocol for the antidepressants was somewhat imbalanced. Specifically, the maximum dose of venlafaxine was 60% of the FDA-approved maximum (i.e., 225/375 mg/day), whereas the maximum dose of fluoxetine was 75% of the approved maximum (i.e., 60/80 mg/day). There may have been better separation of the venlafaxine and fluoxetine groups that had a proportionally comparable dose of venlafaxine (i.e., 300 mg/day) been permitted. Second, by limiting the protocol to 6 weeks and imposing submaximal dosing ceilings for both active treatments, it is likely that the efficacy of both drugs relative to placebo was underestimated. Third, it is noteworthy that only about one half of the patients received the maximally permitted doses of the active study medications, despite the fact that less than one third of the patients remitted. In future studies it would be useful to implement more rigorous protocol monitoring procedures to ensure that patients receive optimal treatment trials. Finally, it should be noted that the twice-daily dosing protocol used in this study is no longer necessary because of the availability of the extended release formulation, and that some of the observed differences in tolerability might have been lessened by use of the newer formulation of venlafaxine.

Disclosure statement

CB Nemeroff – Grants/Research: AFSP, AstraZeneca, Bristol-Myers-Squibb, Forest Laboratories, Janssen Pharmaceutica, NARSAD, NIMH, Pfizer Pharmaceuticals, Wyeth-Ayerst; Consultant: Abbott Laboratories, Acadia Pharmaceuticals, Bristol-Myers-Squibb, Corcept, Cypress Biosciences, Cyberonics, Eli Lilly, Entrepreneur's Fund, Forest Laboratories, GlaxoSmithKline, i3 DLN, Janssen Pharmaceutica, Lundbeck, Otsuka, Pfizer Pharmaceuticals, Quintiles, UCB Pharma, Wyeth-Ayerst. Speakers Bureau: Abbott Laboratories, GlaxoSmithKline, Janssen Pharmaceutica, Pfizer Pharmaceuticals.

Stockholder: Acadia Pharmaceuticals, Corcept, Cypress Biosciences, NovaDel. *Board of Directors:* American Foundation for Suicide Prevention (AFSP), American Psychiatric Institute for Research and Education (APIRE), George West Mental Health Foundation, Novadel Pharma, National Foundation for Mental Health (NFMH). *Patents:* Method and devices for transdermal delivery of lithium (US 6,375,990 B1). Method to estimate serotonin and norepinephrine transporter occupancy after drug treatment using patient or animal serum (provisional filing April, 2001). *Equity:* Reevox, BMC-JR LLC, CeNeRx.

ME Thase – Consultant: AstraZeneca, Bristol-Myers Squibb Company, Cephalon, Inc., Cyberonics, Inc., Eli Lilly & Co., Forest Laboratories, Inc., GlaxoSmithKline, Janssen Pharmaceutica, Novartis, Organon, Inc., Pfizer Pharmaceutical, Sepracor, Inc., Shire US Inc., Wyeth Pharmaceuticals. *Pharmaceutical Grant/Research:* None. *Speakers Bureau:* AstraZeneca, Bristol-Myers Squibb Company, Cyberonics, Inc., Eli Lilly & Co., GlaxoSmithKline, Sanofi Aventis, Wyeth Pharmaceuticals. *Major Stockholder:* None. *Other Financial or Material Support:* None.

Acknowledgements

The following investigators comprised the EPIC 014 Study Group: Robert Bielski, MD, Institute for Health Studies, Okemos, MI; Jacqueline Collins, MD, Department of Psychiatry, University of Cincinnati Clinical Trials Administration, Cincinnati, OH; Lynn Cunningham, MD, Vine Street Clinical Research Center, Springfield, IL; Ned Kalin, MD, Wisconsin Psychiatric Institute and Clinics, Madison, WI; Arifulla Khan, MD, Northwest Clinical Research Center, Bellevue, WA; Charles Nemeroff, MD, PhD, Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA; Teresa Pigott, MD, Houston, TX; Frederic Quitkin, MD, New York State Psychiatric Institute, New York, NY; Frederick Reimherr, MD, Department of Psychiatry, University of Utah, Salt Lake City, UT; Richard C. Shelton, MD, Nashville, TN; Steve Targum, MD, Clinical Studies, Philadelphia, PA; Michael E. Thase, MD, University of Pittsburgh and Western Psychiatric Institution and Clinic, Pittsburgh, PA; John Zajecka, MD, Rush Institute for Mental Well-Being, Chicago, IL.

The authors acknowledge Sherri D. Jones, PharmD, and Diane M. Sloan, PharmD, for their editorial assistance.

References

American Psychiatric Association, Diagnostic and statistical manual of mental disorders: DSM-IVTM. Washington (DC): American Psychiatric Association; 1994.

- Elkin I, Parloff MB, Hadley SW, et al. NIMH Treatment of Depression Collaborative Research Program: background and research plan. *Archives of General Psychiatry* 1985;42:305–16.
- Ellingrod VL, Perry PJ. Venlafaxine: a heterocyclic antidepressant. *American Journal of Hospital Pharmacy* 1994;51:3033–46.
- Entsuah R. ETRANK: a ranking procedure for handling missing data in clinical trials: application to venlafaxine extended-release depression trial. *Journal of Biopharmaceutical Statistics* 1996;6: 457–75.
- Entsuah AR. Randomization procedures for analyzing clinical trial data with treatment related withdrawals. *Communications in Statistics: Theory and Method* 1990;19:3859–80.
- Holliday SM, Benfield P. Venlafaxine: a review of its pharmacology and therapeutic potential in depression. *Drugs* 1995;49: 280–94.
- Khan A, Warner HA, Brown WA. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials. *Archives of General Psychiatry* 2000;57:311–7.
- Khan A, Leventhal RM, Khan SR, Brown WA. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *Journal of Clinical Psychopharmacology* 2002;22:40–5.
- Kirsch I, Moore TJ, Scoboria A, Nicholls SS. The emperor's new drugs: an analysis of antidepressant medication data submitted to the US Food and Drug Administration. *Prevention and Treatment* 1998;1 [Article 0002a].
- Lavori PW, Dawson R, Shera D. A multiple imputation strategy for clinical trials with truncation of patient data. *Statistics of Medicine* 1995;14:1913–25.
- Lenderking WR, Tennen H, Nackley JF, et al. The effects of venlafaxine on social activity level in depressed outpatients. *Journal of Clinical Psychiatry* 1999;60:157–63.
- Muth EA, Haskins JT, Moyer JA, Husbands GE, Nielsen ST, Sigg EB. Antidepressant biochemical profile of the novel bicyclic compound Wy-45,030, an ethyl cyclohexanol derivative. *Biochemical Pharmacology* 1986;35:4493–7.
- 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. *Journal of Affective Disorders* 1999;56:171–81.
- Silverstone PH, Ravindran A. Once-daily venlafaxine extended release (XR) compared with fluoxetine in outpatients with depression and anxiety. Venlafaxine XR 360 Study Group. *Journal of Clinical Psychiatry* 1999;60:22–8.
- Smith D, Dempster C, Glanville J, Freemantle N, Anderson I. Efficacy and tolerability of venlafaxine compared with selective serotonin reuptake inhibitors and other antidepressants: a meta-analysis. *British Journal of Psychiatry* 2002;180: 396–404.
- Thase ME. Comparing the methods used to compare antidepressants. *Psychopharmacological Bulletin* 2002;36(Suppl. 1):1–17.
- Thase ME, Entsuah AR, Rudolph RL. Remission rates during treatment with venlafaxine or selective serotonin reuptake inhibitors. *British Journal of Psychiatry* 2001;178:234–41.
- Ware Jr JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care* 1992;30:473–83.